Analysis and Geometry on Carnot-Carathéodory Spaces

Alexander Nagel

Contents

Preface iv

Part 1. Introduction 1

Chapter 1. Spaces of Homogeneous Type: Definitions and Examples 2
1. Spaces of homogeneous type 3
1.1. Pseudometrics and doubling measures 4
1.2. Covering Lemmas 6
1.3. The Hardy-Littlewood maximal operator 11
1.4. The Caldéron-Zygmund Decomposition 13
2. The Laplace operator and Euclidean Analysis 15
2.1. Symmetries of the Laplace operator 16
2.2. The Newtonian Potential 17
2.3. The role of Euclidean geometry 19
2.4. Hardy-Littlewood-Sobolev estimates 21
2.5. Lipschitz Estimates 23
2.6. L^2-estimates 28
2.7. L^1-estimates 31
3. The heat operator and a non-isotropic metric on $\mathbb{R}^{n+1}$ 33
3.1. Symmetry properties 33
3.2. The initial value problem and a fundamental solution 34
3.3. Parabolic convergence 38
4. Operators on the Heisenberg group 39
4.1. The Siegal upper half space and its boundary 39
4.2. Dilations and translations on $\mathbb{H}_n$ 41
4.3. The sub-Laplacian and its geometry 42
4.4. The space $H^2(\mathbb{H}^n)$ and the Szegö Projection 45
4.5. Weighted spaces of entire functions 46
5. The Grushin plane 49
5.1. The geometry 50
5.2. Fundamental solution for $\mathcal{L}$ 50
5.3. Hermite functions 51

Chapter 2. VECTOR FIELDS 53
1. Vector fields and commutators 54
1.1. Germs of functions, tangent vectors, and vector fields 54
1.2. Commutators 55
1.3. Vector fields of finite type 56
1.4. Derivatives of determinants 58
2. Vector fields and smooth mappings 59

2.1. The differential acting on tangent vectors, vector fields, and commutators 60
2.2. Action on determinants 61
3. Integral curves and the exponential maps 61
3.1. Integral curves 62
3.2. The exponential mapping 63
3.3. Exponential maps of several vector fields 65
4. Composition of flows 67
4.1. Commuting vector fields 67
4.2. Non-commuting vector fields 71
5. Normal coordinates 74
5.1. The definition of Θ_x 74
5.2. The action of $d\Theta_x$ on tangent vectors 75
6. Submanifolds and tangent vectors 76
6.1. Submanifolds 76
6.2. Tangential vector fields 77
6.3. The Frobenius theorem 77
7. Normal forms for vector fields 80
7.1. The case of a single vector field 80
7.2. The case of two vector fields 81
8. Vector fields and Derivations 81
8.1. Factorization 82
8.2. Derivations 83

Chapter 3. CARNOT-CARATHÉODORY METRICS 84
1. Construction of Carnot-Carathéodory metrics 85
1.1. Homogeneous norms 85
1.2. Control systems 86
1.3. Carnot-Carathéodory metrics 88
1.4. Vector fields of finite type 92
2. Examples of metrics and operators 94
2.1. Isotropic Euclidean space 94
2.2. Non-isotropic Euclidean space 95
2.3. The Heisenberg group 96
2.4. The Grushin plane 101
3. Local structure of Carnot-Carathéodory metrics 101
3.1. Notation 102
3.2. Exponential Balls 103
3.3. (x, δ, η)-dominance 105
3.4. The local structure of Carnot-Carathéodory balls 105
4. Proof of the main theorem 107
4.1. Algebraic preliminaries 107
4.2. Estimates at x using (x, δ, κ)-dominance 111
4.3. Estimates on exponential balls 112
4.4. Some topological remarks 115
4.5. Proof of Theorem **??**, part (**??**) 116
4.6. Proof of Theorem **??**, part (**??**) 120
4.7. Proof of Theorem **??**, part (**??**) 122
4.8. Proof of Theorem **??**, part (**??**) 124

5. Smooth Metrics 125
5.1. Differentiable metrics 125
5.2. Scaled bump functions 129

Chapter 4. Subelliptic estimates and hypoellipticity 130
1. Introduction 130
2. Subelliptic estimates 130
2.1. Statement of the main theorem 130
2.2. Hypoellipticity of $\mathcal{L}$ 132
2.3. Commentary on the argument 132
3. Estimates for $\sum_{j=1}^{p} \big|\big| X_j\varphi \big|\big|_s^2$ 133
3.1. Integration by parts 134
3.2. The basic L^2-identity and L^2-inequality 135
3.3. Commutators 136
3.4. The passage from L^2 to H^s 141
4. Estimates for smooth functions 144
4.1. Introduction 144
4.2. The space of sub-elliptic multipliers 145
4.3. The main theorem on sub-elliptic multipliers 147
5. The theorem for distributions 153
5.1. Mollifiers 153
5.2. The main result 154
6. Examples of vector fields Y_0 satisfying hypothesis (H-2) 154

Chapter 5. Estimates for fundamental solutions 160
1. Introduction 160
2. Unbounded operators on Hilbert space 161
2.1. Closed, densely defined operators and their adjoints 162
2.2. The spectrum of an operator 165
2.3. Self-adjoint operators 166
2.4. The spectral theorem for self-adjoint operators 167
3. The initial value problem for the heat operator $\partial_t + \mathcal{L}_x$ 168
3.1. The Friedrich's construction 168
3.2. The heat semigroup $\{e^{-t\mathcal{L}}\}$ 171
4. The heat kernel and heat equation 175
4.1. The heat kernel 175
4.2. The heat equation on $\mathbb{R} \times \Omega$ 178
5. Pointwise estimates for H 180
5.1. Scaling maps 180
5.2. A bound for the scaled initial value problem 180
5.3. Pointwise estimates for H 182
5.4. Action of $e^{-t\mathcal{L}}$ on bump functions 184

Chapter 6. Non-isotropic smoothing operators 185
1. Definitions and properties of NIS operators 185
1.1. Definition 185
1.2. Elementary properties 186

Chapter 7. Algebras 191
1. Associative Algebras 191

1.1. Free Associative Algebras 191
1.2. Algebras of formal series 193
2. Lie Algebras 195
2.1. Iterated Commutators 196
3. Universal Enveloping Algebras 198
3.1. Construction of $\mathfrak{U}(\mathfrak{L})$ 198
3.2. The Poincaré-Birkhoff-Witt Theorem 199
4. Free Lie Algebras 202
4.1. The construction 202
4.2. Characterizations of Lie elements 204
5. The Campbell-Baker-Hausdorff formula 207
6. Hall Bases for Free Lie Algebras 208
6.1. Hall Sets 209
6.2. Two Examples 209
6.3. Proof of Theorem **??** 210

Chapter 8. The $\overline{\partial}$ and $\overline{\partial}$-Neumann Problems 212
1. Differential forms and the d and $\overline{\partial}$-operators 212
1.1. Real differential forms 212
1.2. Notation 213
1.3. Differential forms with coordinates 214
1.4. The exterior derivative d 215
1.5. Complex coordinates and holomorphic functions 217
1.6. Splitting the complexified tangent space 218
1.7. Complex differential forms 219
1.8. The $\overline{\partial}$-operator 220
1.9. The $\overline{\partial}$-problem 222
2. The $\overline{\partial}$-Neumann problem 223
2.1. A finite dimensional analogue 223
2.2. Hilbert spaces 224
2.3. Notation 229
2.4. Computation of $\overline{\partial}$ and its formal adjoint 232
2.5. The basic identity 233
2.6. Further estimates 238
2.7. Hilbert spaces 239
2.8. The Neumann operator 242

Chapter 9. Appendix on background material 243
1. Distributions, Fourier transforms, and Fundamental Solutions 243
1.1. Notation 243
1.2. Spaces of smooth functions 243
1.3. Spaces of distributions 244
1.4. The Fourier Transform 245
1.5. Fundamental solutions and parametrices 246

Chapter 10. Pseudodifferential operators 248
1. Functions, distributions, and Fourier transforms 248
1.1. Spaces of smooth functions 248
1.2. Spaces of distributions 249

2. Pseudodifferential Operators 251
2.1. Results 252

Bibliography 254

Preface

The subject matter of this book lies at the interface of the fields of harmonic analysis, complex analysis, and linear partial differential equations, and has been at the center of considerable research effort since at least the late 1960's. Some aspects of this work are presented in monographs and texts. Any brief list would have to include:

- G.B. Folland and J.J. Kohn's monograph, *The Neumann Problem for the Cauchy-Riemann Complex* [**FK72**], on the $\overline{\partial}$-Neumann problem;
- L. Hörmander's books, *An Introduction to Complex Analysis in Several Variables* [**Hö66**] and *Notions of Convexity* [**Hö94**], on estimates for the $\overline{\partial}$-problem;
- L. Hörmander's encyclopedic *The Analysis of Linear Partial Differential Operators I - IV* [**Hö85**] which contains material on hypoellipticity of sums of squares of vector fields;
- E.M. Stein's short monograph *Boundary Behavior of Holomorphic Functions of Several Complex Variables* [**Ste72**];
- E.M. Stein's definitive classic *Harmonic Analysis: Real-Variables Methods, Orthogonality, and Oscillatory Integrals* [**Ste93**];
- The article by M. Gromov in the collection *Sub-Riemannian Geomerty* [**BR96**] by A. Bellaïche and J.-J. Risler;
- The recent book by S.-C. Chen and M.-C. Shaw *Partial Differential Equations in Several Complex Variables* [**CS01**].

Despite this list of references, many of the developments that have occurred in this subject remain largely unchronicled except in the original papers. Also, many of these results require general techniques that are not fully discussed in the earlier texts. This situation makes it difficult for a student to start work in this area. The most recent papers refer to older papers, which in turn cite earlier work, and a student often becomes discouraged at the prospect of trying to navigate this seemingly infinite regress.

Thus this book has two objectives. The first is to provide an accessible reference for material and techniques from the subject of 'control' or Carnot-Carathéodory metrics. The second is to provide a coherent account of some applications of these techniques to problems in complex analysis, harmonic analysis, and linear partial differential equations. These two objectives are inseparable. One needs the general theory of the geometry of Carnot-Carathéodory metrics in order to deal with certain kinds of problems in complex and harmonic analysis, but the theory by itself is essentially indigestible unless leavened with interesting problems and examples.

Part I of this book provides an introduction to the geometry associated to certain families of vector fields, and to analytic results about a related class of integral of operators. The basic geometric structures go by various names such as *control metrics*, *Carnot-Carathéodory spaces*, or *sub-Riemannian manifolds.* The associated analytic objects, in this general context, are often known as *non-isotropic smoothing* (NIS) operators. These concepts arose in part through attempts to present a more unified description of classical results, and from the need for more flexible tools to deal with new problems and phenomena arising in complex and harmonic analysis and linear partial differential equations.

Part 1

Introduction

CHAPTER 1

Spaces of Homogeneous Type: Definitions and Examples

In Part II of this book we show how it is possible to construct a metric from families of first order partial differential operators on an open subset of $\mathbb{R}^n$, and then how this geometry can be used in the analysis of certain second order operators built from the first order data. This presentation will involve a long and involved geometric construction, as well as technical analytic arguments. To make this material palatable, it is important to keep in mind the examples and objectives that led to the general theory. In this first chapter of Part I we prepare for the later technical work by first presenting basic results about metrics and the associated families of balls, and then discussing four classical examples where we can see more or less directly the connection between the underlying geometry and the analytic problems. In each case we consider the relationship between a particular partial differential operator and a corresponding notion of distance on $\mathbb{R}^n$. By presenting a different kind of result in each example, we hope to motivate the later development of a general theory.

In Section 1, we give the definition of a *space of homogeneous type*, where it makes sense to talk about lengths and volumes. We then establish analogues in this setting of some classical covering lemmas in Euclidean space. Finally, we show that one can establish an analogue of the Hardy-Littlewood maximal theorem and the Caldéron-Zygmund decomposition of integrable functions in this context.

In Section 2, we study the Laplace operator

$$\triangle = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}.$$

This is of course the sum of the squares of the first order operators $\frac{\partial}{\partial x_j}$. Here the appropriate geometry is given by the standard Euclidean metric on $\mathbb{R}^n$. We construct a fundamental solution for $\triangle$ given by convolution with the Newtonian potential, and then show how classical regularity properties of the Laplace operator follow from arguments using Euclidean geometry.

In Section 3, we consider the heat operator

$$\frac{\partial}{\partial t} - \triangle_x,$$

and show that the appropriate geometry is now a non-isotropic metric on $\mathbb{R}^{n+1}$. The heat operator is a first order operator $\dfrac{\partial}{\partial t}$ minus the sum of squares of the operators $\frac{\partial}{\partial x_j}$. In particular, we show that this metric is reflected in the boundary behavior of functions satisfying a classical initial value problem for the heat operator.

In Section 4, we study a model problem arising in complex analysis in several variables. We see that the space $\mathbb{C}^n \times \mathbb{R} = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ can be identified with the boundary of a domain in $\mathbb{C}^{n+1}$, and that on this boundary there is, in a natural way, the structure of a nilpotent Lie group, called the 'Heisenberg group'. The appropriate metric in this case is then invariant under this non-commutative group structure. The first order operators $X_j = \frac{\partial}{\partial x_j} + 2y_j \frac{\partial}{\partial t}$ and $Y_j = \frac{\partial}{\partial y_j} - 2x_j \frac{\partial}{\partial t}$ are left invariant on this group, and a natural analogue of the Laplace operator in this case is the operator

$$\begin{aligned}\mathcal{L} &= \sum_{j=1}^{n} X_j^2 + Y_j^2 \\ &= \sum_{j=1}^{n} \Big[\frac{\partial^2}{\partial x_j^2} + \frac{\partial^2}{\partial y_j^2} + 2y_j \frac{\partial^2}{\partial x_j \partial t} - 2x_j \frac{\partial^2}{\partial y_j \partial t}\Big] + \sum_{j=1}^{n} (x_j^2 + y_j^2) \frac{\partial^2}{\partial t^2}.\end{aligned}$$

In addition, we consider the orthogonal projection from $L^2(\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R})$ to the closed subspace consisting of the functions annihilated by the n complex first order operators $\overline{Z}_j = X + j + iY_j$. We see that this operator, as well as the fundamental solution of $\mathcal{L}$ can be understood in terms of the given geometry.

Finally, in Section 5, we construct a fundamental solution for the operator

$$\frac{\partial^2}{\partial x^2} + x^2 \frac{\partial}{\partial y^2}$$

on $\mathbb{R}^2$. This leads to the study of what is sometimes called the 'Grushin plane'.

Our discussion is not an exhaustive account of the theory of these well-know operators. Rather, we focus on certain results which have analogues in the more general theory developed in later chapters. Our object is to show how symmetry properties of the operators are reflected in the corresponding distances, and, conversely, how the underlying geometry plays a role in various analytic results.

1. Spaces of homogeneous type

In this section we present an abstract framework in which we can discuss both distance and volume. There are many possible approaches to such a discussion, and we choose a definition of 'spaces of homogeneous type' which is a compromise between the competing needs for simplicity and abstractness. As we shall see, the key concept is a space equipped with a distance which defines a family of balls, and a measure which allows us to talk about the volumes of these balls. For our purposes, the most important requirements are:

(a) an "engufing" property which guarantees that if two balls of comparable size intersect, then each is contained in a fixed dilate of the other;
(b) a "doubling property" which guarantees that the volume of a ball of radius 2δ is bounded by a multiple of the volume of the ball of radius δ.

After giving the basic definitions in section 1.1, we prove analogues of the Vitali and Whitney covering theorems in section 1.2. We study the Hardy-Littlewood maximal operator in section 1.3, and we introduce the Caldéron-Zygmund decomposition of L^1 functions in section 1.4.

1.1. Pseudometrics and doubling measures.

The notion of an abstract space equipped with a distance is often formalized through the concept of a metric space. However, in anticipation of later examples, it will be more convenient for us to start with the weaker notion of a *pseudo-metric*.

DEFINITION 1.1. *A* pseudo-metric *on a set X is a function $\rho : X \times X \to \mathbb{R}$ with the following properties.*

(1) *For all $x, y \in X$, $\rho(x, y) \geq 0$, and $\rho(x, y) = 0$ if and only if $x = y$.*
(2) *For all $x, y \in X$, $\rho(x, y) = \rho(y, x)$.*
(3) *There is a constant $A_1 \geq 1$ so that for all $x, y, z \in X$ we have*

$$\rho(x, y) \leq A_1 \big[\rho(x, z) + \rho(z, y)\big].$$

A_1 is called the triangle inequality constant *for the pseudo-metric ρ. If we can take $A_1 = 1$, then ρ is called a* metric.

In our applications it will be possible to use several different metrics or pseudo-metrics, provided that they are equivalent in an appropriate sense. We make this precise as follows.

DEFINITION 1.2. *Let ρ_1 and ρ_2 be pseudo-metrics on a set X.*

(1) *ρ_1 is* globally equivalent *to ρ_2 if there is a constant $C > 0$ so that for all $x, y \in X$ we have*

$$C^{-1}\rho_2(x, y) \leq \rho_1(x, y) \leq C\,\rho_2(x, y).$$

(1) *ρ_1 is* locally equivalent *to ρ_2 if there are constants $C, \delta_0 > 0$ so that for all $x, y \in X$ with $\rho_1(x, y) < \delta_0$ we have*

$$C^{-1}\rho_2(x, y) \leq \rho_1(x, y) \leq C\,\rho_2(x, y).$$

It is easy to check that these are indeed equivalence relations. The notion of global equivalence is perhaps the more natural of the two, but the concept of local equivalence is useful when we are only interested is small distances.

If ρ is a pseudo-metric on a set X, then the ball with center $x \in X$ and radius $\delta > 0$ is the set

$$B_\rho(x; \delta) = \Big\{y \in X \,\Big|\, \rho(x, y) < \delta\Big\}.$$

Note that ρ_1 and ρ_2 are globally equivalent pseudo-metrics with comparability constant C if and only if for all $x \in X$ and $\delta > 0$ we have

$$B_{\rho_2}(x; \delta) \subset B_{\rho_1}(x; C\delta),$$
$$B_{\rho_1}(x; \delta) \subset B_{\rho_2}(x; C\delta).$$

The metrics are locally equivalent if and only if we have these inclusions for all sufficiently small δ.

Many properties of metric spaces cary over to spaces equipped with a pseudo-metric, but some care must be taken with repeated applications of the triangle inequality. In the following proposition, we focus on three easy consequences of the definitions.

PROPOSITION 1.3. *Suppose that ρ is a pseudo-metric on a space X with triangle inequality constant A_1.*

(i) *If* $\{x_0, x_1, \dots, x_m\}$ *are points in* X*, then* $\rho(x_0, x_m) \leq A_1^m \sum_{j=1}^m \rho(x_{j-1}, x_j)$.

(ii) *Let* $x_1, x_2 \in X$ *and* $\delta_1 \geq \delta_2$*. Then*

$$B_\rho(x_1; \delta_1) \cap B_\rho(x_2; \delta_2) \neq \emptyset \Longrightarrow B_\rho(x_2; \delta_2) \subset B_\rho(x_1; 3A_1^2 \delta_1).$$

(iii) *Suppose that* $y \in B_\rho(y_0; \delta_0)$ *and* $x \notin B_\rho(y_0; \eta\, \delta_0)$*. Then if* $\eta > A_1 \geq 1$,

$$\frac{1}{2A_1}\, \rho(x, y) \leq \rho(x, y_0) \leq \left(\frac{\eta\, A_1}{\eta - A_1}\right) \rho(x, y).$$

Part (ii) gives the basic engulfing property of balls, and part (iii) show that well outside a ball, the distance to the center and the distance to an arbitrary point of the ball are comparable.

PROOF. Part (i) follows easily by induction on m. To establish (ii), suppose $B_\rho(x_1; \delta_1) \cap B_\rho(x_2; \delta_2) \neq \emptyset$, and let $z \in B_\rho(x_1; \delta_1) \cap B_\rho(x_2; \delta_2)$. Let $y \in B_\rho(x_2; \delta_2)$. Then

$$\begin{aligned} \rho(x_1, y) &\leq A_1^2 \big[\rho(x_1, z) + \rho(z, x_2) + \rho(x_2, y)\big] \\ &< A_1^2 \big[\delta_1 + \delta_2 + \delta_2\big] \\ &\leq 3A_1^2 \delta_1, \end{aligned}$$

so $B_\rho(x_2; \delta_2) \subset B_\rho(x_1; 3A_1^1 \delta_1)$ as asserted. To establish part (iii), suppose $x \notin B_\rho(y_0; \eta\delta_0)$ and $y \in B_\rho(y_0; \delta_0)$. If $\eta \geq 1$ then $\rho(y_0, y) < \delta_0 \leq \rho(x, y_0)$, so the triangle inequality gives

$$\rho(x, y) \leq A_2 \big[\rho(x, y_0) + \rho(y_0, y)\big] < 2A_1 \rho(x, y_0).$$

On the other hand, we also have $\rho(y, y_0) < \delta_0 \leq \eta^{-1}\, \rho(x, y_0)$ so the triangle inequality gives

$$\rho(x, y_0) \leq A_1 \big[\rho(x, y) + \rho(y, y_0)\big] < A_1\, \rho(x, y) + A_2\, \eta^{-1} \rho(z_0, y),$$

and this gives the second inequality if $\eta > A_1$. □

We next introduce a measure so that we can talk about volumes of balls. Assumption (3) below is the basic doubling property.

DEFINITION 1.4. *Let* X *be a locally compact topological space equipped with a pseudo-metric* ρ *and a positive regular Borel measure* μ*. Then* (X, ρ, μ) *is a* space of homogeneous type *if the following conditions are satisfied:*

(1) *For each* $x \in X$*, the collection of balls* $\big\{B_\rho(x; \delta) : \delta > 0\big\}$ *are open and hence are* μ*-measurable, and they form a basis for the open neighborhoods of* x.

(2) *For all* $x \in X$ *and* $\delta > 0$, $0 < \mu\big(B_\rho(x; \delta)\big) < \infty$.

(3) *There is a constant* $A_2 > 0$ *so that for all* $x \in X$ *and* $\delta > 0$

$$\mu\big(B_\rho(x; 2\delta)\big) \leq A_2\, \mu\big(B_\rho(x; \delta)\big).$$

The constant A_2 *is called the* doubling constant *for the measure* μ.

The following result shows that the volume of $B(x; \delta)$ grows at most polynomially in δ.

PROPOSITION 1.5. *If (X, ρ, μ) is a space of homogeneous type with doubling constant A_2, there is a constant τ so that for $\lambda \geq 2$*

$$\mu\big(B_\rho(x; \lambda\,\delta) \leq \lambda^\tau\,\mu\big(B_\rho(x; \delta)\big).$$

PROOF. Let N be the positive integer such that $\lambda \leq 2^N < 2\lambda$, so that $\log_2(\lambda) \leq N < \log_2(\lambda) + 1$. Then using the doubling property N times yields

$$\mu\big(B_\rho(x; \lambda\,\delta)\big) \leq \mu\big(B_\rho(x; 2^N\,\delta)\big) \leq A_2^N\,\mu\big(B_\rho(x; \delta)\big) \leq \lambda^\tau\,\mu\big(B_\rho(x; \delta)\big)$$

where $\tau = 2\,\log_2(A_2)$. □

1.2. Covering Lemmas.

We now establish a number of results for spaces of homogeneous type which are analogues of standard results in Euclidean analysis. Many of the arguments involve only minor modifications of the classical proofs. We primarily follow the development in [**CW77**] and [**Ste93**]. Throughout this section, (X, ρ, μ) will denote a space of homogeneous type. We let A_1 be the the triangle inequality constant for ρ and let A_2 be the doubling constant for μ.

Our first result deals with estimates for the number of uniformly separated points in a fixed ball. In Euclidean geometry in $\mathbb{R}^n$, the volume of a ball of radius δ is proportional to δ^n. Thus if $0 < \eta < 1$ and if $B_1, \ldots, B_m$ are mutually disjoint balls of radius $\eta\,\delta$ all contained in a ball of radius δ, we must have $m \leq \eta^{-n}$, so m is bounded by a constant depending only on η and the dimension. For spaces of homogeneous type, it is not true in general that the volume of a ball is proportional to a fixed power of the radius. Nevertheless, we have the following result.

LEMMA 1.6. *Let $0 < \eta < 1$, let $x_1, \ldots, x_m \in B_\rho(x_0; \delta)$, and suppose that $\rho(x_j, x_k) \geq \eta\delta$ for all $1 \leq j \neq k \leq m$. Then m is bounded by a constant that depends only on A_1, A_2 and η.*

PROOF. First observe that the balls $\{B_\rho(x_j; \frac{\eta\delta}{2A_1})\}$, $1 \leq j \leq m$, are disjoint, for if $y \in B_\rho(x_j, \frac{\eta\delta}{2A_1}) \cap B_\rho(x_k; \frac{\eta\delta}{2A_1})$, then

$$\rho(x_j, x_k) \leq A_1[\rho(x_j, y) + \rho(y, x_k)] < A_1\Big[\frac{\eta\delta}{2A_1} + \frac{\eta\delta}{2A_1}\Big] = \eta\delta,$$

which is a contradiction. Also, each ball $B_\rho(x_j; \frac{\eta\delta}{2A_1}) \subset B_\rho\big(x_0, 2A_1\,\delta\big)$, for if $y \in B_\rho(x_j; \frac{\eta\delta}{2A_1})$, then

$$\rho(x_0, y) \leq A_1[\rho(x_0, x_j) + \rho(x_j, y)] < A_1[\delta + \frac{\eta\delta}{2A_1}] \leq 2A_1\delta.$$

Finally, $B_\rho(x_0; \delta) \subset B_\rho(x_j; 2A_1\,\delta)$, for if $y \in B_\rho(x_0; \delta)$, then

$$\rho(x_j, y) \leq A_1[\rho(x_j, x_0) + \rho(x_0, y)] < 2A_1\delta.$$

Now let N_1 and N_2 be the smallest positive integers such that $2A_1 \leq 2^{N_1}$ and $(2A_1)^2 \leq 2^{N_2}\eta$. Then by the doubling property, for any $x \in X$ and any $\eta > 0$ we have

$$\mu\big(B_\rho(x_0; 2A_1\delta)\big) \leq A_2^{N_1}\,\mu\Big(B_\rho\big(x_0; \delta\big)\Big), \quad \text{and}$$

$$\mu\big(B_\rho(x; 2A_1\delta)\big) \leq A_2^{N_2}\,\mu\big(B_\rho(x; \frac{\eta\delta}{2A_1})\big).$$

Putting all this together, we have

$$\begin{aligned} m\,\mu\big(B_\rho(x_0;\delta)\big) &\le \sum_{j=1}^{m} \mu\big(B_\rho(x_j;2A_1\,\delta)\big) \\ &\le A_2^{N_2} \sum_{j=1}^{m} \mu\big(B_\rho(x_j;\frac{\eta\delta}{2A_1})\big) \\ &\le A_2^{N_2 1}\,\mu\big(B_\rho\big(x_0;2A_1\delta\big)\big) \\ &\le A_2^{N_1+N_2}\,\mu\big(B_\rho(x_0;\delta)\big). \end{aligned}$$

Thus $m \le A_2^{N_1+N_2}$, which completes the proof. □

Our next result is a variant of the Vitali covering lemma. A discussion of various classical covering lemmas in Euclidean spaces can be found, for example, in [**dG75**], Chapter I. We shall say that a set $E \subset X$ is *bounded* if there exist $y \in X$ and $R > 0$ so that $E \subset B_\rho(y;R)$.

LEMMA 1.7. *Let $E \subset X$ be a set and let $\mathcal{A}$ be an index set. Suppose that for each $\alpha \in \mathcal{A}$, there exist $x_\alpha \in X$ and $\delta_\alpha > 0$ so that $E \subset \bigcup_{\alpha\in\mathcal{A}} B_\rho(x_\alpha;\delta_\alpha)$. Suppose also that one of the following conditions is satisfied:*

(a) *The set E is bounded, and for every $\alpha \in \mathcal{A}$, the point $x_\alpha \in E$.*
(b) *There are no restrictions on the set E or the points x_α, but $\sup_{\alpha\in\mathcal{A}} \delta_\alpha = M < \infty$.*

Then there exists a finite or countable sub-collection of these balls,

$$\big\{B_1 = B_\rho(x_1;\delta_1), \dots, B_k = B_\rho(x_k;\delta_k), \dots\big\},$$

so that:

(1) *The balls are mutually disjoint: $B_j \cap B_k = \emptyset$ if $j \neq k$;*

(2) *If $B_j^* = B_\rho(x_j;3A_1^2\,\delta_j)$, then $E \subset \bigcup_j B_j^*$;*

(3) $\mu(E) \le C \sum_j \mu(B_j)$.

The constant C depends only on A_1 and A_2.

PROOF. If E is bounded, we may suppose that $E \subset B_\rho(y;R)$. Suppose that $\sup_{\alpha\in\mathcal{A}} \delta_\alpha = +\infty$. Then there exists $\alpha \in \mathcal{A}$ with $x_\alpha \in E$ and $\delta_\alpha > 2A_1R$. For any $z \in B_\rho(y,R)$ we have

$$\rho(x_\alpha,z) \le A_1[\rho(x_\alpha,y)+\rho(y,z)] < 2A_1R < \delta_\alpha,$$

so $E \subset B_\rho(y;R) \subset B_\rho(x_\alpha;\delta_\alpha)$. In this case, we can choose the sub-collection to consist of the single ball $B_1 = B_\rho(x_\alpha,\delta_{x_\alpha})$.

Thus from now on we shall assume that $M_1 = \sup_{\alpha\in\mathcal{A}} \delta_x < +\infty$. We select the sequence $\{B_j\}$ as follows. Set $M_1 = M$ and $\mathcal{A}_1 = \mathcal{A}$. Choose $B_1 = B_\rho(x_1;\delta_1)$ to be any ball $B(x_\alpha;\delta_\alpha)$ with $\alpha \in \mathcal{A}_1$ for which $\delta_1 = \delta_\alpha > \frac{1}{2}M_1$. Once B_1 has been picked, let

$$\mathcal{A}_2 = \Big\{\alpha \in \mathcal{A} \,\Big|\, B_\rho(x_\alpha;\delta_\alpha) \cap B_1 = \emptyset\Big\}$$

and let $M_2 = \sup_{\alpha\in\mathcal{A}_2} \delta_\alpha$.

Proceeding by induction, suppose that we have picked balls $B_1, \ldots, B_k$, sets $\mathcal{A} = \mathcal{A}_1 \supset \mathcal{A}_2 \supset \cdots \supset \mathcal{A}_k \supset \mathcal{A}_{k+1}$, and real numbers $M_1 \geq M_2 \geq \cdots \geq M_k \geq M_{k+1}$ so that

(i) $B_j = B_\rho(x_j;\delta_j)$ with $x_j = x_\alpha$ and $\delta_j = \delta_\alpha$ for some $\alpha \in \mathcal{A}_j$;

(ii) $M_j = \sup\limits_{\alpha\in\mathcal{A}_j} \delta_\alpha$ for $1 \leq j \leq k+1$;

(iii) $\delta_j > \frac{1}{2}M_j$ for $1 \leq j \leq k$;

(iv) $\mathcal{A}_{j+1} = \Big\{\alpha \in \mathcal{A}_j \,\Big|\, B_\rho(x_\alpha;\delta_\alpha) \cap B_j = \emptyset\Big\}$ for $1 \leq j \leq k$.

If the set $\mathcal{A}_{k+1}$ is empty, the selection process stops. If $\mathcal{A}_{k+1}$ is not empty, choose B_{k+1} to be a ball $B_\rho(x_{k+1};\delta_{k+1}) = B_\rho(x_\alpha;\delta_\alpha)$ with $x_\alpha \in \mathcal{A}_{k+1}$ and $\delta_\alpha > \frac{1}{2}M_{k+1}$. Once B_{k+1} is chosen, we set

$$\begin{aligned}\mathcal{A}_{k+2} &= \Big\{\alpha \in \mathcal{A}_{k+1} \,\Big|\, B_\rho(x_\alpha;\delta_\alpha) \cap B_{k+1} = \emptyset\Big\} \\ &= \Big\{\alpha \in \mathcal{A} \,\Big|\, B_\rho(x_\alpha;\delta_\alpha) \cap \Big(\bigcup_{j=1}^{k+1} B_j\Big) = \emptyset\Big\}.\end{aligned}$$

This completes the induction step, so with this process we have chosen a finite or countable set of balls $\{B_1, \ldots, B_k, \ldots\}$.

Let $j < k$. The center of B_k is $x_j = x_\alpha$ for some $\alpha \in \mathcal{A}_k$ by condition (i), and since $\mathcal{A}_k \subset \mathcal{A}_j$ it follows from condition (iv) that $B_k \cap B_j = \emptyset$. This establishes assertion (1) of the lemma.

Next we show that if the selection process leads to a countable collection of balls $\{B_k\}$, then $\lim_{k\to\infty} \delta_k = 0$. Clearly $\delta_{k+1} \leq \delta_k$. But if $\delta_k \geq \delta_0 > 0$ for all k, the ball $B_\rho(y;R)$ would contain the countable collection of disjoint balls $\{B_\rho(x_j;\delta_0)\}$, and this contradicts the conclusion of Lemma 1.6.

Now let $x_0 \in E$. Then $x_0 \in B_\rho(x_\alpha;\delta_\alpha)$ for some $\alpha \in \mathcal{A}$. If the selection process results in a finite set of balls, $B_\rho(x_\alpha;\delta_\alpha)$ must intersect one of them, for otherwise the selection process would not have stopped. If the selection process results in a countable set of balls, the above argument shows that there is an integer k so that $\delta_k < \frac{1}{2}\delta_\alpha$. But then $\delta_\alpha > 2\delta_k > M_k = \sup_{\beta\in\mathcal{A}_k} \delta_\beta$, and so $\alpha \notin \mathcal{A}_k$. Thus $B_\rho(x_\alpha,\delta_\alpha)$ must intersect one of the balls $\{B_1, \ldots, B_{k-1}\}$.

Thus in either case, there is a smallest positive integer j so that $B_\rho(x_{,\alpha};\delta_\alpha) \cap B_j \neq \emptyset$. It follows that $\alpha \in \mathcal{A}_j$, and hence $\delta_\alpha \leq M_j < 2\delta_{x_j}$. We also have

$$\emptyset \neq B_\rho(x_\alpha;\delta_\alpha) \cap B_j \subset B_\rho(x_\alpha;\delta_\alpha) \cap B_\rho(x_j, 2\delta_{x_j}).$$

It follows from Proposition 1.3, part (ii) that

$$x_0 \in B_\rho(x_\alpha;\delta_\alpha) \subset B_\rho(x_j;3A_1^2\delta_{x_j})$$

and hence

$$E \subset \bigcup_j B_\rho(x_j;3A_1^2\delta_{x_j}).$$

This is assertion (2).

Let N be the smallest positive integer such that $3A_1^2 \leq 2^N$. Then using the doubling property and assertion (2), we have

$$\mu(E) \leq \sum_j \mu\big(B_\rho(x_j; 3A_1^2\delta_{x_j})\big) \leq A_2^N \sum_j \mu\big(B_\rho(x_j;\delta_{x_j})\big) \tag{C}$$

which gives assertion (3), and completes the proof. □

Our next results are related to the classical Whitney covering lemma, which shows that one can decompose an proper open set $U \subset \mathbb{R}^n$ into a union of cubes Q_j such that the size of Q_j is comparable to the distance from Q_j to the boundary of U. For a general space of homogeneous type X, let $U \subsetneqq X$ be an open set. For any $x \in U$, let

$$d(x) = d_U(x) = \inf_{y \notin U} \rho(x,y) = \sup \Big\{ \delta > 0 \,\Big|\, B_\rho(x;\delta) \subset U \Big\}$$

denote the distance from x to the complement of U. Since $U \neq X$, it follows that $d(x) < \infty$. Since the balls $\{B_\rho(x;\delta)\}$ form a basis for the open neighborhoods of x, it follows that $d(x) > 0$.

We first show that if $U \subsetneqq X$ is open and if $x \in U$, then there is a ball centered at x with radius a small multiple of $d(x)$ so that the distances of all points in this ball to the complement of U are comparable.

Proposition 1.8. *Let $U \subsetneqq X$, and let $0 < \eta \leq \frac{1}{2A_1}$. For any $x \in X$, if $z \in B_\rho(x;\eta\, d(x))$, then*

$$(2A_1)^{-1}\, d(z) \leq d(x) \leq (2A_1)\, d(z).$$

Proof. Let $z \in B_\rho(x,\eta\, d(x))$. Then if $y \notin U$ we have

$$d(z) \leq \rho(z,y) \leq A_1[\rho(z,x) + \rho(x,y)] < A_1[\eta\, d(x) + \rho(x,y)].$$

Taking the infimum over all $y \neq U$ we get

$$d(z) \leq A_1\big[\eta + 1\big]\, d(x) \leq 2A_1\, d(x).$$

On the other hand, we have

$$\begin{aligned} d(x) &\leq \rho(x,y) \leq A_1\big[\rho(x,z) + \rho(z,y)\big] \\ &\leq A_1\eta\, d(x) + A_1\,\rho(z,y) \leq \frac{1}{2}\, d(x) + A_1\,\rho(z,y). \end{aligned}$$

Again taking the infimum over $y \neq U$, we get

$$d(x) \leq 2A_1\, d(z)$$

which completes the proof. □

Corollary 1.9. *Let $U \subsetneqq X$, and let $0 < \eta \leq \frac{1}{2A_1}$. If $x, y \in U$ and if $B_\rho\big(x;\eta\, d(x)\big) \cap B_\rho\big(y;\eta\, d(y)\big) \neq \emptyset$, then*

$$(2A_1)^{-2}\, d(y) \leq d(x) \leq (2A_1)^2\, d(y).$$

We can now establish an analogue of the Whitney covering theorem.

LEMMA 1.10. *Let $U \subsetneqq X$ be an open set. Then there is a collection of balls $\{B_j = B_\rho(x_j;\delta_j)\}$ with $\delta_j = \frac{1}{2}\, d(x_j)$ so that if $B_j^* = B_\rho(x_j;4\delta_j)$ and $B_j^\# = B_\rho\big(x_j;(12A_1^4)^{-1}\,\delta_j\big)$, then*

(1) For each j, $B_j \subset U$, and $B_j^* \cap (X-U) \neq \emptyset$.

(2) The balls $\{B_j^\#\}$ are mutually disjoint.

(3) $U = \bigcup_j B_j$.

(4) $\sum_j \mu(B_j) \leq C\,\mu(U)$.

(5) Each point of U belongs to at most $M < \infty$ of the balls $\{B_j\}$, where M depends only on the constants A_1 and A_2.

PROOF. For each $x \in U$, set $\delta(x) = \eta\, d(x)$ with $\eta = (24A_1^4)^{-1} \leq (2A_1)^{-1}$. The balls $\{B_\rho\big(x;d(x)\big)\}$ form an open cover of U. Let $\{B_\rho\big(x_j;\delta(x_j)\big)\}$ be a maximal disjoint sub-collection of these balls. Put

$$B_j = B_\rho\big(x_j, 12A_1^4\,\delta(x_j)\big) = B_\rho\Big(x_j;\frac{1}{2}d(x_j)\Big),$$

so that $B_j^* = B_\rho\big(x_j;2d(x_j)\big)$ and $B_j^\# = B_\rho\big(x;\delta(x_j)\big)$. It follows that $B_j \subset U$, $B_j^* \cap (X-U) \neq \emptyset$, and $\{B_j^\#\}$ are mutually disjoint. This proves assertions (1) and (2).

For every $x \in U$, the maximality of $\{B_\rho\big(x_j;\delta(x_j)\big)\}$ shows that there exists j so that

$$\emptyset \neq B_\rho\big(x;\delta(x)\big) \cap B_\rho\big(x_j;\delta(x_j)\big) \subset B_\rho\big(x;\delta(x)\big) \cap B_\rho\big(x_j;4A_1^2\,\delta(x_j)\big)$$

By Corollary 1.9, it follows that $d(x) \leq 4A_1^2\, d(x_j)$ and so $\delta(x) \leq 4A_1^2\,\delta(x_j)$. It follows from Proposition 1.3, part (ii), that

$$x \in B_\rho\big(x;\delta(x)\big) \subset B_\rho\big(x_j;12A_1^4\,\delta(x_j)\big) = B_j.$$

Thus $U = \bigcup_j B_j$, and so we have verified assertion (3).

Now let N be the smallest positive integer such that $12A_1^4 \leq 2^N$. Then $\mu(B_j) \leq A_2^N\,\mu(B_\rho\big(x_j;\delta(x_j)\big))$. Since the balls $\{B_\rho\big(x_j;\delta(x_j)\big)\}$ are disjoint and are contained in U, we have $\sum_j \mu(B_j) \leq A_2^N \mu(U)$. This gives assertion (4).

Now let $y \in U$ and suppose $y \in B_j$. It follows from Proposition 1.8 that $d(x_j) \leq 2A_1\, d(y)$ and $d(y) \leq 2A_1\, d(x_j)$. Thus if $z \in B_\rho\big(x_j;\delta(x_j)\big)$ we have

$$\begin{aligned}\rho(y,z) &\leq A_1\big[\rho(y,x_j) + \rho(x_j,z)\big]\\ &\leq A_1\big[12A_1^4 + 1\big]\,\delta(x_j)\\ &\leq 2A_1^2\big[12A_1^4+1\big]\,d(y).\end{aligned}$$

Thus

$$B_\rho\big(x_j;\delta(x_j)\big) \subset B_\rho\big(y;2A_1^2\big[12A_1^4+1\big]\,d(y)\big).$$

Since $d(y) \leq 2A_1 d(x_j) = 48A_1^5\,\delta(x_j)$, it follows that

$$B_\rho\big(x_j;(48A_1^5)^{-1}\,d(y)\big) \subset B_\rho\big(y;2A_1^2\big[12A_1^4+1\big]\,d(y)\big).$$

But for $j \neq k$

$$B_\rho\big(x_j; (48A_1^5)^{-1}\, d(y)\big) \cap B_\rho\big(x_k; (48A_1^5)^{-1}\, d(y)\big) = \emptyset,$$

and so by Lemma 1.6, the number of such balls B_j with $y \in B_j$ is bounded by a constant depending only on A_1 and A_2. This establishes (5) and completes the proof. □

1.3. The Hardy-Littlewood maximal operator.

We can now use the geometric information from the covering lemma to study the Hardy-Littlewood maximal operator. We begin by recalling the definition in this general context.

Definition 1.11. *Let (X, ρ, μ) be a space of homogeneous type. Let f be locally integrable on X. For $x \in X$ put*

$$\mathcal{M}[f](x) = \sup_{\delta>0} \sup_{x \in B_\rho(y;\delta)} \frac{1}{\mu\big(B_\rho(y;\delta)\big)} \int_{B_\rho(y;\delta)} |f(t)|\, dt.$$

Observe that if $\mathcal{M}[f](x) > \lambda$ there exists a ball $B = B_\rho(y;\delta)$ containing x such that the average of $|f|$ over B is greater than λ. But then $\mathcal{M}[f](z) > \lambda$ for all $z \in B$. Since B is open, this shows that $\{x \in X \,\big|\, \mathcal{M}[f](x) > \lambda\}$ is open, and hence $\mathcal{M}$ is lower semi-continuous and in particular measurable. Also, since every average is dominated by the supremum of the function, it is clear that if $f \in L^\infty(X)$, we have

$$\big|\big|\, \mathcal{M}[f] \,\big|\big|_{L^\infty(X)} \leq \big|\big|\, f \,\big|\big|_{L^\infty(X)}. \tag{1.1}$$

The following result is much deeper.

Theorem 1.12 (Hardy and Littlewood). *There is a constant C depending only on A_1 and A_2 so that for $1 \leq p < \infty$, the following statements are true:*

(1) *If $f \in L^1(X, d\mu)$, then*

$$\mu\big(\big\{x \in X \,\big|\, \mathcal{M}[f](x) > \lambda\big\}\big) \leq C\, \lambda^{-1}\, \big|\big|\, f \,\big|\big|_{L^1}.$$

(2) *If $1 < p < \infty$ and if $f \in L^p(X, d\mu)$, then*

$$\big|\big|\, \mathcal{M}[f] \,\big|\big|_{L^p} \leq 2C\, p\, (p-1)^{-1}\, \big|\big|\, f \,\big|\big|_{L^p}.$$

Proof. Let $\lambda > 0$ and let M be a positive integer. Let $E_{\lambda,M}$ denote the set of points $x \in X$ such that there exists $y \in X$ and $0 < \delta < M$ such that $x \in B_\rho(y;\delta)$ and

$$\frac{1}{\mu\big(B_\rho(y;\delta)\big)} \int_{B_\rho(x;\delta)} |f(t)|\, d\mu(t) > \lambda.$$

Then $E_{\lambda,M} \subset E_{\lambda,M+1}$ and

$$\bigcup_{M=1}^{\infty} E_{\lambda,M} = \Big\{x \in X \,\big|\, \mathcal{M}[f](x) > \lambda\Big\}.$$

If $x \in E_{\lambda,M}$, there exists $y \in X$ and $0 < \delta_y < M$ so that $x \in B_\rho(y, \delta_y)$, and

$$\frac{1}{\mu\big(B_\rho(y;\delta_y)\big)} \int_{B_\rho(y;\delta_y)} |f(t)|\, d\mu(t) > \lambda$$

or equivalently

$$\mu\big(B_\rho(y;\delta_y)\big) \leq \frac{1}{\lambda}\int_{B_\rho(y;\delta_y)} |f(t)|\,d\mu(t).$$

The balls $\{B_\rho(y;\delta_y)\}$ cover the set $E_{\lambda,M}$. By using the case of uniformly bounded radii in Lemma 1.7, we can find a sub-collection $\{B_1,\ldots,B_k,\ldots\}$ such that $B_j \cap B_k = \emptyset$ if $j \neq k$, and $\mu(E_{\lambda,M}) \leq C\sum_j \mu\big(B_j\big)$, where C depends only on A_1 and A_2. Then

$$\begin{aligned}\mu(E_{\lambda,M}) &\leq C\sum_{j=1}^N \mu(B_j) \leq \frac{C}{\lambda}\sum_{j=1}^N \int_{B_j} |f(t)|\,d\mu(t)\\ &\leq \frac{C}{\lambda}\int_X |f(t)|\,d\mu(t) = \frac{C}{\lambda}\,\big|\big|\,f\,\big|\big|_{L^1(X)}.\end{aligned}$$

This estimate is independent of M, and we conclude that

$$\mu\Big(\Big\{x\in X\,\Big|\,\mathcal{M}[f](x) > \lambda\Big\}\Big) \leq \frac{C}{\lambda}\,\big|\big|\,f\,\big|\big|_{L^1(X)}.$$

Thus proves assertion (1).

Assertion (2) follows from the Marcinkieicz interpolation theorem, for which one can consult [**Ste93**], pages 272-274. However, we give the proof in this special case. We use the fact that if $f \in L^p(X)$,

$$\big|\big|\,f\,\big|\big|_{L^p(X)}^p = p\int_0^\infty \lambda^{p-1}\,\mu\Big(\Big\{x\in X\,\Big|\,|f(x)| > \lambda\Big\}\Big)\,d\lambda.$$

For any $\lambda > 0$, let us write

$$f_\lambda(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq \lambda/4\\ 0 & \text{if } |f(x)| > \lambda/4\end{cases}, \qquad \text{and} \qquad f^\lambda(x) = \begin{cases} 0 & \text{if } |f(x)| \leq \lambda/4\\ f(x) & \text{if } |f(x)| > \lambda/4\end{cases}.$$

Then $f = f_\lambda + f^\lambda$, and so $\mathcal{M}[f](x) \leq \mathcal{M}[f_\lambda](x) + \mathcal{M}[f^\lambda](x)$. It follows that

$$\begin{aligned}&\mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f](x)| > \lambda\Big\}\Big)\\ &\quad\leq \mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f_\lambda](x)| > \frac{\lambda}{2}\Big\}\Big) + \mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f^\lambda](x)| > \frac{\lambda}{2}\Big\}\Big)\end{aligned} \tag{1.2}$$

Now $f_\lambda \in L^\infty(X)$, with $\big|\big|\,f_\lambda\,\big|\big|_{L^\infty(X)} \leq \frac{1}{4}\lambda$. It follows from equation (1.1) that $\mathcal{M}[f_\lambda](x) \leq \frac{1}{4}\lambda$ for almost all $x \in X$. Hence

$$\mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f_\lambda](x)| > \frac{\lambda}{2}\Big\}\Big) = 0. \tag{1.3}$$

On the other hand, we claim $f^\lambda \in L^1(X)$. In fact, since $f \in L^p(X)$, we have

$$\lambda^p\,\mu\Big(\Big\{x\in X\,\Big|\,|f(x)| > \lambda\Big\}\Big) \leq \int_X |f(x)|^p\,dx = \big|\big|\,f\,\big|\big|_{L^p(X)}^p.$$

Thus using Hólder's inequality and $(p)^{-1} + (p')^{-1} = 1$ we have

$$\begin{aligned}\big|\big|\,f^\lambda\,\big|\big|_{L^1(X)} &= \int_{|f|>\lambda} |f(x)|\,dx \leq \big|\big|\,f\,\big|\big|_{L^p(x)}\,\mu\Big(\Big\{x\in X\,\Big|\,|f(x)| > \lambda\Big\}\Big)^{\frac{1}{p'}}\\ &\leq \big|\big|\,f\,\big|\big|_{L^p(x)}^{1+\frac{p}{p'}}\,\lambda^{-\frac{p}{p'}} = \big|\big|\,f\,\big|\big|_{L^p(x)}^p\,\lambda^{p-1}.\end{aligned}$$

Thus using part (1) of the theorem on the function f^λ, we have

$$\begin{aligned}\mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f^\lambda](x)|>\frac{\lambda}{2}\Big\}\Big) &\leq \frac{2C}{\lambda}\big|\big|\,f^\lambda\,\big|\big|_{L^1(X)}\\ &\leq 2C\,\lambda^{-1}\int_{|f|>\lambda}|f(x)|\,dx.\end{aligned}\tag{1.4}$$

Now using equations (1.2), (1.3), and (1.4), we have

$$\begin{aligned}\big|\big|\,\mathcal{M}[f]\,\big|\big|^p_{L^p(X)} &= p\int_0^\infty \lambda^{p-1}\,\mu\Big(\Big\{x\in X\,\Big|\,|\mathcal{M}[f](x)|>\lambda\Big\}\Big)\,d\lambda\\ &\leq 2C\,p\int_0^\infty \lambda^{p-2}\Big[\int_{|f|>\lambda}|f(x)|\,dx\Big]\,d\lambda\\ &= 2C\,p\int_X |f(x)|\Big[\int_0^{|f(x)|}\lambda^{p-2}\,d\lambda\Big]\,dx\\ &= 2C\Big(\frac{p}{p-1}\Big)\int_X |f(x)|^p\,dx.\end{aligned}$$

which completes the proof. □

COROLLARY 1.13. *Let $f\in L^1_{loc}(X,d\mu)$. Then for μ-almost al $x\in X$,*

$$\lim_{\delta\to 0}\frac{1}{\mu\big(B_\rho(x,\delta)\big)}\int_{B_\rho(x,\delta)} f(t)\,d\mu(t)=f(x),$$

The passage from the weak-type estimates of Theroem 1.12 to the differentiation theorem of Corollary 1.13 is standard. See, for example, [**Ste70**].

1.4. The Caldéron-Zygmund Decomposition.

Another classical tool which can be used in the context of space of homogeneous type is the Calderón-Zygmund decomposition of functions in L^1.

THEOREM 1.14 (Caldéron-Zygmund). *Let $f\in L^1(X,d\mu)$, and let $\alpha>0$ satisfy $\alpha\,\mu(X)>C\,\big|\big|\,f\,\big|\big|_{L^1}$ where C is the constant from the Hardy-Littlewood Maximal Theorem 1.12. Then there exists a sequence of balls $\big\{B_j=B_\rho(x_j;\delta_j)\big\}$ and a decomposition*

$$f=g+\sum_j b_j$$

with the following properties:

(1) The functions g and $\{b_j\}$ all belong to $L^1(X)$.

(2) $|g(x)|\leq A_2^2\,\alpha$ for almost all $x\in X$.

(3) The function b_j is supported in B_j. Moreover

$$\int_X |b_j(x)|\,d\mu(x)\leq 2\,A_2^2\,\alpha\,\mu(B_j),\qquad \textit{and}$$

$$\int_X b_j(x)\,d\mu(x)=0.$$

(4) $\sum_j \mu\big(B_j\big)\leq C\,\alpha^{-1}\,\big|\big|\,f\,\big|\big|_{L^1}$, where C depends only on A_1 and A_2.

Proof. Let $E_\alpha = \Big\{x \in X \,\Big|\, \mathcal{M}[f](x) > \alpha\Big\}$. Then E_α is an open set. Using the hypothesis on α and Theorem 1.12, we have

$$\mu\big(E_\alpha\big) \leq C\,\alpha^{-1}\,\big|\big|\,f\,\big|\big|_{L^1} < \mu(X).$$

Thus $E_\alpha \subsetneqq X$.

Since $E_\alpha \subsetneqq X$, Lemma 1.10 gives us a collection of balls $\{B_j = B_\rho(x_j, \delta_j)\}$ and $\{B_j^* = B_\rho(x_j, 4\delta_j)\}$ such that

(i) For all j, $B_j \subset E_\alpha$ and $B_j^* \cap (X - E_\alpha) \neq \emptyset$.
(ii) $E_\alpha = \bigcup_j B_j$.
(iii) $\sum_j \mu(B_j) \leq C\,\mu(E_\alpha)$, where C depends only on A_1 and A_2.
(iv) Each point of E_α belongs to at most M of the balls M_j.

Now Theorem 1.12 implies that $\mu(E_\alpha) < C\,\alpha^{-1}\,\big|\big|\,f\,\big|\big|_{L^1}$. Hence (iii) gives

$$\sum_j \mu(B_j) \leq \frac{C}{\alpha}\,\big|\big|\,f\,\big|\big|_{L^1},$$

which is assertion (4).

Let χ_j be the characteristic function of B_j. Since each point x belongs to at most M of the balls B_j, it follows that $1 \leq \sum_k \chi_k(x) \leq M$ for all $x \in E_\alpha$. Put

$$\begin{aligned} b_j(x) &= \chi_j(x)\,\Big[\sum_k \chi_k(x)\Big]^{-1} f(x) - \frac{\chi_j(x)}{\mu(B_j)} \int_{B_j} \chi_j(t)\,\Big[\sum_k (t)\Big]^{-1} f(t)\,d\mu(t) \\ &= b_{j,1}(x) - b_{j,2}(x). \end{aligned}$$

Then b_j is supported on B_j. Since b_j is a function on B_j minus its average, it is clear that

$$\int_{B_j} b_j(t)\,d\mu(t) = 0.$$

Next, let $y_j \in B_j^* \cap (X - E_\alpha)$. Then since $B_j \subset B_j^*$, $\mu(B_j^*) \leq A_2^2\,\mu(B_j)$, and $y \in B_j^*$ we have

$$\begin{aligned} \frac{1}{\mu(B_j)} \int_{B_j} \frac{\chi_j(t)}{\sum_k \chi_k(t)} \big|f(t)\big|\,d\mu(t) &\leq \left[\frac{\mu(B_j^*)}{\mu(B_j)}\right] \frac{1}{\mu(B_j)} \int_{B_j^*} \big|f(t)\big|\,d\mu(t) \\ &\leq A_2^2\,\mathcal{M}[f](y_j) \\ &\leq A_2^2\,\alpha. \end{aligned}$$

It follows that

$$\frac{1}{\mu(B_j)} \int_{B_j} |b_{j,l}(t)|\,d\mu(t) \leq A_2^2\,\alpha \qquad \text{for } l = 1,\,2.$$

Hence

$$\frac{1}{\mu(B_j)} \int_{B_j} |b_j(t)|\,d\mu(t) \leq 2A_2^2\,\alpha.$$

This shows that $b_j \in L^1(X, d\mu)$, and also establishes (3).

Now set

$$g(x) = \begin{cases} f(x) & \text{if } x \notin E_\alpha; \\ \\ \frac{\chi_j(x)}{\mu\left(B_j\right)} \int_{B_j} \chi_j(t) \left[\sum_k(t)\right]^{-1} f(t)\, d\mu(t) & \text{if } x \in E_\alpha. \end{cases}$$

Then $f = g + \sum_j b_j$ If $x \notin E_\alpha$, then $\mathcal{M}[f](x) \leq \alpha$. It follows from the differentiation theorem (Corollary 1.13), that $|f(x)| \leq \alpha$ for μ-almost all $xinX - E_\alpha$, and so the same is true for $|g(x)|$. On the other hand, we have already observed that

$$\frac{1}{\mu(B_j)} \int_{B_j} \frac{\chi_j(t)}{\sum_k \chi_k(t)} \big|f(t)\big|\, d\mu(t) \leq A_2^2\, \alpha,$$

so $|g(x)| \leq A_2^2\, \alpha$ on E_α. This shows that assertion (2) is true. Since $\mu(E_\alpha) \leq C\, \alpha^{-1} \big|\big| f \big|\big|_{L^1}$, it follows that $g \in L^1(X, d\mu)$, proving (1). This completes the proof of the theorem if $E_\alpha \neq X$.

□

COROLLARY 1.15. *With the notation of Theorem 1.14, the function g belongs to $L^2(X, d\mu)$, and*

$$\big|\big| g \big|\big|_{L^2}^2 \leq C\, \alpha \big|\big| f \big|\big|_{L^1}$$

where C is a constant depending only on A_1, A_2 and the constant C_1 from the Hardy-Littlewood Maximal Theorem.

PROOF. We have

$$\begin{aligned} \big|\big| g \big|\big|_{L^2}^2 &= \int_{E_\alpha} |g(t)|^2\, d\mu(t) + \int_{X - E_\alpha} |g(t)|^2\, d\mu(t) \\ &\leq A_2^4 \alpha^2 \mu(E_\alpha) + A_2^2\, \alpha \int_X |f(t)|\, dt \\ &\leq \left[A_2^4\, C_1 + A_2^2\right] \alpha \big|\big| f \big|\big|_{L^1}. \end{aligned}$$

□

2. The Laplace operator and Euclidean Analysis

We now turn to our first example, the Laplace operator on $\mathbb{R}^n$, which is given by

$$\triangle[u] = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}.$$

The corresponding inhomogeneous equation, sometimes known as the *Poisson equation*, is $\triangle[u] = g$. This equation plays a fundamental role in mathematical physics, and is perhaps the simplest model of an second order elliptic partial differential equation.

In section 2.1, we discuss the symmetries of the opertaor $\triangle$. In section 2.2 we show that the operator $\triangle$ has a fundamental solution given by convolution with the Newtonian potential N. In section 2.3, we discuss the connection between this fundamental solution and the ordinary Euclidean metric on $\mathbb{R}^n$. In particular, we focus on differential inequalities and cancellation conditions satisfied by N that can be expressed in terms of this metric. Then in sections 2.4, 2.5, 2.6, and 2.7, we show how some of the basic regularity properties of the Poisson equation follow from these properties of N.

2.1. Symmetries of the Laplace operator.

We begin by noting that the Laplace operator $\triangle$ has invariance properties with respect to three families of motions of $\mathbb{R}^n$.

PROPOSITION 2.1.

(a) *The operator $\triangle$ is invariant under translations. Thus for $y \in \mathbb{R}^n$ and $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we define $T_y[\varphi](x) = \varphi(x-y)$. Then*
$$\triangle\big[T_y[\varphi]\big] = T_y\big[\triangle[\varphi]\big].$$

(b) *The operator $\triangle$ is invariant under rotations. Let $O : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal linear transformation, and put $R_O[\varphi](x) = \varphi(Ox)$. Then*
$$\triangle\big[R_O[\varphi]\big] = R_O\big[\triangle[\varphi]\big].$$

(c) *Define the standard Euclidean dilations by setting $D_\lambda[\varphi](x) = \varphi(\lambda^{-1}x)$. Then*
$$\triangle\big[D_\lambda[\varphi]\big] = \lambda^{-2} D_\lambda\big[\triangle[\varphi]\big].$$

PROOF. Assertion (a) follows since $\triangle$ has constant coefficients. Assertion (c) is a simple application of the chain rule. We use the Fourier transform to establish (b). Suppose that $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. Recall that the Fourier transform $\mathcal{F}[\varphi]$ was defined in equation (1.10). Integration by parts shows that
$$\begin{aligned}\mathcal{F}\big[\triangle[\varphi]\big](\xi) &= \int_{\mathbb{R}^n} e^{-2\pi i \xi\cdot x}\, \triangle[\varphi](x)\, dx \\ &= -4\pi^2\,|\xi|^2 \int_{\mathbb{R}^n} e^{-2\pi i \xi\cdot x}\varphi(x)\, dx \\ &= -4\pi^2 |\xi|^2\, \mathcal{F}[\varphi](\xi).\end{aligned}$$
Now suppose that O is an orthogonal transformation, so that $O^{-1} = O^*$ where $Ox\cdot y = x \cdot O^* y$. Then
$$\mathcal{F}\big[R_O[\varphi]\big](\xi) = \int_{\mathbb{R}^n} e^{-2\pi i \xi\cdot O^* x}\varphi(x)\, dx = \int_{\mathbb{R}^n} e^{-2\pi i O\xi\cdot x}\varphi(x)\, dx = R_O\big[\mathcal{F}[\varphi]\big](\xi).$$
Thus
$$\begin{aligned}\mathcal{F}\big[\triangle\big[R_0[\varphi]\big]\big](\xi) &= -4\pi^2|\xi|^2\, R_O\big[\mathcal{F}[\varphi]\big](\xi) \\ &= -4\pi^2 |O\xi|^2 \mathcal{F}[\varphi](O\xi) \\ &= R_O\big[\mathcal{F}\big[\triangle[\varphi]\big]\big](\xi) \\ &= \mathcal{F}\big[R_O\big[\triangle[\varphi]\big]\big](\xi).\end{aligned}$$
Since the Fourier transform $\mathcal{F}$ is one-to-one, it follows that $\triangle\, R_0 = R_O\, \triangle$. □

For comparison with later examples, let us note that the translation invariance and homogeneity of the Laplace operator can be expressed in the following way. Define a diffeomorphism from the unit ball in $\mathbb{R}^n$ to the Euclidean ball centered at a point $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ of radius δ by setting $\Theta_{x,\delta}(u) = \delta(x+u)$. Then we can define the 'push-forward' $\widetilde{\triangle}$ of the operator $\triangle$ under the mapping $\Theta_{x,\delta}$ by setting

$$\widetilde{\triangle}[\varphi] = \triangle\big[\varphi \circ \Theta_{x,\delta}^{-1}\big]$$

for $\varphi \in \mathcal{C}_0^\infty\big(\mathbb{B}_E(x,\delta)\big)$. It follows from (a) and (c) of Proposition 2.1 that

$$\widetilde{\triangle}[\varphi] = \delta^{-2}\,\triangle[\varphi].$$

2.2. The Newtonian Potential.

Given the invariance properties of $\triangle$, it is natural to look for a fundamental solution $\mathcal{N} : \mathcal{D}(\mathbb{R}^n) \to \mathcal{D}'(\mathbb{R}^n)$ for $\triangle$ which enjoys the same three kinds of invariance. If the Schwartz kernel for $\mathcal{N}$ is locally integrable and $\mathcal{N}$ is invariant under translation, we must have

$$\mathcal{N}[\varphi](x) = \int_{\mathbb{R}^n} N(x-y)\,\varphi(y)\,dy.$$

If $D_\lambda \mathcal{N} = \lambda^{-2}\mathcal{N}\,D_\lambda$, we must have

$$N(\lambda x) = \lambda^{2-n}\,N(x).$$

And if $\mathcal{N}$ is invariant under rotations, it would follow that $N(x)$ depends only on $|x|$, and so $N(x) = c_n|x|^{2-n}$ for some constant c. At least when $n > 2$, we see in Lemma 2.3 that this heuristic argument is correct.

DEFINITION 2.2. *The Newtonian potential N is the function given by*

$$N(x) = \begin{cases} \omega_2^{-1}\,\log(|x|) & \text{when } n = 2, \\ \\ \omega_n^{-1}\,(2-n)^{-1}\,|x|^{2-n} & \text{when } n > 2. \end{cases}$$

Here

$$\omega_n = 2\pi^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)^{-1}$$

is the surface measure of the unit sphere in $\mathbb{R}^n$.

The function N is locally in $L^p(\mathbb{R}^n)$ provided that $p < \frac{n}{n-2}$, and in particluar, N is always locally integrable.

LEMMA 2.3. *Convolution with N is a fundamental solution for $\triangle$. Precisely, if $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, define*

$$\mathcal{N}[\varphi](x) = \int_{\mathbb{R}^n} N(x-y)\,\varphi(y)\,dy = \int_{\mathbb{R}^n} N(y)\,\varphi(x-y)\,dy.$$

Then $\mathcal{N}[\varphi] \in \mathcal{C}^\infty(\mathbb{R}^n)$, and

$$\varphi(x) = \mathcal{N}\big[\triangle[\varphi]\big](x),$$
$$\varphi(x) = \triangle\Big[\mathcal{N}[\varphi]\Big](x).$$

PROOF. Let $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. Choose $R > 0$ so large that the support of φ is contained in the open Euclidean ball centered at the origin of radius R. Let $B(\epsilon, R) = \{x \in \mathbb{R}^n \,\big|\, \epsilon < |x| < R\}$. We can use Green's theorem to obain

$$\begin{aligned}\int_{B(\epsilon,R)} N(x)\,\triangle[\varphi](x) - \varphi(x)\,\triangle[N](x)\,dx \\ = \int_{\partial B(\epsilon,R)} \Big[N(\zeta)\frac{\partial\varphi}{\partial n}(\zeta) - \varphi(\zeta)\,\frac{\partial N}{\partial n}(\zeta)\Big]\,d\sigma(\zeta)\end{aligned} \tag{2.1}$$

where $\frac{\partial}{\partial n}$ denotes the outward unit normal derivative on the boundary $\partial B(\epsilon, R)$. The function N is infinitely differentiable away from the origin, and a direct calculation shows that $\triangle[N](x) = 0$ for $x \neq 0$. Thus the left hand side of equation (2.1) reduces to

$$\int_{B(\epsilon,R)} N(x)\,\triangle[\varphi](x)\,dx = \int_{|x|>\epsilon} N(x)\,\triangle[\varphi](x)\,dx.$$

The boundary of $B(\epsilon, R)$ has two connected components: the set S_R where $|x| = R$ and the set S_ϵ where $|x| = \epsilon$. The function φ is identically zero in a neighborhood of S_R, so this part of the boundary gives no contribution. Thus, taking account of orientation, the right hand side of equation (2.1) reduces to

$$-\int_{|\zeta|=\epsilon} \Big[N(\zeta)\frac{\partial\varphi}{\partial n}(\zeta) - \varphi(\zeta)\,\frac{\partial N}{\partial n}(\zeta)\Big]\,d\sigma(\zeta).$$

For $|\zeta| = \epsilon$ we have

$$N(\zeta) = \begin{cases} \frac{1}{2\pi}\log(\epsilon) & \text{if } n = 2, \\ \\ \omega_n^{-1}\,(2-n)^{-1}\epsilon^{2-n} & \text{if } n > 2. \end{cases} \qquad \text{and} \qquad \frac{\partial N}{\partial n}(\zeta) = \omega_n^{-1}\,\epsilon^{1-n}.$$

Thus Green's theorem gives

$$\begin{aligned}\int_{|x|\geq\epsilon} N(x)\,\triangle[\varphi](x)\,dx = \varphi(0) + \omega_n^{-1}\epsilon^{1-n}\int_{|\zeta|=\epsilon} \big[\varphi(\zeta) - \varphi(0)\big]\,d\sigma(\zeta) \\ - \begin{cases} \omega_2^{-1}\,\log(\epsilon)\int_{S_\epsilon}\frac{\partial\varphi}{\partial n}(\zeta)\,d\sigma(\zeta) & \text{if } n = 2 \\ \\ \omega_n^{-1}\,\epsilon^{2-n}\int_{S_\epsilon}\frac{\partial\varphi}{\partial n}(\zeta)\,d\sigma(\zeta) & \text{if } n > 2 \end{cases}.\end{aligned}$$

We now let $\epsilon \to 0$. The Lebesgue dominated convergence theorem shows that

$$\lim_{\epsilon\to 0}\int_{|x|\geq\epsilon} N(x)\,\triangle[\varphi](x)\,dx = \int_{\mathbb{R}^n} N(x)\,\triangle[\varphi](x)\,dx.$$

Also

$$\Big|\omega_n^{-1}\epsilon^{1-n}\int_{S_\epsilon}\Big[\varphi(\zeta) - \varphi(0)\Big]\,d\sigma(\zeta)\Big| \leq \epsilon\,\sup_{x\in\mathbb{R}^n}\big|\nabla\varphi(x)\big| \to 0;$$

$$\Big|\omega_2^{-1}\,\log(\epsilon)\int_{S_\epsilon}\frac{\partial\varphi}{\partial n}(\zeta)\,d\zeta\Big| \leq \epsilon\,\log(\epsilon)\,\sup_{x\in\mathbb{R}^2}\big|\nabla\varphi(x)\big| \to 0;$$

$$\Big|\omega_n^{-1}\,(2-n)^{-1}\,\epsilon^{2-n}\int_{S_\epsilon}\frac{\partial\varphi}{\partial n}(\zeta)\,d\zeta\Big| \leq \epsilon\,(2-n)^{-1}\,\sup_{x\in\mathbb{R}^n}\big|\nabla\varphi(x)\big| \to 0.$$

Thus we have shown that $\varphi(0) = \int_{\mathbb{R}^n} N(y)\,\triangle[\varphi](y)\,dy = \mathcal{N}\big[\triangle[\varphi]\big](0)$. But now we can use the translation invariance of $\triangle$. If $\varphi \in \mathcal{C}_0^\infty(\Omega)$, we have

$$\int_{\mathbb{R}^n} N(x-y)\,\triangle[\varphi](y)\,dy = \int_{\mathbb{R}^n} N(y)\,\triangle[\varphi](y+x)\,dy$$
$$= \int_{\mathbb{R}^n} N(y)\,T_{-x}\big[\triangle[\varphi]\big](y)\,dy$$
$$\& = \int_{\mathbb{R}^n} N(y)\,\triangle\big[[T_{-x}[\varphi]\big](y)\,dy$$
$$= T_{-x}[\varphi](0)$$
$$= \varphi(x),$$

so $\varphi(x) = \mathcal{N}\big[\triangle[\varphi]\big](x)$.

Finally, since $\mathcal{N}[\varphi](x) = \int_{\mathbb{R}^n} N(y)\,\varphi(x-y)\,dy$, the local integrability of N shows that we can differentiate under the integral sign, and it follows that $\mathcal{N}[\varphi]$ is infinitely differentiable. In particular

$$\triangle\Big[\mathcal{N}[\varphi]\Big](x) = \int_{\mathbb{R}^n} N(y)\,\triangle[\varphi](x-y)\,dy = \varphi(x),$$

which completes the proof. □

2.3. The role of Euclidean geometry.

An obvious example of a space of homogeneous type is the set $\mathbb{R}^n$ with the standard Euclidean metric

$$d_E(x,y) = |x-y| = \Big(\sum_{j=1}^n (x_j - y_j)^2\Big)^{\frac{1}{2}}$$

and the measure given by Lebesgue measure. Then at least when $n \geq 3$, there is an explicit connection between the Newtonian potential N and d_E. Let

$$\mathbb{B}_E(x,\delta) = \Big\{y \in \mathbb{R}^n \,\Big|\, d_E(x,y) < \delta\Big\}$$

denote the standard Euclidean ball centered at x with radius δ. The volume of this ball is then

$$\big|\mathbb{B}_E(x,\delta)\big| = \frac{\omega_n}{n}\,\delta^n = \pi^{\frac{n}{2}}\,\Gamma\left(\frac{n+2}{2}\right)^{-1}\delta^n.$$

Then for $n \geq 3$,

$$N(x,y) = \left(\frac{n}{n-2}\right)\frac{d_E(x,y)^2}{\big|\mathbb{B}_E\big(x, d_E(x,y)\big)\big|}.$$

Moreover, we can formulate the essential estimates for derivatives of N rather simply in terms of the metric d_E.

LEMMA 2.4. *Suppose $n \geq 3$. Then for all multi-indices α and β with $|\alpha|+|\beta| \geq 0$ there is a constant $C_{\alpha,\beta} > 0$ so that for all $x, y \in \mathbb{R}^n$*

$$\big|\partial_x^\alpha\,\partial_y^\beta\,N(x,y)\big| \leq C_{\alpha,\beta}\,\frac{d_E(x,y)^{2-|\alpha|-|\beta|}}{\big|\mathbb{B}_E\big(x,d_E(x,y)\big)\big|} = C'_{\alpha,\beta}\,|x-y|^{-n+2-|\alpha|-|\beta|}.$$

The same inequality holds when $n = 2$ provided that $|\alpha| + |\beta| > 0$.

PROOF. The function N is homogeneous of degree $-n+2$, and hence $\partial_x^\alpha N$ is homogeneous of degree $-n-\alpha+2$. It follows that the function $x \to |x|^{n+\alpha-2}\,\partial_x^\alpha N(x)$ is homogeneous of degree zero. It is also continuous on the unit sphere, and hence bounded there, and this gives the required estimate. □

A great deal is known about the operator $\mathcal{N}$. In general, the passage from f to $\mathcal{N}[f]$ increases regularity. Roughly speaking the smoothness of $\mathcal{N}[f]$ is improved by two orders when measured in appropriate norms. Moreover, the operator $\mathcal{N}$ is a paradigm for the properties of an appropriate parametrix of any elliptic operator. There are many good references for a detailed discussion of these matters[1], and it is not our objective to give an exhaustive account of this material. Rather, we want to indicate that many of these results do not depend on the explicit formula for N, and are true for any operator whose Schwartz kernel satisfies differential inequalities and cancellation conditions that can be formulated in terms of the Euclidean metric.

The appropriate size conditions are already suggested by Lemma 2.4. We shall consider an operator $\mathcal{K}$ given by

$$\mathcal{K}[f](x) = \int_{\mathbb{R}^n} K(x,y)\,f(y)\,dy, \tag{2.2}$$

and we shall impose the hypothesis that K is smooth away from the diagonal in $\mathbb{R}^n \times \mathbb{R}^n$ and that there are constants $C_{\alpha,\beta}$ so that for $x \neq y$ we have

$$\left|\partial_x^\alpha \partial_y^\beta K(x,y)\right| \leq C_{\alpha,\beta}\, d_E(x,y)^{m-|\alpha|-|\beta|}\left|\mathbb{B}\big(x; d_E(x,y)\big)\right|^{-1}. \tag{2.3}$$

Notice that $N(x,y) = N(x-y)$ satisfies these estimates with $m = 2$.

Size estimates alone are not sufficient to establish the regularity results we have in mind. If we want to show that $\mathcal{N}[f]$ is two orders smoother than f, it is natural to look at a second partial derivatives of $\mathcal{N}[f]$, and try to show that this has the same regularity as f. To understand whether or not this is true, it is tempting to differentiate formally under the integral sign to obtain

$$\frac{\partial^2\left[\mathcal{N}[f]\right]}{\partial x_j \partial x_k}(x) = \int_{\mathbb{R}^n} \frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\,f(y)\,dy.$$

However the function $\partial^2_{j,k}N = \dfrac{\partial^2 N}{\partial x_j \partial x_k}$ is not locally integrable, so the integral is not absolutely convergent, even if $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$.

Note that $\partial^2_{j,k}N$ is a kernel satisfying the hypotheses (2.3) with $m = 0$, and $\partial^2_{j,k}\mathcal{N}$ is a typical example of a *singular integral operator*. The regularity of such operators depends also on certain cancellation conditions on the kernel. In the case of the Newtonian potential, we have the following.

PROPOSITION 2.5. *For any* $R > 0$ *and all* $1 \leq j,k \leq n$ *we have*

$$\int_{|\zeta|=R} \frac{\partial N}{\partial x_j}(\zeta)\,d\sigma(\zeta) = 0 = \int_{|\zeta|=R} \frac{\partial^2 N}{\partial x_j \partial x_k}(\zeta)\,d\sigma(\zeta).$$

[1]See, for example, books on elliptic partial differential equations such as [**BJS64**], [**GT83**], or books on pseudo-differential operators such as [**Tre80**].

PROOF. We calculate

$$\frac{\partial N}{\partial x_j}(x) = \partial_j N(x) = \omega_n^{-1}\, x_j\, |x|^{-n}$$

and

$$\frac{\partial^2 N}{\partial x_j \partial x_k}(x) = \partial^2_{j,k} N(x) = \begin{cases} -n\,\omega_n^{-1}\, x_j x_k |x|^{-n-2} & \text{if } j \neq k, \\ \omega_n^{-1}\left[x_1^2 + \cdots + x_n^2 - n x_j^2\right] |x|^{-n-2} & \text{if } j = k. \end{cases}$$

Thus each function $\partial_j N(x)$ is odd, and this gives the first equality. If $j \neq k$, the function $\partial^2_{j,k} N$ is odd in the variables x_j and x_k separately, which gives the second equality in this case. Finally, observe that by symmetry

$$\int_{|\zeta|=R} \zeta_j^2 \, d\sigma(\zeta) = \int_{|\zeta|=R} \zeta_k^2 \, d\sigma(\zeta) = \frac{\omega_n}{n} R^{n+1},$$

and this proves the second equality when $j = k$. □

This suggests that in addition to the differential inequalities (2.3), we also require that when $m = 0$, for all $0 < R_1 < R_2$ we have

$$\int_{R_1 < |x-y| < R_2} K(x,y)\, dx = \int_{R_1 < |x-y| < R_2} K(x,y)\, dy = 0 \tag{2.4}$$

To illustrate the utility of the estimates (2.3) and cancellation conditions (2.4), in the next four sections we discuss four kinds of classical regularity results.

(I) The Hardy-Littlewood Sobolev theorem on fractional integration. Since this deals with operators for which the order of smoothing $m > 0$, cancellation conditions are not needed.

(II) Lipschitz estimates for $\partial^2_{j,k}\big[\mathcal{N}[f]\big]$. This involves a singular integral operator, and we need the cancellation hypothesis.

(III) L^2-estimates for solutions $\partial^2_{j,k}\big[\mathcal{N}[f]\big]$. Instead of using the Fourier transform, we shall see that the cancellation condition can be used to give an 'almost orthogonal' decomposition of the operator.

(IV) L^1-estimates for solutions $\partial^2_{j,k}\big[\mathcal{N}[f]\big]$. Here we will combine the L^2 estimates and the Caldéron-Zygmund decomposition of L^1-functions established in Theorem 1.14.

2.4. Hardy-Littlewood-Sobolev estimates.

The operator $\mathcal{N}$ improves the integrability of functions. More precisely, suppose $1 < p, q < \infty$ with

$$\frac{1}{q} = \frac{1}{p} - \frac{2}{n}.$$

(Note that in particular this means that $q > p$). Then if $f \in L^p(\mathbb{R}^n)$ we will show that $\mathcal{N}[f] \in L^q(\mathbb{R}^n)$, and there is a constant $C_{p,q}$ independent of f so that

$$\big|\big|\, \mathcal{N}[f]\, \big|\big|_{L^q(\mathbb{R}^n)} \leq C_{p,q}\, \big|\big|\, f\, \big|\big|_{L^p(\mathbb{R}^n)}.$$

In fact, we shall show the following more general result. Let $0 < m < n$, and let $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ be a measurable function such that

$$|K(x,y)| \leq C_0 \frac{d_E(x,y)^m}{\big|\mathbb{B}_E(x; d(x,y))\big|} = C_1\,|x-y|^{-n+m}$$

for some constants C_0 and C_1. Define an operator $\mathcal{K}$ by setting

$$\mathcal{K}[f](x) = \int_{\mathbb{R}^n} K(x,y)\,f(y)\,dy$$

whenever the integral is convergent.

Theorem 2.6 (Hardy-Littlewood-Sobolev). *Let $1 \leq p < \frac{n}{m}$. If $f \in L^p(\mathbb{R}^n)$ then the integral defining $\mathcal{K}$ converges for almost every $x \in \mathbb{R}^n$. Moreover, if $p > 1$ and if*

$$\frac{1}{q} = \frac{1}{p} - \frac{m}{n} > 0,$$

there is a constant $C_{p,m}$ so that for every $f \in L^p(\mathbb{R}^n)$ we have

$$\big|\big|\,\mathcal{K}[f]\,\big|\big|_{L^q(\mathbb{R}^n)} \leq C_{p,m}\,\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}.$$

Proof. Let $x \in \mathbb{R}^n$ and $\lambda > 0$. We have

$$\begin{aligned}
|\mathcal{K}[f](x)| &\leq C_1 \int_{\mathbb{R}^n} |x-y|^{-n+m}\,|f(y)|\,dy = C_1 \int_{\mathbb{R}^n} |y|^{-n+m}\,|f(x-y)|\,dy \\
&= C_1 \int_{|y|\leq\lambda} |y|^{-n+m}\,|f(x-y)|\,dy + C_1 \int_{|y|>\lambda} |y|^{-n+m}\,|f(x-y)|\,dy \\
&= I_\lambda(x) + II_\lambda(x).
\end{aligned}$$

Let $R_{j,\lambda} = \big\{y \in \mathbb{R}^n \,\big|\, 2^{-(j+1)}\lambda < |y| \leq 2^{-j}\lambda\big\}$. Then

$$\begin{aligned}
I_\lambda(x) &= C_1 \sum_{j=0}^{\infty} \int_{R_{j,\lambda}} |y|^{-n+m}\,|f(x-y)|\,dy \\
&\leq C_0 \sum_{j=0}^{\infty} (2^{-j-1}\lambda)^{-n+m} \int_{|y|\leq 2^{-j}\lambda} |f(x-y)|\,dy \\
&= 2^{n-m}\,C_0\,n^{-1}\,\omega_n\,\lambda^m \sum_{j=0}^{\infty} 2^{-jm} \frac{1}{\big|\mathbb{B}(x;2^{-j}\lambda)\big|} \int_{\mathbb{B}(x;2^{-j}\lambda)} |f(y)|\,dy \\
&\leq \left[\frac{2^{n-m}\,C_0\,\omega_n}{n(1-2^{-m})}\right] \lambda^m\,\mathcal{M}[f](x)
\end{aligned}$$

where $\mathcal{M}[f]$ is the Hardy-Littlewood maximal operator applied to f and $n^{-1}\omega_n$ is the volume of the unit Euclidean ball. On the other hand, using Hölder's inequality, we have

$$\begin{aligned}
II(x) &\leq C_1\,\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)} \left[\int_{|y|>\lambda} |y|^{(-n+m)p'}\,dy\right]^{\frac{1}{p'}} \\
&= C_1 \left(\frac{\omega_n}{(m-n)p'+n}\right)^{\frac{1}{p'}} \big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}\,\lambda^{m-\frac{n}{p}},
\end{aligned}$$

where $(p)^{-1} + (p')^{-1} = 1$. We have used the fact that $\frac{1}{p} - \frac{m}{n} > 0$.

Thus we see that there is a constant C depending only on n, p, and m so that

$$|\mathcal{K}[f](x)| \leq C\left[\lambda^m\,\mathcal{M}[f](x) + \lambda^{m-\frac{n}{p}}\,\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}\right].$$

Since $\mathcal{M}[f](x) < \infty$ for almost all $x \in \mathbb{R}^n$, it follows that the integral defining $\mathcal{K}[f](x)$ converges absolutely for almost all x. Now let

$$\lambda = \left(\frac{\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}}{\mathcal{M}[f](x)}\right)^{\frac{p}{n}}.$$

Then it follows that

$$|\mathcal{K}[f](x)| \leq C\,\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}^{\frac{mp}{n}}\,\mathcal{M}[f](x)^{1-\frac{mp}{n}} = C\,\big|\big|\,f\,\big|\big|_{L^p(\mathbb{R}^n)}^{1-\frac{p}{q}}\,\mathcal{M}[f](x)^{\frac{p}{q}},$$

and so $\big|\big|\,\mathcal{K}[f]\,\big|\big|_{L^q} \leq C\,\big|\big|\,f\,\big|\big|_{L^p}$, which completes the proof. □

2.5. Lipschitz Estimates.

Next we show that the operator $\mathcal{N}$ increases differentiability by two orders when measured in an appropriate way. Thus for $0 < \alpha < 1$ we let $\Lambda_\alpha(\mathbb{R}^n)$ denote the space of complex-valued continuous functions f on $\mathbb{R}^n$ for which

$$\big|\big|\,f\,\big|\big|_{\Lambda_\alpha} = \sup_{x_1 \neq x_2 \in \mathbb{R}^n} \frac{\big|f(x_2) - f(x_1)\big|}{\big|x_2 - x_1\big|^\alpha} < \infty.$$

The quantity $\big|\big|\,\cdot\,\big|\big|_{\Lambda_\alpha}$ is not a norm, since $\big|\big|\,f\,\big|\big|_{\Lambda_\alpha} = 0$ if f is a constant function. We say that $f \in \Lambda_\alpha$ satisfies a *Lipschitz condition* of order α. Next, if m is a non-negative integer, then $\Lambda_\alpha^m(\mathbb{R}^n)$ is the space of m-times continuously differentiable complex-valued functions f on $\mathbb{R}^n$ such that every partial derivative of f of order m satisfies a Lipschitz condition of order α. We put

$$\big|\big|\,f\,\big|\big|_{\Lambda_\alpha^m} = \sum_{|\beta|=m} \big|\big|\,\partial_x^\beta f\,\big|\big|_{\Lambda_\alpha}.$$

Note that $\big|\big|\,p\,\big|\big|_{\Lambda_\alpha^m} = 0$ for every polynomial p of order less than or equal to m.

The increased smoothness of $\mathcal{N}[f]$ can be expressed by the fact that $\mathcal{N}[f]$ is $m+2$-times continuously differentiable, and every derivative of order $m+2$ satisfies a Lipschitz condition of order α. In fact, there is a constant C so that if $f \in \Lambda_\alpha^m$ and (say) has compact support, then

$$\big|\big|\,\mathcal{N}[f]\,\big|\big|_{\Lambda_\alpha^{m+2}} \leq C\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha^m}.$$

A key point here is that since $\mathcal{N}$ is a convolution operator, it commutes with differentiation. Thus if $f \in \Lambda_\alpha^m$ and if $|\beta| = m$ we have

$$\partial_x^\beta\Big[\mathcal{N}[f]\Big](x) = \mathcal{N}\big[\partial_x^\beta[f]\big](x),$$

and the function $g = \partial_x^\beta[f]$ satisfies a Lipschitz condition of order α. Thus the crux of the matter is the following result.

THEOREM 2.7. *Let $0 < \alpha < 1$, and let $f \in \Lambda_\alpha$ have compact support. Then $\mathcal{N}[f]$ is twice continuously differentiable, and there is a constant C independent of f and its support so that*

$$\sum_{j,k=1}^{n} \Big|\Big| \frac{\partial^2 \big[\mathcal{N}[f]\big]}{\partial x_j \partial x_k} \Big|\Big|_{\Lambda_\alpha} \leq C \,||\, f \,||_{\Lambda_\alpha}.$$

Before proving Theorem 2.7, we first establish the following preliminary result.

PROPOSITION 2.8. *Let f be continuous with compact support. Then $\mathcal{N}[f]$ is continuously differentiable and*

$$\frac{\partial\big[\mathcal{N}[f]\big]}{\partial x_j}(x) = \int_{\mathbb{R}^n} \frac{\partial N}{\partial x_j}(y)\, f(x-y)\, du.$$

PROOF. Choose $\chi \in \mathcal{C}^\infty(\mathbb{R})$ so that $0 \leq \chi(t) \leq 1$ for all t and

$$\chi(t) = \begin{cases} 0 & \text{if } t \leq 1, \\ 1 & \text{if } t \geq 2. \end{cases}$$

For $x \in \mathbb{R}^n$ and $\epsilon > 0$ put $\varphi_\epsilon(x) = \chi\big(\epsilon^{-1}|x|\big)$. Then φ_ϵ is supported where $|x| \geq \epsilon$, $\nabla\varphi_\epsilon$ is supported where $\epsilon \leq |x| \leq 2\epsilon$, and $|\nabla\varphi_\epsilon(x)| \leq C\epsilon^{-1} \leq 2C\,|x|^{-1}$. Put

$$\mathcal{N}_\epsilon[f](x) = \int_{\mathbb{R}^n} N(x-y)\,\varphi_\epsilon(x-y)\, f(y)\, dy.$$

Then

$$\begin{aligned}\big|\mathcal{N}[f](x) - \mathcal{N}_\epsilon[f](x)\big| &= \Big|\int_{\mathbb{R}^n} N(x-y)\,\big[1-\varphi_\epsilon(x-y)\big]\, f(y)\, dy\Big| \\ &\leq C_n \sup_{x\in\mathbb{R}^n} |f(x)| \int_{|y|<2\epsilon} |y|^{-n+2}\, dy \\ &\leq C_n \sup_{x\in\mathbb{R}^n} |f(x)|\,\epsilon^2.\end{aligned}$$

Thus $\mathcal{N}_\epsilon[f] \to \mathcal{N}[f]$ uniformly on $\mathbb{R}^n$ as $\epsilon \to 0$.

On the other hand, the function $x \to N(x-y)\,\varphi_\epsilon(x-y)$ is infinitely differentiable. Thus $\mathcal{N}_\epsilon[f]$ is infinitely differentiable, and

$$\frac{\partial\big[\mathcal{N}_\epsilon[f]\big]}{\partial x_j}(x) = \int_{\mathbb{R}^n} \frac{\partial[N\varphi_\epsilon]}{\partial x_j}(x-y)\, f(y)\, dy.$$

Put

$$\mathcal{N}_j[f](x) = \int_{\mathbb{R}^n} \frac{\partial N}{\partial x_j}(x-y)\, f(y)\, dy.$$

Then

$$\begin{aligned}\big|\mathcal{N}_j[f](x) - \partial_j\big[\mathcal{N}_\epsilon[f]\big](x)\big| &= \Big|\int_{\mathbb{R}^n} \frac{\partial}{\partial x_j}\big[N(1-\varphi_\epsilon)\big](x-y)\, f(y)\, dy\Big| \\ &\leq C_n \sup_{x\in\mathbb{R}^n} \int_{|x-y|<2\epsilon} \big(|x-y|^{-n+1} + \epsilon^{-1}|N(x-y)|\big)\, dy \\ &\leq \begin{cases} C_n \sup_{x\in\mathbb{R}^n} |f(x)|\,\epsilon & \text{if } n > 2, \\ \\ A_2 \sup_{x\in\mathbb{R}^2} |f(x)|\,\epsilon\big[1+\log\big(\epsilon^{-1}\big)\big] & \text{if } n = 2. \end{cases}\end{aligned}$$

Thus $\partial_j\big[\mathcal{N}_\epsilon[f]\big] \to \mathcal{N}_j[f]$ uniformly on $\mathbb{R}^n$ as $\epsilon \to 0$. Combined with our first observation, this shows that $\mathcal{N}[f]$ is differentiable and $\partial_j\mathcal{N}[f] = \mathcal{N}_j[f]$, completing the proof. □

In attempting to show that $\mathcal{N}[f]$ is twice continuously differentiable when $f \in \Lambda_\alpha$, we have already observed that we cannot simply differentiate under the integral sign since $\partial^2_{j,k}N$ is not locally integrable. However if $f \in \Lambda_\alpha$, we do have

$$\left|\frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\big[f(y)-f(x)\big]\right| \leq C_n \,\big|\big|\, f \,\big|\big|_{\Lambda_\alpha} |y-x|^{\alpha-n}$$

and so $\partial^2_{j,k}N(x-y)\,\big[f(y)-f(x)\big]$ is locally integrable as a function of y. To obtain a correct formula for $\partial^2_{j,k}[\mathcal{N}[f]](x)$, we need to make use of cancellation properties $\partial^2_{j,k}N$.

Lemma 2.9. *Let $f \in \Lambda_\alpha$ have compact support. Then $\mathcal{N}[f]$ is twice continuously differentiable, and for any $0 < R < \infty$ we have*

$$\begin{aligned}\frac{\partial^2\big[\mathcal{N}[f]\big]}{\partial x_j \partial x_k}(x) = \frac{\delta_{j,k}}{n}\, f(x) + \int_{|x-y|<R} \frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\big[f(y)-f(x)\big]\,dy \\ + \int_{|x-y|\geq R} \frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\, f(y)\,dy,\end{aligned} \tag{2.5}$$

where both integrals are absolutely convergent, and $\delta_{j,k} = \begin{cases} 1 & \text{if } j=k, \\ 0 & \text{if } j \neq k.\end{cases}$

Proof. With φ_ϵ defined as in Proposition 2.8, put

$$\begin{aligned}\mathcal{N}_j[f](x) &= \int_{\mathbb{R}^n} \frac{\partial N}{\partial x_j}(x-y)\, f(y)\,dy \\ \mathcal{N}_{j,\epsilon}[f](x) &= \int_{\mathbb{R}^n} \frac{\partial N}{\partial x_j}(x-y)\,\varphi_\epsilon(x-y)\, f(y)\,dy \\ \mathcal{N}_{j,k}[f](x) &= \frac{\delta_{j,k}}{n}\, f(x) + \int_{|x-y|<R} \frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\big[f(y)-f(x)\big]\,dy \\ &\qquad + \int_{|x-y|\geq R} \frac{\partial^2 N}{\partial x_j \partial x_k}(x-y)\, f(y)\,dy.\end{aligned}$$

Then since $(1-\varphi_\epsilon)$ is supported on the ball centered at the origin of radius 2ϵ, we have

$$\begin{aligned}\big|\mathcal{N}_j[f](x) - \mathcal{N}_{j,\epsilon}[f](x)\big| &= \Big|\int_{\mathbb{R}^n} \frac{\partial N}{\partial x_j}(x-y)\,(1-\varphi_\epsilon)(x-y)\, f(y)\,dy\Big| \\ &\leq \int_{|x-y|<2\epsilon} \Big|\frac{\partial N}{\partial x_j}(x-y)\Big|\,|f(y)|\,dy \\ &\leq C_n \sup_{x\in\mathbb{R}^n} |f(x)| \int_{|y|<2\epsilon} |y|^{-n+1}\,dy \\ &\leq C'_n \sup_{x\in\mathbb{R}^n} |f(x)|\,\epsilon.\end{aligned}$$

Thus $\mathcal{N}_{j,\epsilon}[f] \to \mathcal{N}_j[f]$ uniformly on $\mathbb{R}^n$ as $\epsilon \to 0$.

But now $\mathcal{N}_{j,\epsilon}[f]$ is infinitely differentiable, and

$$\begin{aligned}\frac{\partial\big[\mathcal{N}_{j,\epsilon}[f]\big]}{\partial x_k}(x) &= \int_{\mathbb{R}^n} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,f(y)\,dy \\ &= \int_{|x-y|<R} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,[f(y)-f(x)]\,dy \\ &\qquad + f(x)\int_{|x-y|<R} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,dy \\ &\qquad\qquad + \int_{|x-y|\geq R} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,f(y)\,dy\end{aligned}$$

Now $\varphi_\epsilon(x-y)\equiv 1$ when $|x-y|>2\epsilon$. Thus if $R>2\epsilon$ we have

$$\int_{|x-y|\geq R} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,f(y)\,dy = \int_{|x-y|\geq R} \frac{\partial^2 N}{\partial x_j\partial x_k}\,(x-y)\,f(y)\,dy.$$

Also, if $\epsilon<R$ we have, by the divergence theorem,

$$\begin{aligned}\int_{|x-y|<R} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,\varphi_\epsilon\Big](x-y)\,dy &= \int_{|\zeta|=R} \frac{\partial N}{\partial x_j}(\zeta)\,\varphi_\epsilon(\zeta)\,\zeta_k\,d\sigma(\zeta) \\ &= \int_{|\zeta|=R} \frac{\partial N}{\partial x_j}(\zeta)\,\zeta_k\,d\sigma(\zeta) \\ &= \begin{cases} 0 & \text{if } j\neq k, \\ \frac{1}{n} & \text{if } j=k.\end{cases}\end{aligned}$$

Thus

$$\begin{aligned}&\Big|\mathcal{N}_{j,k}[f](x) - \frac{\partial\big[\mathcal{N}_{j,\epsilon}[f]\big]}{\partial x_k}(x)\Big| \\ &\qquad = \Big|\int_{|x-y|\leq 2\epsilon} \frac{\partial}{\partial x_k}\Big[\frac{\partial N}{\partial x_j}\,(1-\varphi_\epsilon)\Big](x-y)\,[f(y)-f(x)]\,dy\Big| \\ &\qquad \leq C_n\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\,\epsilon^\alpha\end{aligned}$$

Hence $\partial_k\big[\mathcal{N}_{j,\epsilon}[f]\big]\to\mathcal{N}_{j,k}[f]$ uniformly on $\mathbb{R}^n$ as $\epsilon\to 0$. This shows that $\mathcal{N}[f]$ is twice continuously differentiable, and is given by equation (2.5). □

To prove Theorem 2.7, it suffices to show that if $f\in\Lambda_\alpha$ has compact support, and if

$$F(x) = \int_{|x-y|<1} \frac{\partial^2 N}{\partial x_j\partial x_k}(x-y)\,\big[f(y)-f(x)\big]\,dy + \int_{|x-y|<1} \frac{\partial^2 N}{\partial x_j\partial x_k}(x-y)\,f(y)\,dy$$

then $F\in\Lambda_\alpha$ and $\big|\big|\,F\,\big|\big|_{\Lambda_\alpha}\leq C\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}$ for some constant C independent of f and its support. In fact, we will show that this holds if the kernel $\partial^2_{j,k}N$ is replaced by any function having the same size and cancellation conditions.

Let K be smooth away from the diagonal in $\mathbb{R}^n\times\mathbb{R}^n$ and suppose

(1) For all $x\neq y$, $|K(x,y)|\leq C_0|x-y|^{-n}$.

(2) For all $x\neq y$, $|\nabla_x K(x,y)|+|\nabla_y K(x,y)|\leq C_1|x-y|^{-n-1}$.

(3) For all $0 < R_1 < R_2$ we have

$$\int_{R_1<|x-y|<R_2} K(x,y)\,dx = \int_{R_1<|x-y|<R_2} K(x,y)\,dy = 0.$$

If $f \in \Lambda_\alpha$ has compact support, we can define

$$\mathcal{K}[f](x) = \int_{|x-y|<1} K(x,y)\big[f(y)-f(x)\big]\,dy + \int_{|x-y|\geq 1} K(x,y)\,f(y)\,dy.$$

and both integrals converge absolutely.

THEOREM 2.10. *Let $0 < \alpha < 1$. There is a constant C_α depending only on C_0 and C_1 so that for all $f \in \Lambda_\alpha$ with compact support we have*

$$\sup_{x_1\neq x_2\in\mathbb{R}^n} \frac{|\mathcal{K}[f](x_2)-\mathcal{K}[f(x_1)|}{|x_2-x_1|^\alpha} \leq C_\alpha\,\big|\big|\,f\,\big|\big|_{\Lambda^\alpha}.$$

PROOF. Let $x_1 \neq x_2 \in \mathbb{R}^n$ and put $\delta = |x_2 - x_1|$. Since f has compact support, there exists $R > 3\delta$ (depending on f) so that $|x_j - y| \geq R$ implies $f(y) = 0$. Using the fact that the function $y \to K(x,y)$ has mean value zero on the set $1 \leq |x-y| \leq R$, it follows that

$$\mathcal{K}[f](x_j) = \int_{|x_j-y|<R} K(x_j,y)\,\big[f(y)-f(x_j)\big]\,dy.$$

Let $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ be a radial function such that $\psi(x) = 1$ if $|x| \leq R$ and $0 \leq \psi(x) \leq 1$ and $|\nabla\psi(x)| \leq |x|^{-1}$ for all $x \in \mathbb{R}^n$. Then we can write

$$\mathcal{K}[f](x_j) = \int_{R^n} K(x_j,y)\,\psi(x_j-y)\,\big[f(y)-f(x_j)\big]\,dy.$$

Then

$$\begin{aligned}
&\mathcal{K}[f](x_2)-\mathcal{K}[f](x_1)\\
&\qquad= \int_{\mathbb{R}^n} (K\,\psi)(x_2,y)\,\big(f(y)-f(x_2)\big) - (K\,\psi)(x_1,y)\,\big(f(y)-f(x_1)\big)\,dy\\
&\qquad= \int_{\mathbb{B}(x_2,2\delta)} K(x_2,y)\,\big(f(y)-f(x_2)\big) - K(x_1,y)\,\big(f(y)-f(x_1)\big)\,dy\\
&\qquad\quad+ \int_{\mathbb{B}(x_2,2\delta)^c} (K\,\psi)(x_2,y)\,\big(f(y)-f(x_2)\big) - (K\,\psi)(x_1,y)\,\big(f(y)-f(x_1)\big)\,dy\\
&\qquad= I + II.
\end{aligned}$$

Now $B(x_2, 2\delta) \subset B(x_1, 3\delta)$. Thus using the size estimates (1) for K we have

$$\begin{aligned}
\big|I\big| &\leq C_0\big|\big|\,f\,\big|\big|_{\Lambda_\alpha} \int_{|y-x_2|<2\delta} |y-x_2|^\alpha |K(x_2,y)|\,dy\\
&\qquad + C_0\big|\big|\,f\,\big|\big|_{\Lambda_\alpha} \int_{|y-x_1|<3\delta} |y-x_1|^\alpha |K(x_1,y)|\,dy\\
&\leq \frac{C_0\omega_n}{\alpha}\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\big[(2\delta)^\alpha+(3\delta)^\alpha\big] = C(n,\alpha)\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\,\delta^\alpha.
\end{aligned}$$

To deal with II, we rewrite the integrand as

$$\big(f(y)-f(x_1)\big)\big[(K\,\psi)(x_2,y)-(K\,\psi)(x_1,y)\big] + \big[f(x_1)-f(x_2)\big](K\,\psi)(x_2,y).$$

Since ψ is radial, we can use the cancellation condition (3) to conclude that

$$\int_{B(x_2 2\delta)^c} \big[f(x_1) - f(x_2)\big](K\,\psi)(x_2, y)\,dy = 0.$$

Thus

$$\big|II\big| \leq \big|\big|\,f\,\big|\big|_{\Lambda^\alpha} \int_{|y-x_2|>2\delta} |y - x_1|^\alpha |(K\,\psi)(x_2, y) - (K\,\psi)(x_1, y)|\,dy$$

If $|y - x_2| > 2\delta$, it follows that

$$|y - x_1| \leq |y - x_2| + |x_2 - x_1| = |y - x_2| + \delta < |y - x_2| + \frac{1}{2}|y - x_2|,$$

and so

$$|y - x_1| < \frac{3}{2}|y - x_1|.$$

Also, using the mean value theorem, it follows that

$$\begin{aligned} |(K\,\psi)(x_2, y) - (K\,\psi)(x_1, y)| &= |x_2 - x_1|\,|\nabla_x (K\,\psi)(\lambda x_2 + (1-\lambda)x_1, y)| \\ &\leq \frac{3}{2}\,(C_0 + C_1)\,\delta\,|y - x_2|^{-n-1}. \end{aligned}$$

Thus

$$\begin{aligned} \big|II\big| &\leq \left(\frac{3}{2}\right)^{1+\alpha} (C_0 + C_1)\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\,\delta \int_{|y-x_2|>2\delta} |y - x_2|^{-n-1+\alpha} \\ &= \left(\frac{3}{2}\right)^{1+\alpha} (C_0 + C_1)\,\frac{\omega_n}{1-\alpha}\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\,\delta^\alpha = C(n, \alpha)\,\big|\big|\,f\,\big|\big|_{\Lambda_\alpha}\,\delta^\alpha. \end{aligned}$$

This completes the proof. □

2.6. L^2-estimates.

We have seen that the operator which takes a function f to $\partial^2_{j,k}\big[\mathcal{N}[f]\big]$ can be written as an operator of the form

$$\mathcal{K}[\varphi](x) = \int_{|y|<R} K(y)\big[\varphi(x-y) - \varphi(x)\big]\,dy + \int_{|y|\geq R} K(y)\,\varphi(x-y)\,dy$$

where $K \in \mathcal{C}^1(\mathbb{R}^n - \{0\})$ and we assume that

(1) For all $x \neq 0$, $|K(x)| \leq C_0|x|^{-n}$.
(2) For all $x \neq 0$, $|\nabla K(x)| \leq C_1|x|^{-n-1}$.
(3) For all $0 < R_1 < R_2$ we have $\displaystyle\int_{R_1<|x|<R_2} K(x)\,dx = 0$.

We want to show that the operator $\mathcal{K}$, defined for $\varphi \in \mathcal{C}^\infty_0(\mathbb{R}^n)$, has a bounded to $L^2(\mathbb{R}^n)$. Since this operator is given by convolution with a distribution K, it is natural to use the Fourier transform and the Plancherel theorem to reduce the problem to showing that the Fourier transform of the distribution is uniformly bounded. This can certainly be done, but since we will not have the Fourier transform available in later examples, we prefer to use a method with wider application. This is based on an 'almost orthogonality' argument, and the key result is the following.

THEOREM 2.11 (Cotlar-Stein). *Let $\{T_j\}$, $j \in \mathbb{Z}$, be bounded operators on a Hilbert space $\mathcal{H}$. Assume there are constants C and $\epsilon > 0$ so that for all $j, k \in \mathbb{Z}$ we have*

(1) $\big|\big| T_j \big|\big| \leq C$.

(2) $\big|\big| T_j^* T_k \big|\big| \leq C\, 2^{-\epsilon|j-k|}$.

(3) $\big|\big| T_j T_k^* \big|\big| \leq C\, 2^{-\epsilon|j-k|}$.

There is a constant A so that for all N

$$\Big|\Big| \sum_{j=-N}^{N} T_j \Big|\Big| \leq A.$$

This is proved, for example, in [**Ste93**], pages 279-281, so we do not reproduce the argument here.

Now let $\chi \in \mathcal{C}_0^\infty(\mathbb{R})$ satisfy

(i) $0 \leq \chi(t) \leq 1$ for all $t \in \mathbb{R}$;

(ii) We have

$$\chi(t) = \begin{cases} 0 & \text{if } t \leq \frac{1}{8}, \\ 1 & \text{if } \frac{1}{4} \leq t \leq \frac{1}{2}, \\ 0 & \text{if } 1 \leq t. \end{cases}$$

Put $\chi_j(t) = \chi(2^{-j}t)$. Then each $t > 0$ is in the support of at most 4 of the functions $\{\chi_j\}$. Put

$$\Psi_j(t) = \big[\sum_k \chi_k(t)\big]^{-1} \chi_j(t).$$

Then Ψ_j is supported on $2^{j-3} \leq t \leq 2^j$, and

(iii) $\displaystyle\sum_{j=-\infty}^{\infty} \Psi_j(t) \equiv 1$ for all $t > 0$.

(iv) $\displaystyle\sum_{j=-N}^{N} \Psi_j$ is supported on $2^{-N-3} \leq t \leq 2^N$, and is identically 1 for $2^{-N} \leq t \leq 2^{N-3}$.

Now put

$$K_j(x) = \Psi_j(|x|)\, K(x).$$

Then it is not difficult to check that $\{K_j\}$, $j = 0, \pm 1, \pm 2, \ldots$ is a doubly infinite sequence of continuously differentiable functions on $\mathbb{R}^n$, and there is a constant C so that

(1) For all $j \in \mathbb{Z}$ and all $x \in \mathbb{R}^n$ we have $|K_j(x)| \leq C\, 2^{-nj}$.

(2) For all $j \in \mathbb{Z}$ and all $x \in \mathbb{R}^n$ we have $|\nabla K_j(x)| \leq C\, 2^{-(n+1)j}$.

(3) For all $j \in \mathbb{Z}$, K_j is supported in the ball $\mathbb{B}_E(0; 2^j)$.

(4) For all $j \in \mathbb{Z}$, $\displaystyle\int_{\mathbb{R}^n} K_j(x)\, dx = 0$.

Moreover

(5) $\displaystyle\sum_{j=-\infty}^{\infty} K_j(x) = K(x)$ for $x \neq 0$.

(6) $K^{[N]} = \sum_{j=-N}^{N} K_j$ is supported $2^{-N-3} \leq |x| \leq 2^N$, and is identically equal to K for $2^{-N} \leq |x| \leq 2^{N-3}$. Moreover, the kernel $K^{[N]}$ satisfies the same conditions (1), (2), and (3) as K with constants that are independent of N.

Set

$$\mathcal{K}_j[f](x) = K_j * \varphi(x) = \int_{\mathbb{R}^n} K_j(y)\, f(x-y)\, dy.$$

It follows that if $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, we have

$$\begin{aligned}\mathcal{K}[\varphi](x) &= \int_{|y|<R} K(y)\, \big[\varphi(x-y) - \varphi(x)\big]\, dy + \int_{|y|\geq 1} K(y)\, \varphi(x-y)\, dy \\ &= \lim_{N\to\infty} \int_{\mathbb{R}^n} \sum_{j=-N}^{N} K_j(y)\, \varphi(x-y)\, dy \\ &= \lim_{N\to\infty} \sum_{j=-N}^{N} \mathcal{K}_j[\varphi](x).\end{aligned}$$

Thus if we can show that the operators $\{\mathcal{K}_j\}$ satisfy the almost orthogonality conditions of Theorem 2.11, it follows from Fatou's lemma that for $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we have

$$\begin{aligned}\big|\big|\, \mathcal{K}[\varphi]\, \big|\big|^2_{L^2(\mathbb{R}^n} &= \int_{\mathbb{R}^n} \lim_{N\to\infty} \Big| \sum_{j=-N}^{+N} \mathcal{K}_j[\varphi](x) \Big|^2 dx \\ &\leq \limsup_{N\to\infty} \int_{\mathbb{R}^n} \Big| \sum_{j=-N}^{+N} \mathcal{K}_j[\varphi](x) \Big|^2 dx \\ &\leq A^2 \big|\big|\, \varphi\, \big|\big|^2_{L^2(\mathbb{R}^n)}.\end{aligned}$$

It follows from (1) and (3) that

$$\big|\big|\, K_j\, \big|\big|_{L^1(\mathbb{R}^n)} = \int_{|x|<2^j} \big|K_j(x)\big|\, dx \leq C\, 2^{-nj}\, \big|\mathbb{B}_E(0; 2^j)\big| \leq C\, n^{-1}\, \omega_n$$

since $n^{-1}\omega_n$ is the volume of the Euclidean unit ball in $\mathbb{R}^n$. Now if $f, g \in L^1(\mathbb{R}^n)$, we always have

$$\big|\big|\, f*g\, \big|\big|_{L^1} = \int_{\mathbb{R}^n} |f*g(x)|\, dx \leq \iint_{\mathbb{R}^n\times\mathbb{R}^n} |f(x-y)|\, |g(y)|\, dy\, dx = \big|\big|\, f\, \big|\big|_{L^1}\, \big|\big|\, g\, \big|\big|_{L^1},$$

and consequently $\big|\big|\, K_j * K_k\, \big|\big|_{L^1} \leq (C\, n^{-1}\, \omega_n)^2$. However, the differential inequality (2) and the cancellation property (4) allow us to get a better estimate when $j \neq k$.

Proposition 2.12. *For all $j, k \in \mathbb{Z}$ we have*

$$\big|\big|\, K_j * K_k\, \big|\big|_{L^1} \leq 2^n\, (C\, n^{-1}\, \omega_n)^2\, 2^{-|j-k|}.$$

PROOF. Without loss of generality, assume that $j < k$. Then

$$\begin{aligned}
\big|K_j * K_k(x)\big| &= \Big|\int_{\mathbb{R}^n} K_k(x-y)\,K_j(y)\,dy\Big| \\
&= \Big|\int_{\mathbb{R}^n} \big[K_k(x-y) - K_k(x)\big]\,K_j(y)\,dy\Big| && \text{(using assumption (4))} \\
&\leq \int_{\mathbb{B}_E(0;2^j)} \big|K_k(x-y) - K_k(x)\big|\,\big|K_j(y)\big|\,dy \\
&\leq \int_{\mathbb{B}_E(0;2^j)} |y| \sup_{z\in\mathbb{R}^n} \big|\nabla K_k(z)\big|\,\big|K_j(y)\big|\,dy && \text{(Mean Value Theorem)} \\
&\leq C\,2^{j-(n+1)k}\big|\big|\,K_j\,\big|\big|_{L^1} && \text{(using estimate (3)).}
\end{aligned}$$

On the other hand, if $K_j * K_k(x) \neq 0$ we must have $|y| \leq 2^j$ and $|x-y| \leq 2^k$, so $|x| \leq |x-y| + |y| \leq 2 \cdot 2^k$. Thus

$$\big|\big|\,K_j * K_k\,\big|\big|_{L^1} \leq C\,2^{j-(n+1)k}\big|\big|\,K_j\,\big|\big|_{L^1}\big|\mathbb{B}(0;2\cdot 2^k)\big| \leq 2^n\,(C\,n^{-1}\,\omega_n)^2\,2^{j-k}.$$

This completes the proof. □

LEMMA 2.13. *Let $\{K_j\}$ be functions satisfying conditions (1) through (4), and for each $j \in \mathbb{Z}$ define an operator T_j by setting*

$$T_j[f](x) = K_j * f(x) = \int_{\mathbb{R}^n} K_j(x-y)\,f(y)\,dy.$$

There exists a constant C so that for any integer N we have

$$\Big|\Big|\sum_{j=-N}^{N} T_j[f]\Big|\Big|_{L^2(\mathbb{R}^n)} \leq C\,\big|\big|\,f\,\big|\big|_{L^2(\mathbb{R}^n)}.$$

2.7. L^1-estimates.

It is not true that the operator $\mathcal{K}$ from section 2.6, defined on the space $\mathcal{C}_0^\infty(\mathbb{R}^n)$, extends to a bounded operator on $L^1(\mathbb{R}^n)$, but we do have the following replacement, which is called a *weak type (1,1)* estimate. Let us write $\mathcal{K}^{[N]} = \sum_{j=-N}^{+N} \mathcal{K}_j$. We will need the following estimate, which is sometimes called the Caldéron-Zygmund estimate.

LEMMA 2.14. *Let $\eta > A_2$, let $B = \mathbb{B}_\rho(x_0;\delta_0)$, and let $B^* = \mathbb{B}_\rho(x_0;\eta\,\delta_0)$. Suppose $x_1, x_2 \in B$. Then*

$$\int_{\mathbb{R}^n - B^*} \big|K^{[N]}(y-x_1) - K^{[N]}(y-x_2)\big|\,dy \leq C.$$

THEOREM 2.15. *There is a constant A independent of N so that if $f \in L^1(\mathbb{R}^n)$, then*

$$\Big|\Big\{x \in \mathbb{R}^n\,\Big|\,\big|\mathcal{K}^{[N]}[f](x) > \alpha\Big\}\Big| \leq \frac{A}{\lambda}\,\big|\big|\,f\,\big|\big|_{L^1(\mathbb{R}^n)}.$$

PROOF. We apply the Caldéron-Zygmund decomposition of Theorem 1.14 to the function f. Let $E_\alpha = \Big\{x \in \mathbb{R}^n\,\Big|\,\mathcal{M}[f](x) > \alpha\Big\}$. Then we can write $f = g + \sum_j b_j = g + b$ where g and $\{b_j\}$ satisfy

(1) b_j is supported on a ball $\mathbb{B}_E(x_j;\delta_j) \subset E_\alpha$ and $\int_{\mathbb{R}^n} b_j(x)\,dx = 0$ while $\big|\big| b_j \big|\big|_{L^1(\mathbb{R}^n)} \leq C\,\alpha\,\big|B_j\big|$;

(2) If $x \notin E_\alpha$, then $|x - x_j| \geq 2\delta_j$;

(3) $\sum_j \big|\mathbb{B}_E(x_j;\delta_j)\big| \leq C\,\big|E_\alpha\big| \leq \frac{C}{\alpha}\,\big|\big| f \big|\big|_{L^1(\mathbb{R}^n)}$;

(4) $|g(x)| \leq C\,\alpha$ for almost all $x \in \mathbb{R}^n$.

Since $|\mathcal{K}^{[N]}[f](x)| \leq |\mathcal{K}^{[N]}[g](x)| + |\mathcal{K}^{[N]}[b](x)|$, we have

$$\begin{aligned}
&\Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[f](x) > \alpha\Big\}\Big| \\
&\qquad \leq \Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[g](x) > \frac{\alpha}{2}\Big\}\Big| \cup \Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[b](x) > \frac{\alpha}{2}\Big\}\Big|.
\end{aligned}$$

Now $\mathcal{K}^{[N]}$ is a bounded operator on $L^2(\mathbb{R}^n)$ with norm A independent of N. Thus using Corollary 1.15 we have

$$\begin{aligned}
\frac{\alpha^2}{4}\Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[g](x) > \frac{\alpha}{2}\Big\}\Big| &\leq \int_{\mathbb{R}^n} |\mathcal{K}^{[N]}[g](x)|^2\,dx \\
&\leq A\big|\big| g \big|\big|_{L^2}^2 \\
&\leq A\,C\,\alpha\,\big|\big| f \big|\big|_{L^1}
\end{aligned}$$

and so

$$\Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[g](x) > \frac{\alpha}{2}\Big\}\Big| \leq \frac{4AC}{\alpha}\,\big|\big| f \big|\big|_{L^1}.$$

Now suppose we can show that

$$\int_{\mathbb{R}^n - E_\alpha} |\mathcal{K}^{[N]}[b](x)|\,dx \leq C\big|\big| f \big|\big|_{L^1}. \tag{2.6}$$

Then

$$\begin{aligned}
&\Big|\Big\{x \in \mathbb{R}^n \,\Big|\, \big|\mathcal{K}^{[N]}[b](x) > \frac{\alpha}{2}\Big\}\Big| \\
&\qquad\qquad \leq \Big|E_\alpha\Big| + \Big|\Big\{x \in \mathbb{R}^n - E_\alpha \,\Big|\, \big|\mathcal{K}^{[N]}[b](x) > \frac{\alpha}{2}\Big\}\Big| \\
&\qquad\qquad \leq \frac{C_1}{\alpha}\big|\big| f \big|\big|_{L^1} + \frac{2}{\alpha}\int_{\mathbb{R}^n - E_\alpha} |\mathcal{K}^{[N]}[b](x)|\,dx \\
&\qquad\qquad \leq \frac{C'}{\alpha}\,\big|\big| f \big|\big|_{L^1},
\end{aligned}$$

which would complete the proof. Thus the key is to prove the estimate is equation (2.6).

Suppose $x \notin E_\alpha$. Since the integral of b_j is zero, we have

$$\begin{aligned}
\big|\mathcal{K}^{[N]}[b_j](x)\big| &= \Big|\int_{\mathbb{R}^n} K^{[N]}(x-y)\,b_j(y)\,dy\Big| \\
&= \Big|\int_{\mathbb{R}^n} \big[K^{[N]}(x-y) - K(x-x_j)\big]\,b_j(y)\,dy\Big| \\
&\leq \int_{\mathbb{R}^n} \big|K^{[N]}(x-y) - K^{[N]}(x-x_j)\big|\,|b_j(y)|\,dy
\end{aligned}$$

and hence

$$\begin{aligned}\int_{\mathbb{R}^n - E_\alpha} \big|\mathcal{K}^{[N]}[b_j](x)\big|\,dx \\ &\leq \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n - E_\alpha}\big|K^{[N]}(x-y) - K^{[N]}(x-x_j)\big|\,dx\right]\,|b_j(y)|\,dy \\ &\leq C\int_{\mathbb{R}^n}|b_j(y)|\,dy \\ &\leq C'\,\alpha\,\big|B_j\big|.\end{aligned}$$

It follows that

$$\begin{aligned}\int_{\mathbb{R}^n - E_\alpha}\big|\mathcal{K}^{[N]}[b](x)\big|\,dx &\leq \sum_j \int_{\mathbb{R}^n - E_\alpha}\big|\mathcal{K}^{[N]}[b_j](x)\big|\,dx \\ &\leq C'\,\alpha\sum_j\big|B_j\big| \\ &\leq C''\big|\big|\,f\,\big|\big|_{L^1}.\end{aligned}$$

□

3. The heat operator and a non-isotropic metric on $\mathbb{R}^{n+1}$

We introduce coordinates $(t,x) = (t,x_1,\ldots,x_n)$ on $\mathbb{R}\times\mathbb{R}^n = \mathbb{R}^{n+1}$. Then the heat operator is

$$\big(\partial_t - \triangle_x\big)[u] = \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x_1^2} - \cdots - \frac{\partial^2 u}{\partial x_n^2}.$$

The inhomogeneous equation $\partial_t[u] - \triangle_x[u] = g$ describes heat flow in the presence of heat sources specified by g. The heat operator is a typical example of a parabolic partial differential operator.

In section 3.1 we review the symmetry properties of this operator, and introduce a non-isotropic metric on $\mathbb{R}^{n+1}$ which turns out to be the appropriate analogue of the standard Euclidean metric for the Laplace operator. In section 3.2 we study the initial value problem for the heat operator, find an explicit formula for the heat kernel, and use this to construct a fundamental solution for the heat operator. In section 3.3, we see how the non-isotropic geometry is reflected in the boundary behavior of solutions to the initial value problem.

3.1. Symmetry properties.

Since $\partial_t - \triangle_x$ has constant coefficients, it follows that the heat operator, like the Laplace operator, is invariant under translations. However, the heat operator is not homogeneous with respect to the standard Euclidean dilations, and this is perhaps the first sign that the underlying geometry is different.

We can define a family of *non-isotropic* dilations on $\mathbb{R}^{n+1}$ by setting

$$D_{H,\lambda}(t,x_1,\ldots,x_n) = (\lambda^2 t, \lambda x_1,\ldots,\lambda x_n).$$

The corresponding action on functions is given by

$$D_{H,\lambda}[\varphi](t,x) = \varphi(\lambda^{-2}t, \lambda^{-1}x).$$

Like Euclidean dilations, each mapping $D_{H,\delta} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is an automorphism of the vector space structure; *i.e.*

$$D_{H,\delta}\big[(t,x)+(s,y)\big] = D_{H,\delta}[(t,x)] + D_{H,\delta}[(s,y)].$$

Also $D_{H,\delta_1} \circ D_{H,\delta_2} = D_{H,\delta_1\delta_2}$.

The reason for introducing these non-isotropic dilations is that we have

$$\big[\partial_t - \triangle_x\big]\big[D_{H,\lambda}[\varphi]\big] = \lambda^{-2}\big[\partial_t - \triangle_x\big][\varphi], \tag{3.1}$$

so that the heat operator is homogeneous with respect to these new dilations. This is clearly the analogue of part (c) of Proposition 2.1.

We can also introduce a pseudo-metric d_H on $\mathbb{R}^{n+1}$ which is compatible with this family of dilations. (Here the notation d_H stands for a 'heat' distance). Put

$$d_H\big((t,x),(s,y)\big) = \big[(t-s)^2 + |x-y|^4\big]^{\frac{1}{4}}.$$

It is not hard to check that d_H is a pseudo-metric. Let us put

$$\mathbb{B}_H\big((x,t),\delta\big) = \left\{(s,y) \in \mathbb{R}^{n+1} \,\middle|\, d_H\big((t,x),(s,y)\big) < \delta\right\}.$$

Then the volume of such a ball of radius δ is a constant times δ^{n+2}.

The balls in this new metric are quite different from the standard Euclidean balls. Note that if $d_H\big((t,x),(s,y)\big) = \big[(t-s)^2 + |x-y|^4\big]^{\frac{1}{4}} < \delta$, then $|t-s| < \delta^2$ and $|x-y| < \delta$. Conversely, if $|t-s| < \delta^2$ and $|x-y| < \delta$, then $d_H\big((t,x),(s,y)\big) < 2^{\frac{1}{4}}\,\delta$. Thus $B_H\big((t,x);\delta\big)$ is essentially the Cartesian product of a Euclidean ball of radius δ in the x-variables with an interval of length δ^2 in the t-variable. For small δ, the balls are much smaller in the t-direction than in the x-directions, which for δ large, the situation is reversed. Thus the balls are non-isotropic.

In general, if A is a symmetric $n \times n$ real matrix, we can define a family of dilation on $\mathbb{R}^n$ by setting

$$D_{A,\delta}[x] = e^{\log(\delta)A}[x].$$

Then $D_{A,1}$ is the identity operator, and $D_{A,\delta_1} \circ D_{A,\delta_2} = D_{A,(\delta_1\delta_2)}$. If all the eigenvalues of A are positive, we have $\lim_{\delta\to 0^+} D_{A,\delta}[x] = 0$. If A is the identity matrix, then $D_{A,\delta}$ is the usual family of Euclidean dilations. If A is the $(n+1)\times(n+1)$ diagonal matrix with entries $\{2,1,\dots,1\}$, then $D_{A,\delta} = D_{H,\delta}$.

We say that the *homogeneous dimension* of $\mathbb{R}^n$ under the family of dilations $\{D_{A,\delta}\}$ is $Tr[A]$, the trace of the matrix A or equivalently the sum of the eigenvalues of A. Then the homogeneous dimension of ordinary Euclidean dilations on $\mathbb{R}^n$ is n, while the homogeneous dimension of $\mathbb{R}^{n+1}$ under the dilations $\{D_{H,\delta}\}$ is $n+2$. Thus both N_0 and H_0 are homogeneous (on $\mathbb{R}^n$ or $\mathbb{R}^{n+1}$) of degree 2 minus the relevant homogeneous dimension.

3.2. The initial value problem and a fundamental solution.

We begin by considering the following initial value problem for the heat operator. Given $f \in L^2(\mathbb{R}^n)$, we want to find a function $F \in \mathcal{C}^\infty(\mathbb{R}\times\mathbb{R}^n)$ such that:

(1) $\displaystyle\frac{\partial F}{\partial t}(t,x) - \sum_{j=1}^n \frac{\partial^2 F}{\partial x_j^2}(t,x) = 0$ for $t > 0$ and $x \in \mathbb{R}^n$.

(2) If we put $F_t(x) = F(t,x)$, then $\lim_{t\to 0} F_t = f$ with convergence in $L^2(\mathbb{R}^n$.

To motivate the solution, we argue informally as follows. Suppose F were a solution. Let

$$\mathcal{F}_x[F](t,\xi) = \widehat{F}(t,\xi) = \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, F(t,x)\,dx,$$

be the partial Fourier transform of F in the x-variables. Then if we could integrate by parts and there were no boundary terms, the partial differential equation in (1) would become the ordinary differential equation

$$\frac{d\widehat{F}}{dt}(t,\xi) = -4\pi^2|\xi|^2\,\widehat{F}(t,\xi).$$

This suggests that $\widehat{F}(t,\xi) = C(\xi)\,\exp(-4\pi^2|\xi|^2 t)$, and if F is to satisfy the initial condition given in (2), then we should take $C(\xi) = \widehat{f}(\xi)$, or

$$\widehat{F}(t,\xi) = \widehat{f}(\xi)\,e^{-4\pi^2|\xi|^2\,t}.$$

Taking the inverse Fourier transform, we would get

$$F(t,x) = \int_{\mathbb{R}^n} H_t(x-y)\,f(y)\,dy$$

where

$$H_t(x) = \int_{\mathbb{R}^n} e^{2\pi i x\cdot\xi}\,e^{-4\pi\,|\xi|^2 t}\,d\xi = (4\pi\,t)^{-\frac{n}{2}}\,e^{-\frac{|x|^2}{4\,t}},$$

the inverse partial Fourier transform of $e^{-4\pi^2|\xi|^2 t}$. This informal argument is justified by the following result. Define

$$H_0(t,x) = H_t(x) = \begin{cases} (4\pi\,t)^{-\frac{n}{2}}\,e^{-\frac{|x|^2}{4\,t}} & \text{if } t>0,\\ 0 & \text{if } t\le 0.\end{cases} \tag{3.2}$$

THEOREM 3.1. *Let $1\le p<\infty$, and let $f\in L^p(\mathbb{R}^n)$. Put*

$$F(t,x) = H_t * f(x) = \int_{\mathbb{R}^n} (4\pi\,t)^{-\frac{n}{2}}\,e^{-\frac{|x-y|^2}{4\,t}}\,f(y)\,dy.$$

Then $F\in\mathcal{C}^\infty\big((0,\infty)\times\mathbb{R}^n\big)$ and

(1) $\displaystyle\frac{\partial F}{\partial t}(t,x) - \sum_{j=1}^n \frac{\partial^2 F}{\partial x_j^2}(t,x) = 0$ *for $t>0$ and $x\in\mathbb{R}^n$;*

(2) $\displaystyle\lim_{t\to 0}\big|\big|\,H_t * f - f\,\big|\big|_{L^p(\mathbb{R}^n)} = 0.$

(3) If $\varphi\in\mathcal{C}_0^\infty(\mathbb{R}^n)$ then $\displaystyle\lim_{t\to 0}\big|\big|\,H_t*\varphi-\varphi\,\big|\big|_{L^\infty(\mathbb{R}^n)} = 0.$

PROOF. A direct calculation shows that $\dfrac{\partial H_0}{\partial t}(t,x) - \triangle_x[H_0](t,x) = 0$ for $t>0$ and $x\in\mathbb{R}^n$. For $t>0$, the rapid decay of $H_0(t,x)$ as $|x|\to\infty$ justifies differentiating under the integral sign, an so $\dfrac{\partial F}{\partial t}(t,x) - \triangle_x[F](t,x) = 0$ for $t>0$ and $x\in\mathbb{R}^n$. Next,

$$\int_{\mathbb{R}^n} H_t(x)\,dx = \pi^{-\frac{n}{2}}\int_{\mathbb{R}^n} e^{-|x|^2}\,dx = 1.$$

Thus if $\varphi\in\mathcal{C}_0^\infty(\mathbb{R}^n)$ we have

$$H_t*\varphi(x) - \varphi(x) = \int_{\mathbb{R}^n} (4\pi\,t)^{-\frac{n}{2}}\,e^{-\frac{|y|^2}{4t}}\,\big[\varphi(x-y)-\varphi(x)\big]\,dy$$

It follows from Minkowski's inequality for integrals that

$$\big|\big|\, H_t * \varphi - \varphi \,\big|\big|_{L^p(\mathbb{R}^n)} \leq \int_{\mathbb{R}^n} (4\pi\, t)^{-\frac{n}{2}} e^{-\frac{|y|^2}{4t}}\, \big|\big|\, T_y[\varphi] - \varphi \,\big|\big|_{L^p(\mathbb{R}^n)}\, dy$$

where $T_y\varphi(x) = \varphi(x-y)$ is the translation operator. Translation is continuous in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. Thus if $\epsilon > 0$, there exists $\eta > 0$ so that $|y| \leq \eta$ implies $\big|\big|\, T_y[\varphi] - \varphi \,\big|\big|_{L^p(\mathbb{R}^n)} < \epsilon$. It follows that

$$\begin{aligned}
&\big|\big|\, H_t * \varphi - \varphi \,\big|\big|_{L^p(\mathbb{R}^n)} \\
&\qquad \leq \epsilon \int_{|y|<\eta} (4\pi\, t)^{-\frac{n}{2}} e^{-\frac{|y|^2}{4t}}\, dy + 2\big|\big|\, \varphi \,\big|\big|_{L^p(\mathbb{R}^n)} \int_{|y|>\eta} (4\pi\, t)^{-\frac{n}{2}} e^{-\frac{|y|^2}{4t}}\, dy \\
&\qquad < \epsilon + 2\big|\big|\, \varphi \,\big|\big|_{L^p(\mathbb{R}^n)}\, \pi^{-\frac{n}{2}} \int_{|y|>\eta/(2\sqrt{t})} e^{-|y|^2}\, dy.
\end{aligned}$$

The second term goes to zero as $t \to 0$, and so $\lim_{t\to 0} \big|\big|\, H_t * \varphi - \varphi \,\big|\big|_{L^p(\mathbb{R}^n)} = 0$. Since the space $\mathcal{C}_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, this completes the proof. □

We now show that a fundamental solution for the heat operator on $\mathbb{R} \times \mathbb{R}^n$ is given by $H\big((t,x),(s,y)\big) = H_0(t-s, x-y)$.

LEMMA 3.2. *Convolution with H is a fundamental solution for the heat operator on $\mathbb{R} \times \mathbb{R}^n$. Precisely, if $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^{n+1})$, define*

$$\mathcal{H}[\varphi](t,x) = \iint_{\mathbb{R}^{n+1}} H(t-s, x-y)\, \varphi(s,y)\, ds\, dy.$$

Then $\mathcal{H}[\varphi] \in \mathcal{C}^\infty(\mathbb{R}^{n+1})$ and

$$\begin{aligned}
\varphi(t,x) &= \mathcal{H}\big[(\partial_t - \triangle_x)[\varphi]\big](t,x), \\
\varphi(t,x) &= \big(\partial_t - \triangle_x\big)\big[\mathcal{H}[\varphi]\big](t,x).
\end{aligned}$$

PROOF. Integrating by parts, and using the fact that $\partial_t H(t,x) = \triangle_x H(t,x)$ for $t > 0$, we have

$$\begin{aligned}
\int_{\mathbb{R}^n}\int_\epsilon^\infty H(s,y)\, \frac{\partial\varphi}{\partial t}(t-s, x-y)\, ds\, dy &= \int_{\mathbb{R}^n}\int_\epsilon^\infty \frac{\partial H}{\partial s}(s,y)\, \varphi(t-s,x-y)\, ds\, dy \\
&\qquad + \int_{\mathbb{R}^n} H(\epsilon,y)\varphi(t-\epsilon, x-y)\, dy \\
&= \int_{\mathbb{R}^n}\int_\epsilon^\infty \triangle_y H(s,y)\, \varphi(t-s,x-y)\, ds\, dy \\
&\qquad + \int_{\mathbb{R}^n} H(\epsilon,y)\varphi(t-\epsilon, x-y)\, dy \\
&= \int_{\mathbb{R}^n}\int_\epsilon^\infty H(s,y)\, \triangle_x\varphi(t-s,x-y)\, ds\, dy \\
&\qquad + \int_{\mathbb{R}^n} H(\epsilon,y)\varphi(t-\epsilon, x-y)\, dy.
\end{aligned}$$

Thus

$$\begin{aligned}\int_{\mathbb{R}^n}\int_\epsilon^\infty & H(s,y)\,\big[\frac{\partial\varphi}{\partial t}-\triangle_x[\varphi]\big](t-s,x-y)\,ds\,dy\\ &=\int_{\mathbb{R}^n} H(\epsilon,y)\varphi(t-\epsilon,x-y)\,dy\\ &=\int_{\mathbb{R}^n} H(\epsilon,y)\varphi(t,x-y)\,dy+\int_{\mathbb{R}^n} H(\epsilon,y)[\varphi(t-\epsilon,x-y)-\varphi(t,x-y)]\,dy\end{aligned}$$

If we take the limit as $\epsilon\to 0$, the first integral on the right converges to $\varphi(t,x)$, and easy estimates show that the second integral tends to zero. It follows that

$$\int_{\mathbb{R}^n}\int_0^\infty H(s,y)\,\big[\frac{\partial\varphi}{\partial t}-\triangle_x[\varphi]\big](t-s,x-y)\,ds\,dy=\varphi(t,x).$$

We now ask whether the behavior of the heat kernel $H(t,x)$ can be described in terms of a metric on $\mathbb{R}\times\mathbb{R}^n$. Let us write

$$H\big((t,x),(s,y)\big)=H(t-s,x-y).$$

One *cannot* express the size of H in terms of the standard Euclidean metric. For example, if $x=y$, the Euclidean distance between (t,x) and (s,y) is $|t-s|$ and for such points we have

$$H\big((t,x),(y,s)\big)\sim d_E\big((t,x),(s,y)\big)^{-\frac{n}{2}}.$$

On the other hand, if $t-s=|x-y|^2$ and if $|x-y|$ is small, we have

$$d_E\big((t,x),(s,y)\big)=\sqrt{|x-y|^2+|t-s|^2}=\sqrt{|x-y|^2+|x-y|^4}\approx|x-y|,$$

and for such points we have

$$H\big((t,x),(y,s)\big)\approx d_E\big((t,x),(s,y)\big)^{-n}.$$

Thus $H\big((t,x),(y,s)\big)$ is not comparable to any fixed power of the Euclidean distance.

Now one can check that

$$H\big((t,x),(s,y)\big)\leq C_0\,\frac{d_H((t,x),(s,y))^2}{\big|\mathbb{B}_H\big((x,t),d_H((t,x),(s,y))\big)\big|},\tag{3.3}$$

and more generally

$$\big|\partial_x^\alpha\partial_y^\beta\partial_t^\gamma\partial_s^\delta H\big((t,x),(s,y)\big)\big|\leq C_{\alpha,\beta,\gamma,\delta}\,\frac{d_H((t,x),(s,y))^{2-|\alpha|-|\beta|-2|\gamma|-2|\delta|}}{\big|\mathbb{B}_H\big((x,t),d_H((t,x),(s,y))\big)\big|}.\tag{3.4}$$

It is clear that the introduction of the metric d_H allows us to write estimates for the fundamental solutions N and H in equations (2.6) and (3.3) which have exactly the same form. The estimates for derivatives in equations (**??**) and (3.4) are very similar, but there is an important difference. When dealing with the heat equation, each s or t derivative of the fundamental solution introduces a factor of $d_H((t,x),(s,y))^{-2}$, and thus behaves as though it were two derivatives in x or y.

□

3.3. Parabolic convergence.

If we were to replace the heat operator by the Laplace operator $-\frac{\partial^2}{\partial t^2} - \triangle_x$, we could pose an initial value problem similar to that discussed in the last section. Now the problem amounts to finding a harmonic function $u(t,x)$ on $(0,\infty) \times \mathbb{R}^n$ such that $\lim_{t\to 0} u(t,\,\cdot\,)$ is a prescribed function $f \in L^p(\mathbb{R}^n)$. This is called the Dirichlet problem, and if one imitates the proof of Theorem 3.1, one can check that the solution is given by the Poisson integral

$$u(t,x) = \Gamma\left(\frac{n+1}{2}\right) \pi^{-\frac{n+1}{2}} \int_{\mathbb{R}^n} \frac{t}{|x-y|^2 + t^2} f(y)\, dy.$$

In addition to convergence in norm, it is also a classical fact that $u(t,y)$ converges pointwise to $f(x)$ for almost all $x \in \mathbb{R}^n$ when (y,t) lies in a non-tangential approach region with vertex at x. Thus for each $\alpha > 0$ consider the conical region

$$C_\alpha(x) = \left\{(t,y) \in (0,\infty) \times \mathbb{R}^n \,\middle|\, |y-x| < \alpha\, t\right\}.$$

The result is that if $f \in L^p(\mathbb{R}^n)$, then except for x belonging to a set of measure zero, we have

$$\lim_{\substack{(t,y)\to(0,x)\\(t,y)\in C_\alpha(x)}} u(t,y) = f(x).$$

It is well known[2] that this results depends on estimating the non-tangential maximal function in terms of the Hardy-Littlewood maximal function:

$$\sup_{(t,y)\in C_\alpha(x)} |u(t,y)| \leq A_\alpha\, \mathcal{M}[f](x).$$

We can clearly see the difference between the standard Euclidean setting and the non-isotropic geometry associated to the heat operator when we study pointwise convergence as $t \to 0$ of the solution $H_t * f$. For each $x \in \mathbb{R}^n$ and each $\alpha > 0$, consider the *parabolic* approach region

$$\Gamma_\alpha(x) = \left\{(t,y) \in (0,\infty) \times \mathbb{R}^n \,\middle|\, |y-x| < \alpha\sqrt{t}\right\}.$$

The analogue of the statement about pointwise convergence of the Poisson integral is then:

Let $1 \leq p \leq \infty$ and let $f \in L^p(\mathbb{R}^n)$. Then there is a set $E \subset \mathbb{R}^n$ with Lebesgue measure zero so that for all $x \notin E$ and all $\alpha > 0$

$$\lim_{\substack{(t,y)\to(0,x)\\(t,y)\in \Gamma_\alpha(x)}} H_t * f(y) = f(x).$$

This result depends on the following estimate for the parabolic maximal function.

LEMMA 3.3. *For each $\alpha > 0$ there is a constant $C = C(n,\alpha) > 0$ depending only on the dimension n and on α so that if $f \in L^1(\mathbb{R}^n)$ and if $F(t,x) = H_t * f(x)$, then*

$$\sup_{(t,y)\in\Gamma_\alpha(x)} |F(t,y)| \leq C(n,\alpha)\, \mathcal{M}[f](x).$$

---

[2]See for example, [**Ste70**], page 197.

PROOF. Let

$$R_0 = \left\{ z \in \mathbb{R}^n \,\middle|\, |z| < 2\alpha\sqrt{t} \right\}, \qquad \text{and}$$
$$R_j = \left\{ z \in \mathbb{R}^n \,\middle|\, 2^j\alpha\sqrt{t} \le |z| < 2^{j+1}\alpha\sqrt{t} \right\} \qquad \text{for } j \ge 1.$$

Note that if $|x-y| < \alpha\sqrt{t}$ and if $|z| \ge 2^j\alpha\sqrt{t}$ for some $j \ge 1$, then

$$\frac{|z-(y-x)|}{2\sqrt{t}} \ge \frac{|z|-|x-y|}{2\sqrt{t}} \ge 2^{j-1}\,\alpha.$$

Then if $(t,y) \in \Gamma_\alpha(x)$ we have

$$\begin{aligned}
|F(t,y)| &\le (4\pi\,t)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{-\frac{|(y-x)-z|^2}{4t}}\,|f(x+z)|\,dz \\
&= (4\pi\,t)^{-\frac{n}{2}} \sum_{j=0}^{\infty} \int_{R_j} e^{-\frac{|(y-x)-z|^2}{4t}}\,|f(x+z)|\,dz \\
&= \frac{\omega_n}{n}\left(\frac{\alpha}{\sqrt{\pi}}\right)^n \Bigg[|B(x,2\alpha\sqrt{t})|^{-1} \int_{B(x,2\alpha\sqrt{t})} |f(z)|\,dz \\
&\qquad \sum_{j=1}^{\infty} 2^{nj} e^{-2^{j-1}\alpha}\,|B(x,2^{j+1}\alpha\sqrt{t})|^{-1} \int_{B(x,2^{j+1}\alpha\sqrt{t})} |f(z)|\,dz \Bigg] \\
&\le C(n,\alpha)\,\mathcal{M}[f](x),
\end{aligned}$$

as asserted. □

4. Operators on the Heisenberg group

Our next example is motivated by problems arising in complex analysis in several variables.

4.1. The Siegal upper half space and its boundary.

In the complex plane $\mathbb{C}$, the upper half plane $U = \left\{ z = x+iy \,\middle|\, y > 0 \right\}$ is biholomorphically equivalent to the open unit disk $\mathbb{D} = \left\{ w \in \mathbb{C} \,\middle|\, |w| < 1 \right\}$ via the mapping

$$U \ni z \longrightarrow w = \frac{z-i}{z+i} \in \mathbb{D}.$$

In $\mathbb{C}^{n+1}$, the domain analogous to U is the Siegel upper half-space

$$\mathcal{U}_{n+1} = \left\{ (z_1,\ldots,z_n,z_{n+1}) \in \mathbb{C}^{n+1} \,\middle|\, \Im m[z_{n+1}] > \sum_{j=1}^{n} |z_j|^2 \right\},$$

which is biholomorphic to the open unit ball

$$\mathbb{B}_{n+1} = \left\{ (w_1,\ldots,w_n,w_{n+1}) \in \mathbb{C}^{n+1} \,\middle|\, \sum_{j=1}^{n+1} |w_j|^2 < 1 \right\}$$

via the mapping $\mathcal{U}_{n+1} \ni z \longrightarrow w \in \mathcal{B}_{n+1}$ given by

$$w = (w_0, w_1, \ldots, w_N) = \left(\frac{2z_1}{z_{n+1}+i}, \ldots, \frac{2z_n}{z_{n+1}+i}, \frac{z_{n+1}-i}{z_{n+1}+i} \right).$$

We can identify the boundary of $\mathcal{U}_{n+1}$ with $\mathbb{C}^n \times \mathbb{R}$ via the mapping given by

$$\partial\mathcal{U}_{n+1} \ni \big(z_1, \ldots, z_n, t + i\sum_{j=1}^{n} |z_j|^2\big) \longleftrightarrow (z_1, \ldots, z_n, t) \in \mathbb{C}^n \times \mathbb{R}.$$

The boundary of the upper half plane $U \subset \mathbb{C}$ is the set of real numbers $\mathbb{R}$. We can identify this boundary with a subgroup of the group of biholomorphic mappings of $\mathbb{C}$ to itself given by translations. Thus for each $a \in \mathbb{C}$, let $T_a(z) = a + z$. Then $\Im m[T_a(z)] = \Im m[a] + \Im m[z]$, so T_a carries horizontal lines to horizontal lines. If $a \in \mathbb{R}$, then T_a carries each horizontal line to itself, and carries U to U. The set of translations $\big\{T_a\big\}_{a\in\mathbb{R}}$ is a group under composition, which is isomorphic to the additive group structure of $\mathbb{R}$.

To find the analogue in several variables, we proceed as follows. We write elements $z \in \mathbb{C}^{n+1}$ as $z = (z', z_{n+1})$ where $z' \in \mathbb{C}^n$ and $z_{n+1} \in \mathbb{C}$. For each $a = (a', a_{n+1}) \in \mathbb{C}^{n+1}$ consider the affine mapping

$$T_a(z) = T_{(a',a_{n+1})}(z', z_{n+1}) = \big(a' + z', a_{n+1} + z_{n+1} + 2i < z', a' > \big)$$

where $< z', a' >= \sum_{j=1}^{n} z_j \bar{a}_j$ is the Hermitian inner produce on $\mathbb{C}^n$. Note that T_0 is the identity map, and that the collection of mappings $\{\{T_z\big\}_{z\in\mathbb{C}^{n+1}}$ is closed under composition and taking inverses. In fact

$$T_{(a',a_{n+1})} \circ T_{(b',b_{n+1})} = T_{(a'+b',a_{n+1}+b_{n+1}+2i<b',a'>)}$$
$$(T_{(a',a_{n+1})})^{-1} = T_{(-a',-a_{n+1}+2i<a',a'>)}.$$

It follows that if we set

$$(a', a_{n=1}) \cdot (b', b_{n=1}) = (a' + b', a_{n+1} + b_{n+1} + 2i\ \ < b', a' >), \tag{4.1}$$

then $\mathbb{C}^{n+1}$ becomes a group with this product, and

$$T_{(a',a_{n+1})}(z', z_{n+1}) = (a', a_{n+1}) \cdot (z', z_{n+1}).$$

The analogue of the height function $\Im m[z]$ in one variable is the function

$$\rho(z) = \rho(z', z_{n+1}) = \Im m[z_{n+1}] - \sum_{j=1}^{n} |z_j|^2.$$

A simple calculation shows that

$$\rho\big((a', a_{n+1}) \cdot (b', b_{n+1})\big) = \rho(a', a_{n+1}) + \rho(b', b_{n+1}). \tag{4.2}$$

Thus T_a maps each level surface $M_t = \{\rho(z) = t\}$ into the level surface $M_{t+\rho(a)}$. Now M_0 is the boundary of $\mathcal{U}_{n+1}$, and if $a \in M_0$, then T_a maps each level surface M_t into itself, and maps $\mathcal{U}_{n+1}$ into itself.

It follows from equation (4.2) that the collection of mappings $\big\{T_z\big\}_{z\in M_0}$ is a subgroup of $\mathbb{C}^{n+1}$ with multiplication "$\cdot$", and if we use the coordinates $(z, t) \in \mathbb{C}^n \times \mathbb{R}$, this multiplication is given by

$$(z, t) \cdot (w, s) = \big(z + w, t + s + 2\Im m[< w, z >]\big). \tag{4.3}$$

$\mathbb{C}^n \times \mathbb{R}$ with this multiplication is called the n-dimensional *Heisenberg group* $\mathbb{H}_n$. The point $(0, 0)$ is the identity, and $(-z, -t)$ is the inverse of (z, t). It is easy to check that $(z, t) \cdot (w, s) = (w, s) \cdot (z, t)$ if and only if $\Im m[< z, w >] = 0$, so the group is not commutative. It is sometimes convenient to use real coordinates on

$\mathbb{H}_n = \mathbb{C}^n \times \mathbb{R} = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. Thus we write $z_j = x_j + iy_j$, and write $(z,t) = (x,y,t)$. Then the Heisenberg multiplication is given by

$$(x,y,t)\cdot(u,v,s) = (x+u, y+v, t+s+2[< y,u > - < x,v >])$$

where now $< y,u >$ and $< x,v >$ stand for the standard Euclidean inner product on $\mathbb{R}^n$.

4.2. Dilations and translations on $\mathbb{H}_n$.

We introduce a family of dilations on $\mathbb{C}^n \times \mathbb{R} = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ by setting

$$D_{\mathbb{H},\delta}(x,y,t) = (\delta x, \delta y, \delta^2 t).$$

Then $D_{\mathbb{H},\delta}\big[(x,y,t)\cdot(u,v,s)\big] = D_{\mathbb{H},\delta}[(x,y,t)]\cdot D_{\mathbb{H},\delta}[(u,v,s)]$, so that these dilations are group automorphisms. Thus in the Heisenberg group, the t variable has order 2, much as the time variable t has order 2 when studying the heat equation. The homogeneous dimension of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ under this family of dilations is $2n+2$.

We can define translation operators for the Heisenberg group in analogy with Euclidean translation, but since the group is not commutative, it is important to define the multiplication correctly. Thus define the operator of left translation by setting

$$\begin{aligned} L_{(u,v,s)}[f](x,y,t) &= f\big((u,v,s)^{-1}\cdot(x,y,t)\big) \\ &= f\big(x-u, y-v, t-s-2(< x,v > - < y,u >)\big). \end{aligned}$$

If P is an operator acting on functions on $\mathbb{H}^n$, P is *left-invariant* if $P\,L_{(u,v,s)} = L_{(u,v,s)}\,P$ for all $(u,v,s) \in \mathbb{H}_n$.

Define $(2n+1)$ special first order partial differential operators on $\mathbb{H}^n$ as follows:

$$\begin{cases} X_j = \frac{\partial}{\partial x_j} + 2y_j\,\frac{\partial}{\partial t} \\ \\ Y_j = \frac{\partial}{\partial y_j} - 2x_j\,\frac{\partial}{\partial t} \end{cases} \qquad \text{for } 1 \leq j \leq n, \text{ and} \qquad T = \frac{\partial}{\partial t}.$$

Then

$$\begin{aligned} L_{(u,v,s)}\,X_j[f](x,y,t) &= X_j[f]\big(x-u, y-v, t-s-2(< x,v > - < y,u >)\big) \\ &= \frac{\partial f}{\partial x_j}\big(x-u, y-v, t-s-2(< x,v > - < y,u >)\big) \\ &\quad + 2(y_j - v_j)\frac{\partial f}{\partial t}\big(x-u, y-v, t-s-2(< x,v > - < y,u >)\big) \end{aligned}$$

On the other hand, since $L_{(u,v,s)}[f](x,y,t) = f\big(x-u, y-v, t-s-2(< x,v > - < y,u >)\big)$, it follows that

$$
\begin{aligned}
X_j\, L_{(u,v,s)}[f](x,y,t)
&= \frac{\partial f}{\partial x_j}\big(x-u, y-v, t-s-2(<x,v>-<y,u>)\big)\\
&\quad - 2v_j \frac{\partial f}{\partial t}\big(x-u, y-v, t-s-2(<x,v>-<y,u>)\big)\\
&\qquad + 2y_j \frac{\partial f}{\partial t}\big(x-u, y-v, t-s-2(<x,v>-<y,u>)\big)\\
&= \frac{\partial f}{\partial x_j}\big(x-u, y-v, t-s-2(<x,v>-<y,u>)\big)\\
&\quad + 2(y_j - v_j)\frac{\partial f}{\partial t}\big(x-u, y-v, t-s-2(<x,v>-<y,u>)\big)\\
&= L_{(u,v,s)}\, X_j[f](x,y,t).
\end{aligned}
$$

Thus X_j is a left-invariant operator. A similar calculation shows that Y_j and T are also left-invariant.[3]

4.3. The sub-Laplacian and its geometry.

In the Euclidean situation, the first order partial derivatives $X_j = \frac{\partial}{\partial x_j}$ are all (Euclidean) translation invariant, and the Laplace operator is obtained by taking the sum of the squares of all n of these operators. In analogy, we now consider a family of second order operators $\mathcal{L}_\alpha$ on the Heisenberg group given by:

$$
\begin{aligned}
\mathcal{L}_\alpha &= \frac{1}{4}\sum_{j=1}^{n}\big(X_j^2 + Y_j^2\big) + i\alpha\, T\\
&= \frac{1}{4}\sum_{j=1}^{n}\Big[\frac{\partial^2}{\partial x_j^2} + \frac{\partial^2}{\partial y_j^2} + 2y_j\frac{\partial^2}{\partial x_j \partial t} - 2x_j\frac{\partial^2}{\partial y_j\partial t} + \big(x_j^2 + y_j^2\big)\frac{\partial^2}{\partial t^2}\Big] + i\alpha\,\frac{\partial}{\partial t}.
\end{aligned}
$$

When $\alpha = 0$, this is sometimes called the *sub-Laplacian*. These operators arise in studying the $\overline{\partial}_b$-complex[4] on the boundary of $\mathcal{U}_{n+1}$.

Except when $\alpha = \pm n, \pm(n+1), \ldots$, the operator $\mathcal{L}_\alpha$ has a fundamental solution K_α. The operator $\mathcal{L}_\alpha$ is a combination of the left-invariant operators $\{X_j, Y_j, T\}$, and one may hope that the operator $\mathcal{K}_\alpha$ which inverts $\mathcal{L}_\alpha$ also has this property. If so, this would mean that if

$$
\mathcal{K}_\alpha[f](z,t) = \int_{\mathbb{C}^n\times\mathbb{R}} K_\alpha\big((z,t),(w,s)\big)\, f(w,s)\, dw\, ds,
$$

---

[3] If we define right translation by

$$
R_{(u,v,s)}[f](x,y,t) = f\big((x,y,t)\cdot(u,v,s)^{-1}\big) = f\big(x-u, y-v, t-s+2(<x,v>-<y,u>)\big)
$$

and right-translation invariance in the natural way, then X_j and Y_j are *not* right-invariant. Instead, the corresponding right-invariant operators are

$$
\widetilde{X}_j = \frac{\partial}{\partial x_j} - 2y_j\frac{\partial}{\partial t}, \quad \widetilde{Y}_j = \frac{\partial}{\partial y_j} + 2x_j\frac{\partial}{\partial t}, \quad \text{and} \quad T = \frac{\partial}{\partial t}.
$$

[4] See [**Ste93**] for definitions and an extensive discussion of these matters. We will consider the $\overline{\partial}_b$-complex on more general domains in Chapter **??** below.

then we would have

$$\begin{aligned}
\mathcal{K}_\alpha[f](z,t) &= L_{(z,t)^{-1}}\,\mathcal{K}_\alpha[f](0,0)\\
&= \mathcal{K}_\alpha\, L_{(z,t)^{-1}}[f](0,0)\\
&= \int_{\mathbb{C}^n\times\mathbb{R}} K_\alpha\big((0,0),(w,s)\big)L_{(z,t)^{-1}}[f](w,s)\,dw\,ds\\
&= \int_{\mathbb{C}^n\times\mathbb{R}} K_\alpha\big((0,0),(w,s)\big)f\big((z,t)\cdot(w,s)\big)\,dw\,ds\\
&= \int_{\mathbb{C}^n\times\mathbb{R}} K_\alpha\big((0,0),(z,t)^{-1}\cdot(w,s)\big)f(w,s)\,dw\,ds\\
&= \int_{\mathbb{H}^n} f(w,s)\,k_\alpha\big((w,s)^{-1}\cdot(z,t)\big)\,ds\,ds\\
&= \int_{\mathbb{H}^n} f(w,s)\,L_{(w,s)}[k_\alpha](z,t)\,dw\,ds
\end{aligned}$$

where $k_\alpha(z,t) = K_\alpha\big((0,0),(z,t)^{-1}\big)$. But this is just the convolution of f with the function k_α on the Heisenberg group $\mathbb{H}^n$.

THEOREM 4.1 (Folland and Stein [**FS74**]). *Suppose that* $\alpha \neq \pm n, \pm(n+1), \ldots$. *Put*

$$k_\alpha(z,t) = \frac{2^{n-2}}{\pi^{n+1}}\,\Gamma\left(\frac{n+\alpha}{2}\right)\Gamma\left(\frac{n-\alpha}{2}\right)\big(|z|^2 - it\big)^{-\frac{n+\alpha}{2}}\big(|z|^2 + it\big)^{-\frac{n-\alpha}{2}}.$$

Then $\mathcal{K}_\alpha[f] = f * k_\alpha$ *is a fundamental solution for* $\mathcal{L}_\alpha$. *Explicitly, if* $\varphi \in \mathcal{C}_0^\infty(\mathbb{H}^n)$, *we have*

$$\begin{aligned}
\varphi(z,t) &= \mathcal{K}_\alpha\big[\mathcal{L}_\alpha[\varphi]\big](z,t)\\
\varphi(z,t) &= \mathcal{L}_\alpha\big[\mathcal{K}_\alpha[\varphi]\big](z,t),
\end{aligned}$$

We shall not prove this result, but we now show that there a distance $d_{\mathbb{H}}$ on $\mathbb{R}^3$ so that we can characterize the size of the fundamental solution K and its derivatives in terms of $d_{\mathbb{H}}$ and the volumes of the corresponding balls as we did for the Laplace operator and the heat operator in equations (2.6) - (3.4). Define a norm on $\mathbb{R}^3$ by setting

$$\big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}} = \big((x^2+y^2)^2 + t^2\big)^{\frac{1}{4}},$$

and let

$$\mathbb{B}_{\mathbb{H}}(0;\delta) = \Big\{(x,y,t)\in\mathbb{H}\,\Big|\,\big|\big|\,(x,y,t)\,\big|\big| < \delta\Big\}$$

be the corresponding family of balls at the origin $(0,0,0)$. We have

$$\big|\big|\,D_{\mathbb{H},\delta}[(x,y,t)]\,\big|\big|_{\mathbb{H}} = \delta\,\big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}}.$$

and consequently

$$\big|\mathbb{B}_H(0;\delta)\big| = \delta^4\,\big|\mathbb{B}(0;1)\big|.$$

Since $K_0(x,y,t)$ is homogeneous of degree -2 with respect to the dilations $\{D_{\mathbb{H},\delta}\}$ and is continuous and non-vanishing away from $(0,0,0)$, it follows that

$$K_0(x,y,t) \approx \big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}}^{-2} \approx \big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}}^{+2}\,\big|\mathbb{B}_{\mathbb{H}}\big(0;\big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}}\big)\big|^{-1},$$

where the symbol $\approx$ means that the ratio of the two sides is bounded and bounded away from 0 by constants which are independent of (x, y, t). Thus if we take the distance from the point $(0, 0, 0)$ to the point (x, y, t) to be

$$d_{\mathbb{H}}\big((x,y,t),(0,0,0)\big) = \big|\big|\,(x,y,t)\,\big|\big|_{\mathbb{H}} = \big((x^2+y^2)^2+t^2\big)^{\frac{1}{4}},$$

we have

$$\begin{aligned} K\big((x,y,t),(0,0,0)\big) &= K_0(x,y,t) \\ &\approx d_{\mathbb{H}}\big((x,y,t),(0,0,0)\big)^2 \left|\mathbb{B}_{\mathbb{H}}\big(0; d_{\mathbb{H}}\big((x,y,t),(0,0,0)\big)\big)\right|^{-1}, \end{aligned}$$

Moreover, $K\big((x,y,t),(u,v,s)\big) = K_0\big((u,v,s)^{-1}\cdot(x,y,t)\big)$. This suggests we should put

$$\begin{aligned} d_{\mathbb{H}}\big((x,y,t),(u,v,s)\big) &= d_{\mathbb{H}}\big((u,v,s)^{-1}\cdot(x,y,t),(0,0,0)\big) \\ &= \Big(\big((x-u)^2+(y-v)^2\big)^2 + \big(t-s-2(xv-yu)\big)^2\Big)^{\frac{1}{4}}, \end{aligned}$$

and the corresponding balls

$$\mathbb{B}_{\mathbb{H}}\big((x,y,t);\delta\big) = \left\{(u,v,s)\in\mathbb{R}^3 \,\middle|\, d_{\mathbb{H}}\big((x,y,t),(u,v,s)\big) < \delta\right\}.$$

One can now show that the function $d_{\mathbb{H}}$ has the following properties:

(1) $d_{\mathbb{H}}\big((x,y,t),(u,v,s)\big) \geq 0$ and $d_{\mathbb{H}}\big((x,y,t),(u,v,s)\big) = 0$ if and only if $(x,y,t) = (u,v,s)$.

(2) $d_{\mathbb{H}}\big((x,y,t),(u,v,s)\big) = d_{\mathbb{H}}\big((u,v,s),(x,y,t)\big)$.

(3) There is a constant $C \geq 1$ so that if $(x_j, y_j, t_j) \in \mathbb{R}^3$ for $1 \leq j \leq 3$ then

$$\begin{aligned} &d_{\mathbb{H}}\big((x_1,y_1,t_1),(x_3,y_3,t_3)\big) \\ &\qquad \leq C\left[d_{\mathbb{H}}\big((x_1,y_1,t_1),(x_2,y_2,t_2)\big) + d_{\mathbb{H}}\big((x_2,y_2,t_2),(x_3,y_3,t_3)\big)\right]. \end{aligned}$$

Note that the ball $\mathbb{B}_{\mathbb{H}}(0;\delta)$ is comparable to the set

$$\left\{(u,v,s)\,\big|\,|u|<\delta, \quad |v|<\delta, \quad |s|<\delta^2\right\},$$

and thus has the same non-isotropic nature as the ball $\mathbb{B}_H(0;\delta)$ we used for the heat equation. However, the ball $\mathbb{B}_{\mathbb{H}}\big((x,y,t);\delta\big)$ is the Heisenberg translate of the ball at the origin, not the Euclidean translate, to the point (x,y,t). Thus in addition to being non-isotropic, the ball $\mathbb{B}_{\mathbb{H}}\big((x,y,t);\delta\big)$ has a 'twist' as well. The ball is comparable to the set

$$\left\{(u,v,s)\,\big|\,|u-x|<\delta, \quad |v-y|<\delta, \quad |s-t+2(xv-yu)|<\delta^2\right\},$$

and thus has size δ in the u and v directions, and size δ^2 along the plane $s = t - 2(vx - uy)$.

We also have estimates for the fundamental solution K in terms of this geometry. Let us write $p = (x,y,t)$ and $q = (u,v,s)$. Then

$$K(p,q) \approx d_{\mathbb{H}}(p,q)^2 \left|\mathbb{B}_{\mathbb{H}}\big(p; d_{\mathbb{H}}(p,q)\big)\right|^{-1}, \tag{4.4}$$

which is the analogue of equations (2.6) and (3.3). We formulate estimates for derivatives of K not in terms of derivatives with respect to x, y, and t, but rather with respect to the operators X, Y, and T. Note that the operators X and Y do not commute, so the order in which the operators are applied makes a difference. Nevertheless, we have the following statement, which is the analogue of equations

(**??**) and (3.4). Let $P^\alpha(X,Y)$ be a non-commuting polynomial of order α in the operators X and Y, which we allow to act either on the $p=(x,y,t)$ or $q=(u,v,s)$ variables. Then there is a constant C_α so that

$$\big|P^\alpha(X,Y)K(p,q)\big| \leq C_\alpha\, d_{\mathbb{H}}(p,q)^{2-\alpha}\,\big|\mathbb{B}_{\mathbb{H}}\big(p;d_{\mathbb{H}}(p,q)\big)\big|^{-1}. \tag{4.5}$$

We have not explicitly indicated the effect of differentiation with respect to T. However, a key point is that T can be written in terms of X and Y. We have

$$XY - YX = -4T, \tag{4.6}$$

and so the action of T is the difference of two second order monomials in X and T. Thus we can formulate a more general statement about differentiation that follows from (4.5) and (4.6). Let $P^{\alpha,\beta}(X,Y,T)$ be a noncommuting polynomial of order α in X and Y, and of order β in T. These operators can act either on the $p=(x,y,t)$ or $q=(u,v,s)$ variables. Then there is a constant $C_{\alpha,\beta}$ so that

$$\big|P^{\alpha,\beta}(X,Y,T)K(p,q)\big| \leq C_{\alpha,\beta}\, d_{\mathbb{H}}(p,q)^{2-\alpha-2\beta}\,\big|\mathbb{B}_{\mathbb{H}}\big(p;d_{\mathbb{H}}(p,q)\big)\big|^{-1}. \tag{4.7}$$

Thus $d_{\mathbb{H}}$ is very much like a metric, but satisfies the weaker form of the triangle inequality given in (3). This suffices for most purposes, and we will eventually see that there is a true metric such that the metric balls are equivalent to the balls defined by $d_{\mathbb{H}}$.

4.4. The space $H^2(\mathbb{H}^n)$ and the Szegö Projection.

We now define the analogue of the classical Hardy space $H^2(\mathbb{D})$ in the unit disk. Consider the n complex first order partial differential operators on $\mathbb{H}^n$ given by

$$\bar{Z}_j = \frac{1}{2}\big[X_j + iY_j\big] = \frac{1}{2}\left[\frac{\partial}{\partial x_j} + i\frac{\partial}{\partial y_j}\right] - i(x_j+iy_j)\,\frac{\partial}{\partial t} = \frac{\partial}{\partial \bar{z}_j} - iz_j\,\frac{\partial}{\partial t},$$

for $1\leq j\leq n$, where we now write $z_j = x_j+iy_j$, $\bar{z}_j = x_j - iy_j$, and

$$\frac{\partial}{\partial z_j} = \frac{1}{2}\Big[\frac{\partial}{\partial x_j} - i\frac{\partial}{\partial y_j}\Big],$$
$$\frac{\partial}{\partial \bar{z}_j} = \frac{1}{2}\Big[\frac{\partial}{\partial x_j} + i\frac{\partial}{\partial y_j}\Big].$$

If $f\in L^2(\mathbb{H}^n)$, we say $\bar{Z}_j[f]=0$ if this equations holds in the distributional sense; *i.e.* for all $\varphi\in\mathcal{C}_0^\infty(\mathbb{H}^n)$ we have

$$\int_{\mathbb{H}^n} f(z,t)\,\overline{Z_j[\varphi](z,t)}\,dz\,dt = 0.$$

We set

$$H^2(\mathbb{H}^n) = \Big\{f\in L^2(\mathbb{H}^n)\,\Big|\,\overline{Z}_j[f]=0,\quad 1\leq j\leq n\Big\}. \tag{4.8}$$

It follows from this definition that $H^2(\mathbb{H}^n)$ is a closed subspace of $L^2(\mathbb{H}^n)$. Also one can check that if we set

$$f_\alpha(z,t) = \big(1+|z|^2+it\big)^{-\alpha} = \Big(1+\sum_{j=1}^n |z_j|^2 + it\Big)^{-\alpha},$$

then $f_\alpha\in H^2(\mathbb{H}^n)$ for all $\alpha > n+1$. Thus the space $H^2(\mathbb{H}^n)$ is non-zero, and in fact is infinite dimensional. The *Szegö projection* $\mathcal{S}: L^2(\mathbb{H}^n)\to H^2(\mathbb{H}^n)$ is the orthogonal projection of $L^2(\mathbb{H}^n)$ onto $H^2(\mathbb{H}^n)$. Our object in this section is to describe S as a singular integral operator.

For $\varphi \in \mathcal{S}(\mathbb{C}^n \times \mathbb{R})$, define the partial Fourier transform $\mathcal{F}$ in the t-variable by setting

$$\mathcal{F}[\varphi](z,\tau) = \widehat{\varphi}(z,\tau) = \int_{\mathbb{R}} e^{-2\pi i t\,\tau} \varphi(z,t)\,dt.$$

Then the inversion formula gives

$$\varphi(z,t) = \mathcal{F}^{-1}[\widehat{\varphi}](z,t) = \int_{\mathbb{R}} e^{+2\pi i t\,\tau} \widehat{\varphi}(z,\tau)\,d\tau.$$

The partial Fourier transform maps the Schwartz space $\mathcal{S}(\mathbb{C}^n \times \mathbb{R})$ to itself, and is one-to-one and onto. It follows from the Plancherel formula that it extends to an isometry of $L^2(\mathbb{C}^n \times \mathbb{R})$. Using this partial Fourier transform, we show that the space $H^2(\mathbb{C}^n \times \mathbb{R})$ can be identified with weighted spaces of holomorphic functions.

Let us put $\lambda(z,\tau) = e^{2\tau\,|z|^2}\,dz\,d\tau$, and define the operator

$$M[\psi](z,\tau) = e^{-\tau |z|^2}\,\psi(z,\tau).$$

Then $M\,\mathcal{F} : L^2(\mathbb{C}^n \times \mathbb{R}) \to L^2(\mathbb{C}^n \times \mathbb{R}; d\lambda)$ is an isometry. Let us set

$$B^2 = M\mathcal{F}\big[H^2(\mathbb{C}^n \times \mathbb{R}^n)\big].$$

Proposition 4.2. *A measurable function g on $\mathbb{C}^n \times \mathbb{R}$ belongs to B^2 if and only if*

(1) $\displaystyle\int_{\mathbb{C}^n \times \mathbb{R}} |g(z,\tau)|^2\, e^{2\tau|z|^2}\,dz\,d\tau = \big|\big|\,g\,\big|\big|^2 < +\infty.$

(2) *For almost every $\tau \in \mathbb{R}$, the function $z \to g(z,\tau)$ is an entire holomorphic function on $\mathbb{C}^n$.*

Proof. If $\varphi \in \mathcal{S}(\mathbb{C}^n \times \mathbb{R})$, we have

$$\overline{Z}_j[\varphi](z,t) = \int_{\mathbb{R}} e^{+2\pi i t\,\tau}\, e^{\tau|z|^2} \frac{\partial}{\partial \bar{z}_j}\Big[e^{-\tau|z|^2}\,\widehat{\varphi}\Big](z,\tau)\,d\tau.$$

It follows that

$$\overline{Z}_j = \mathcal{F}^{-1}\,M^{-1}\Big[\frac{\partial}{\partial \bar{z}_j}\Big]\,M\,\mathcal{F},$$

so the operator $\overline{Z}_j$ is conjugate to the operator $\frac{\partial}{\partial \bar{z}_j}$ acting on the space $L^2(\mathbb{C}^n \times \mathbb{R}; d\lambda)$. The proposition follows. □

4.5. Weighted spaces of entire functions.

For each $\tau \in \mathbb{R}$, let $L^2_\tau(\mathbb{C}^n) = L^2_\tau$ denote the space of (equivalence classes of) measurable functions $g : \mathbb{C}^n \to \mathbb{C}$ such that

$$\big|\big|\,g\,\big|\big|^2_\tau = \int_{\mathbb{C}^n} |g(z)|^2 e^{-2\tau|z|^2}\,dm(z) < +\infty.$$

L^2_τ is a Hilbert space, and we consider the subspace

$$B^2_\tau(\mathbb{C}^n) = B^2_\tau = \Big\{ g \in L^2_\tau(\mathbb{C}^n) \,\Big|\, g \text{ is holomorphic} \Big\}.$$

Using the mean value property of holomorphic functions, it follows that for any $z \in \mathbb{C}^n$ and any $g \in B^2_\tau$ we have

$$|g(z)| \leq \frac{2n}{\omega_{2n}} \int_{|z-w|<1} |g(w)|\, dm(w) \leq \frac{2n}{\omega_{2n}} \Big[\int_{|z-w|<1} e^{2\tau|w|^2}\, dm(w) \Big]^{\frac{1}{2}} \big|\big| \, g \, \big|\big|_\tau.$$

It follows that B^2_τ is a closed subspace of L^2_τ.

PROPOSITION 4.3. *If $\tau \leq 0$, the space $B^2_\tau = (0)$. If $\tau > 0$,*

(1) *Each monomial $z^\alpha \in B^2_\tau$, $\big|\big| \, z^\alpha \, \big|\big|^2_\tau = \pi^n \, \alpha! \, (2\tau)^{-|\alpha|-n}$, and $\big\langle z^\alpha, z^\beta \big\rangle_\tau = 0$ if $\alpha \neq \beta$. If*

$$c^2_\alpha = (\pi^n \, \alpha! \, (2\tau)^{-|\alpha|-n})^{-1},$$
$$\varphi_\alpha(z) = c_\alpha \, z^\alpha,$$

then $\{\varphi_\alpha\}$ is an orthonormal sequence in B^2_α.

(2) *If $g \in B^2_\tau$, then for every $z \in \mathbb{C}^n$*

$$g(z) = \sum_\alpha \big\langle g, \varphi_\alpha \big\rangle \, \varphi_\alpha(z)$$

where the series converges uniformly on compact subsets of $\mathbb{C}^n$ and the series $\sum_\alpha \big\langle g, \varphi_\alpha \big\rangle \, \varphi_\alpha$ converges to g in the Hilbert space B^2_α.

PROOF. Suppose $\tau \leq 0$ and $g \in B^2_\tau$. For $z_0 \in \mathbb{C}^n$, let $B(z_0, R)$ be the Euclidean ball centered at z_0 of radius R. Holomorphic functions are harmonic, and satisfy the mean value property on any ball. Using this and the Schwarz inequality we have

$$\begin{aligned} g(z_0) &= \big|B(z_0,R)\big|^{-1} \int_{B(z_0,R)} |f(w)|\, dm(w) \\ &\leq \big|B(z_0,R)\big|^{-1} \Big[\int_{B(z_0,R)} e^{+2\tau|w|^2}\, dm(w) \Big]^{\frac{1}{2}} \big|\big| \, g \, \big|\big|_\tau \\ &\leq \big|B(z_0,R)\big|^{-\frac{1}{2}} \big|\big| \, g \, \big|\big|_\tau. \end{aligned}$$

Letting $R \to \infty$, this shows that $g(z_0) = 0$. Thus $B^2_\tau = (0)$.

Now suppose that $\tau > 0$. Using polar coordinates we compute

$$\big|\big| \, z^\alpha \, \big|\big|^2_\tau = (2\pi)^n \prod_{j=1}^n \int_0^\infty r^{2\alpha_j+1} e^{-2\tau r^2}\, dr = \pi^n \, \alpha! \, (2\tau)^{-|\alpha|-n}.$$

Also, if we compute the inner product of two monomials z^α and z^β, then if $\alpha_j \neq \beta_j$ for some j, the integral in the θ_j-variable will vanish. Thus $\big\langle z^\alpha, z^\beta \big\rangle_\tau = 0$ if $\alpha \neq \beta$. If we put $c^2_\alpha = (\pi^n \, \alpha! \, (2\tau)^{-|\alpha|-n})^{-1}$ and $\varphi_\alpha(z) = c_\alpha \, z^\alpha$, it now follows that $\{\varphi_\alpha\}$ is an orthonormal sequence in B^2_τ.

Next, if $g \in B^2_\tau$, then g is an entire function, and hence the Taylor series for g converges absolutely and uniformly to g on any compact subset of $\mathbb{C}^n$. Write

$g(z) = \sum_\beta a_\beta \, z^\beta$. Then

$$\begin{aligned} c_\alpha \left\langle g, \varphi_\alpha \right\rangle_\tau &= \lim_{R\to\infty} c_\alpha \int_{|z|<R} g(z)\, \overline{\varphi_\alpha(z)}\, e^{-2\tau|z|^2}\, dm(z) \\ &= \lim_{R\to\infty} \sum_\beta a_\alpha \int_{|z|<R} \varphi_\beta(z)\, \overline{\varphi_\alpha(z)}\, dm(z) \\ &= \lim_{R\to\infty} a_\alpha \int_{|z|<R} |\varphi_\alpha(z)|^2 e^{-2\tau|z|^2}\, dm(z) \\ &= a_\alpha. \end{aligned}$$

Thus for any $z \in \mathbb{C}^n$ we have

$$g(z) = \sum_\alpha \left\langle g, \varphi_\alpha \right\rangle \varphi_\alpha(x).$$

Finally,

$$\begin{aligned} \left|\left| \, g \, \right|\right|_\tau^2 &= \lim_{\mathbb{R}\to\infty} \int_{|z|\le R} |g(z)|^2 e^{-2\tau|z|^2}\, dm(z) \\ &= \lim_{\mathbb{R}\to\infty} \sum_{\alpha,\beta} a_\alpha \, \overline{a_\beta} \int_{|z|\le R} z^\alpha \overline{z}^\beta \, e^{-2\tau|z|^2}\, dm(z) \\ &= \lim_{\mathbb{R}\to\infty} \sum_\alpha |a_\alpha|^2 \int_{|z|\le R} |z^\alpha|^2 e^{-2\tau|z|^2}\, dm(z) \\ &= \sum_\alpha |\left\langle g, \varphi_\alpha \right\rangle|^2. \end{aligned}$$

Thus $\displaystyle\sum_\alpha \left\langle g, \varphi_\alpha \right\rangle \varphi_\alpha$ converges to g in B_τ^2. □

LEMMA 4.4. *Let $P_\tau : L_\tau^2 \to B_\tau^2$ be the orthogonal projection. For $h \in L_\tau^2$,*

$$P_\tau[h](z) = \left(\frac{2}{\pi}\right)^n \tau^n \int_{\mathbb{C}^n} h(w)\, e^{2\tau<z,w>-2\tau|w|^2}\, dm(w).$$

PROOF. Since $\{\varphi_\alpha\}$ is a complete orthonormal sequence in B_τ^2, it follows that if $h \in L_\tau^2$ we have $P_\tau[h] = \sum_\alpha \left\langle h, \varphi_\alpha \right\rangle \varphi_\alpha$ with convergence in B_τ^2. Also, for any $z, w \in \mathbb{C}^n$ we have

$$\sum_\alpha \varphi_\alpha(z)\, \overline{\varphi_\alpha(w)} = \left(\frac{2\tau}{\pi}\right)^n \sum_\alpha \frac{1}{\alpha!}\, (2\tau)^\alpha z^\alpha \bar{w}^\alpha = \left(\frac{2\tau}{\pi}\right)^n e^{2\tau<z,w>}$$

with uniform convergence on compact subsets of $\mathbb{C}^n \times \mathbb{C}^n$. Thus for any $z \in \mathbb{C}^n$ we have

$$\begin{aligned} P_\tau[h](z) &= \sum_\alpha \left\langle h, \varphi_\alpha \right\rangle \varphi_\alpha(z) \\ &= \sum_\alpha \int_{\mathbb{C}^n} h(w)\, \varphi_\alpha(z)\, \overline{\varphi_\alpha(w)} e^{-2\tau|w|^2}\, dm(w) \\ &= \left(\frac{2}{\pi}\right)^n \tau^n \int_{\mathbb{C}^n} h(w)\, e^{2\tau<z,w>-2\tau|w|^2}\, dm(w), \end{aligned}$$

as asserted. □

Then he space $B^2_\tau = (0)$ for all $\tau \geq 0$. If $\tau < 0$, then $z^\alpha \in B^2_\tau$, and we have

$$||\, z^\alpha \,|| = \prod_{j=1}^{n} \int_{\mathbb{C}} |z_j|^{2\alpha_j}\, e^{\lambda |z_j|^2}\, dz_j = \pi^n\, \alpha!\, \lambda^{-|\alpha|-n}.$$

It follows that the orthogonal projection $P_\tau : L^2(\mathbb{C}^n, e^{2\tau|z|^2} dz) \to B^2_\tau$ is given by

$$P_\tau[g](z) = \frac{\tau^n}{\pi^n} \int_{\mathbb{C}^n} g(w)\, e^{-2\tau<z,w>+2\tau|w|^2}\, dw.$$

Now

$$\begin{aligned}
\mathcal{S}[f](z,t) &= \mathcal{F}^{-1}\, M^{-1}\, P\, M\, \mathcal{F}[f](z,t)\\
&= \int_{\mathbb{R}} e^{+2\pi i t\tau} M^{-1}\, P\, M\, \mathcal{F}[f](z,\tau)\, d\tau\\
&= \int_{\mathbb{R}} e^{+2\pi i t\tau} e^{\tau|z|^2}\, P\, M\, \mathcal{F}[f](z,\tau)\, d\tau\\
&= \int_{-\infty}^{0} e^{+2\pi i t\tau} e^{\tau|z|^2}\, P_\tau\big[(M\,\mathcal{F}[f])_\tau\big](z)\, d\tau\\
&= \frac{1}{\pi^n} \int_{-\infty}^{0} \int_{\mathbb{C}^n} e^{+2\pi i t\tau} e^{\tau|z|^2}\, e^{-2\tau<z,w>+2\tau|w|^2} M\mathcal{F}[f](w,\tau)\, dw\, \tau^n\, d\tau\\
&= \frac{1}{\pi^n} \int_{-\infty}^{0} \int_{\mathbb{C}^n} e^{+2\pi i t\tau} e^{\tau|z|^2}\, e^{-2\tau<z,w>+2\tau|w|^2} e^{-\tau|w|^2}\, \mathcal{F}[f](w,\tau)\, dw\, \tau^n\, d\tau\\
&= \frac{1}{\pi^n} \int_{-\infty}^{0} \int_{\mathbb{C}^n} \int_{\mathbb{R}} e^{+2\pi i t\tau}\, e^{-2\pi i s\tau}\, e^{\tau|z|^2-2\tau<z,w>+\tau|w|^2}\, f(w,s)\, ds\, dw\tau^n\, , d\tau\\
&= \iint_{\mathbb{C}^n\times\mathbb{R}} f(w,s)\, S\big((z,t),(w,s)\big)\, dw\, ds
\end{aligned}$$

where

$$\begin{aligned}
S\big((z,t),(w,s)\big) &= \frac{1}{\pi^n} \int_{-\infty}^{0} e^{2\pi i(t-s)\tau} e^{\tau[|z|^2-2<z,w>+|w|^2}\, \tau^n\, d\tau\\
&= \frac{(-1)^n}{\pi^n} \int_{0}^{\infty} e^{-2\pi\,\tau\big[|z-w|^2+i[t-s-2\Im m<z,w>]\big]}\, \tau^n\, d\tau\\
&= \frac{(-1)^n}{\pi^n} \big[|z-w|^2 + i[t-s-2\Im m<z,w>]\big]^{-n-1}\\
&= S\big((w,s)^{-1}\cdot(z,t)\big)
\end{aligned}$$

where

$$S(z,t) = \frac{(-1)^n}{\pi^n}\, \big[|z|^2 + it\big]^{-n-1},$$

5. The Grushin plane

As a final example, we study the geometry associated to the second order partial differential operator on $\mathbb{R}^2$ given by

$$\mathcal{L} = \frac{\partial^2}{\partial x^2} + x^2 \frac{\partial^2}{\partial t^2}. \tag{5.1}$$

This is an example of a class of non-elliptic, hypoelliptic operators studied by V.V. Grushin [**Gru71**].

5.1. The geometry. For $(x,t)\in\mathbb{R}^2$ and $\delta>0$, consider the set

$$\mathbb{B}_G\big((x,t);\delta\big)=\left\{(y,s)\in\mathbb{R}^2\,\middle|\,|x-y|<\delta,\qquad |t-s|<\delta^2+|x|\,\delta\right\}. \tag{5.2}$$

Note that for $|x|\le\delta$, this ball is essentially the non-isotropic (δ,δ^2) rectangle

$$\left\{(y,s)\in\mathbb{R}^2\,\middle|\,|x-y|<\delta,\qquad |t-s|<\delta^2\right\} \tag{5.3}$$

while for $|x|>\delta$, this is essentially

$$\left\{(y,s)\in\mathbb{R}^2\,\middle|\,|x-y|<\delta,\qquad |t-s|<|x|\,\delta\right\} \tag{5.4}$$

which is the translation to (x,t) of the standard Euclidean dilation by δ of the box $\{|y|<1,|s|<|x|\}$. If we set

$$d_G\big((x,t),(y,s)\big)=\max\left\{|x-y|,\min\left\{|t-s|^{\frac12},|x|^{-1}|t-s|\right\}\right\}, \tag{5.5}$$

then

$$\mathbb{B}_G\big((x,t);\delta\big)\approx\left\{(y,s)\in\mathbb{R}^2\,\middle|\,d_G\Big((x,t),(y,s)\Big)<\delta\right\}. \tag{5.6}$$

5.2. Fundamental solution for $\mathcal{L}$.

We begin by conjugating the operator $\mathcal{L}$ with the partial Fourier transform in the t-variable. Thus let us define

$$\mathcal{F}[f](x,\tau)=\int_{\mathbb{R}}e^{-2\pi it\tau}f(x,t)\,dt=\widehat{f}(x,\tau), \tag{5.7}$$

so that

$$f(x,t)=\int_{\mathbb{R}}e^{2\pi it\tau}\widehat{f}(x,\tau)\,d\tau=\mathcal{F}^{-1}[\widehat{f}]. \tag{5.8}$$

$\mathcal{F}$ is an isometry on $L^2(\mathbb{R}^2)$ and we have $\mathcal{L}=\mathcal{F}^{-1}\widehat{\mathcal{L}}\,\mathcal{F}$ where

$$\widehat{\mathcal{L}}[g](x,\tau)=\frac{\partial^2 g}{\partial x^2}(x,\tau)-4\pi^2\tau^2x^2g(x,\tau). \tag{5.9}$$

If $\widehat{\mathcal{K}}$ is an inverse for $\widehat{\mathcal{L}}$, then $\mathcal{K}=\mathcal{F}^{-1}\widehat{\mathcal{K}}\,\mathcal{F}$ is an inverse for $\mathcal{L}$. Put

$$g_\tau(x)=g(x,\tau) \tag{5.10}$$

$$\widehat{\mathcal{L}}_\tau[g]=\frac{d^2g}{\partial x^2}(x)-4\pi^2\tau^2x^2g(x), \tag{5.11}$$

so that

$$\widehat{\mathcal{L}}[g](x,\tau)=\widehat{\mathcal{L}}_\tau[g_\tau](x). \tag{5.12}$$

If $\widetilde{\mathcal{K}}_\tau$ is an inverse for $\widehat{\mathcal{L}}_\tau$ on $\mathcal{L}^2(\mathbb{R},dx)$, then $\widehat{\mathcal{K}}[g]=\widehat{\mathcal{K}}_\tau[g_\tau]$ is the desired inverse for $\widehat{\mathcal{L}}$. Thus we are reduced to finding an inverse for the ordinary differential operator

$$\widehat{\mathcal{L}}_\eta=\frac{\partial^2}{\partial x^2}-\eta^2x^2 \tag{5.13}$$

where we have set $2\pi\tau=\eta$.

5.3. Hermite functions.

The Hermite functions can be defined by

$$H_n(x) = \left(2^n n!\sqrt{\pi}\right)^{-\frac{1}{2}} (-1)^n \, e^{\frac{1}{2}x^2} \, \frac{d^n}{dx^n}\left(e^{-x^2}\right). \tag{5.14}$$

The collection of functions $\{H_n\}$ is a complete orthonormal basis for $L^2(\mathbb{R})$, and of particular significance for us, they are eigenfunctions of the operator $\widehat{\mathcal{L}}_1$ with eigenvalue $-(2n+1)$:

$$\frac{d^2 H_n}{dx^2}(x) - x^2 H_n(x) = -(2n+1) H_n(x). \tag{5.15}$$

We shall need the following two additional facts. The first is that H_n is essentially its own Fourier transform:

$$\int_{\mathbb{R}} e^{-2\pi i x\xi} H_n(\sqrt{2\pi}x)\, dx = (-i)^n \, H_n(\sqrt{2\pi}\xi). \tag{5.16}$$

The second is Mehler's formula: for $|z| < 1$ we have

$$\sum_{n=0}^{\infty} z^n H_n(x)\, H_n(y) = \left(\pi(1-z^2)\right)^{-\frac{1}{2}} \, \exp\left[\frac{4xyz - (x^2+y^2)(1+z^2)}{2(1-z^2)}\right]. \tag{5.17}$$

We can produce eigenfunctions for the opertor $\widehat{\mathcal{L}}_\eta$ with a simple change of variables. Put

$$H_n^\eta(x) = \eta^{\frac{1}{4}} \, H_n(\eta^{\frac{1}{2}} x). \tag{5.18}$$

Then $\{H_n^\eta\}$ is a complete orthonormal basis for $L^2(\mathbb{R})$, and we have

$$\widehat{\mathcal{L}}_\eta[H_n^\eta] = -\eta(2n+1)\, H_n^\eta. \tag{5.19}$$

If $g \in L^2(\mathbb{R})$ we can write $g = \sum_{n=0}^{\infty} (g, H_n^\eta)\, H_n^\eta$ where the sum converges in norm and

$$\big|\big|\, g \,\big|\big|_{L^2}^2 = \sum_{n=0}^{\infty} \left|(g, H_n^\eta)\right|^2. \tag{5.20}$$

Thus if we put

$$\widetilde{\mathcal{K}}_\eta[g] = -\eta^{-1} \sum_{n=0}^{\infty} (2n+1)^{-1} (g, H_n^\eta)\, H_n^\eta, \tag{5.21}$$

then $\widehat{\mathcal{K}}_\eta$ is a bounded operator on $L^2(\mathbb{R})$ which is a right inverse for $\widetilde{\mathcal{L}}_\eta$ on $L^2(\mathbb{R})$ and a left inverse on its natural domain, the subspace

$$D_{\widehat{\mathcal{L}}_\eta} = \left\{ g \in L^2(\mathbb{R}) \,\Big|\, \sum_{n=0}^{\infty} (2n+1)^2 \left|(g, H_n^\eta)\right|^2 < \infty \right\}. \tag{5.22}$$

But now, at least formally,

$$\widehat{\mathcal{K}}_\eta[g](x) = \int_{\mathbb{R}} \widetilde{K}_\eta(x,y)\, g(y)\, dy \tag{5.23}$$

where

$$\widehat{K}_\eta(x,y) = -\eta^{-1} \sum_{n=0}^{\infty} \frac{H_n^\eta(x)\, H_n^\eta(y)}{2n+1} = -\eta^{-\frac{1}{2}} \sum_{n=0}^{\infty} \frac{H_n(\eta^{\frac{1}{2}} x)\, H_n(\eta^{\frac{1}{2}} y)}{(2n+1)}. \tag{5.24}$$

Using Mehler's formula, we can write this as

$$\widehat{K}_\eta(x,y) = -\frac{1}{2}(\pi\eta)^{-\frac{1}{2}} \int_0^1 \exp\Big[\frac{4\eta xyr - \eta(x^2+y^2)(1+r^2)}{2(1-r^2)}\Big] \frac{dr}{\sqrt{r(1-r^2)}} \tag{5.25}$$

Now we compute

$$\begin{aligned}
\mathcal{K}[f](x,t) &= \mathcal{F}^{-1}\widehat{\mathcal{K}}\,\mathcal{F}[f](x,t) \\
&= \int_{\mathbb{R}} e^{2\pi i t\tau}\widehat{\mathcal{K}}\mathcal{F}[f](x,\tau)\,d\tau \\
&= \int_{\mathbb{R}} e^{2\pi i t\tau}\widehat{\mathcal{K}}_\tau[\mathcal{F}[f]_\tau](x)\,d\tau \\
&= \int_{\mathbb{R}} e^{2\pi i t\tau}\int_{\mathbb{R}} \widehat{K}_{2\pi\tau}(x,y)\,\mathcal{F}[f]_\tau(y)\,dy\,d\tau \\
&= \int_{\mathbb{R}} e^{2\pi i t\tau}\int_{\mathbb{R}} \widehat{K}_{2\pi\tau}(x,y)\,\mathcal{F}[f](y,\tau)\,dy\,d\tau \\
&= \iint_{\mathbb{R}^2} f(y,s)\Big[\int_{\mathbb{R}} e^{2\pi i\tau(t-s)}\widehat{K}_{2\pi\tau}(x,y)\,d\tau\Big]\,dy\,ds.
\end{aligned} \tag{5.26}$$

Thus the inverse to $\mathcal{L}$ has distribution kernel

$$K\big((x,t),(y,s)\big) = \int_{\mathbb{R}} e^{2\pi i\tau(t-s)}\widehat{K}_{2\pi\tau}(x,y)\,d\tau \tag{5.27}$$

which equals

$$-\frac{1}{\sqrt{4\pi}}\int_0^1\int_{\mathbb{R}} e^{2\pi i\tau(t-s)}\exp\Big[2\pi|\tau|\frac{4xyr-(x^2+y^2)(1+r^2)}{2(1-r^2)}\Big]\frac{d\tau}{|\tau|^{\frac{1}{2}}}\,\frac{dr}{\sqrt{r(1-r^2)}}. \tag{5.28}$$

Let

$$A(x,y,r) = -\frac{4xyr-(x^2+y^2)(1+r^2)}{2(1-r^2)} \tag{5.29}$$

Then the inner integral is equal to

$$\begin{aligned}
&\int_0^\infty e^{-2\pi\tau\left[A(x,y,r)+i(t-s)\right]}\frac{d\tau}{\tau^{\frac{1}{2}}} + \int_0^\infty e^{-2\pi\tau\left[A(x,y,r)-i(t-s)\right]}\frac{d\tau}{\tau^{\frac{1}{2}}} \\
&\quad= \frac{1}{\sqrt{2}}\big[A(x,y,r)+i(t-s)\big]^{-\frac{1}{2}} + \frac{1}{\sqrt{2}}\big[A(x,y,r)-i(t-s)\big]^{-\frac{1}{2}} \\
&\quad= \frac{\sqrt{1-r^2}}{\sqrt{(x^2+y^2)(1+r^2)-4xyr+2i(1-r^2)(t-s)}} \\
&\qquad\quad+ \frac{\sqrt{1-r^2}}{\sqrt{(x^2+y^2)(1+r^2)-4xyr-2i(1-r^2)(t-s)}}
\end{aligned} \tag{5.30}$$

Hence, up to a non-vanishing constant, we have

$$K\big((x,t),(y,s)\big) = K_+\big((x,t),(y,s)\big) + K_-\big((x,t),(y,s)\big), \tag{5.31}$$

where

$$K_\pm\big((x,t),(y,s)\big) = \int_0^1 \frac{1}{\sqrt{(x^2+y^2)(1+r^2)-4xyr\pm 2i(1-r^2)(t-s)}}\,\frac{dr}{\sqrt{r}}. \tag{5.32}$$

CHAPTER 2

VECTOR FIELDS

Vector fields can be thought of either as geometric or as analytic objects, and their importance is due at least in part to the interplay of these different points of view. The main topic of this book is the theory and applications of Carnot-Carathéodory metrics. While the metrics are geometric objects, the applications we have in mind are mainly to various analytic questions. The link between them is the concept of a vector field. The object of this chapter is to review the basic results about vector fields, commutators, and flows that will be used throughout this book.

Vector fields are often introduced geometrically as quantities, such as force, velocity or pressure, which have magnitude and direction, and which vary from point to point. There are many different ways of making this notion mathematically precise. Thus for example, a single vector is commonly represented in Euclidean space $\mathbb{R}^n$ as an n-tuple of real numbers. A vector field X on an open set $\Omega \subset \mathbb{R}^n$ should then assign an n-tuple of numbers to each point $x \in \Omega$. This of course is the same as prescribing an n-tuple of real-valued functions $X = (a_1, \ldots, a_n)$ defined on Ω; the vector assigned to x is then the n-tuple $X_x = \big(a_1(x), \ldots, a_n(x)\big)$.

Although quite straight forward, this approach leaves an important question unanswered. A quantity having magnitude and direction should have an intrinsic meaning, and its definition should not depend on a particular choice of coordinates. Thus if we adopt the simple-minded definition of the preceding paragraph, we still need to explain how the n-tuple of functions changes if we choose a different set of coordinates on Ω. Thus it is natural to look for a definition of a vector field that is coordinate free. This can be done by introducing the notion of the tangent space at each point $x \in \Omega$, and then thinking of a vector field as a (smooth) assignment of a tangent vector at each point. To proceed rigorously in this way one needs to introduce a fair amount of machinery from differential geometry, including the notions of tangent space and tangent bundle.

We shall proceed by following an intermediate route. In Section 1 we define tangent vectors as directional derivatives and vector fields as first order partial differential operators. Working with a fixed coordinate system on an open set in $\mathbb{R}^n$, we define and study the concept of commutators of two vector fields, and the determinant of a set of n vector fields. In Section 2 we see how these concepts behave under smooth mapping or smooth changes of variables. In Section 6 we explain what it means for a vector field to be tangent to a submanifold, and we review the Frobenius theorem. In Section 3 we define the flow along a vector field and the corresponding exponential map. We also begin the study of the Taylor series expansion of a function along a flow. In Section 7 we show that, locally, a single non-vanishing vector field always has a very simple form after a change of variables, and we consider the corresponding problem for two vector fields. Finally, in Section 8 we give a completely coordinate free characterization of vector fields by showing that they are the same as derivations. This allows one to study vector fields in more global situations such as on manifolds.

1. Vector fields and commutators

Throughout this section, Ω denotes an open subset of $\mathbb{R}^n$, and $\mathcal{E}(\Omega)$ denotes the the algebra of real-valued infinitely differentiable functions on Ω.

1.1. Germs of functions, tangent vectors, and vector fields.

It is convenient to define tangent vectors or directional derivatives at a point x as linear maps on certain spaces of functions which are defined near x. It does not really make sense to require that the functions be defined in a fixed open neighborhood U_1, since we could apply the directional derivative to a function defined on a smaller neighborhood $U_2 \subset U_1$. On the other hand, if two functions, defined on neighborhoods U_1 and U_2, agree in some smaller neighborhood $U_3 \subset U_1 \cap U_2$, the value of the directional derivative applied to each function will be the same. Thus the domain of the linear mapping should be some space of functions which involve only the values in arbitrarily small neighborhood of x and which regards two functions as equal if they agree in some small neighborhood. To make this precise, it is convenient to introduce the notion of a *germ* of a function at a point x. We begin by explaining this concept.

Let $x \in \mathbb{R}^n$, and consider pairs (U, f) where U is an open neighborhood of x and $f \in \mathcal{E}(U)$. Define an equivalence relation $\sim$ on the set of all such pairs by requiring that $(U_1, f_1) \sim (U_2, f_2)$ if and only if there is an open neighborhood U_3 of the point x such that $U_3 \subset U_1 \cap U_2$ and $f_1(y) = f_2(y)$ for all $y \in U_3$.

Definition 1.1. *A* germ *of an infinitely differentiable function at x is an equivalence class of pairs, and we denote the set of all germs at x by $\mathcal{E}_x$.*

If $f \in \mathcal{E}_x$ is a germ of a function, it is convenient to think of f in terms of one of its representatives: a real-valued infinitely differentiable function defined in some open neighborhood U of x. We can make $\mathcal{E}_x$ into an algebra in an obvious way. For example, if $f, g \in \mathcal{E}_x$, let (U, f) and (V, g) be pairs in the corresponding equivalence classes. We then define $f + g$ to be the equivalence class of the pair $(U \cap V, f + g)$. This is independent of the choice of representatives of f and g. Multiplication of germs is defined similarly.

If $f \in \mathcal{E}_x$, it makes sense to talk about the value of f or any of its derivatives at x, since these quantities only depend on values in an arbitrarily small neighborhood of x. It also makes sense to talk about the formal Taylor series of f at the point x. However, it does not make sense to talk about the value of f at any point other than x.

We can now define tangent vectors and vector fields.

Definition 1.2.

(1) *Let $x \in \mathbb{R}^n$. A tangent vector (or directional derivative) at x is a linear mapping $L : \mathcal{E}_x \to \mathbb{R}$ of the form*

$$L[f] = \sum_{j=1}^{n} c_j \frac{\partial f}{\partial x_j}(x) = \sum_{j=1}^{n} c_j\, \partial_{x_j}[f](x)$$

where $(c_1, \ldots, c_n) \in \mathbb{R}^n$. The space of all tangent vectors at x is denoted by T_x, and is called the tangent space at x.

(2) *If Ω is an open subset of $\mathbb{R}^n$, a smooth vector field on Ω is a first order partial differential operator*

$$X = \sum_{j=1}^{n} a_j \frac{\partial}{\partial x_j} = \sum_{j=1}^{n} a_j \, \partial_{x_j}$$

where each $a_j \in \mathcal{E}(\Omega)$. We denote the space of all smooth vector fields on Ω by $T(\Omega)$.

(3) *If $X = \sum_{j=1}^{n} a_j \, \partial_{x_j} \in T(\Omega)$ is a smooth vector field and $x \in \Omega$, then $X_x : T_x \to \mathbb{R}$ given by $X_x[f] = \sum_{j=1}^{n} a_j(x) \, \partial_{x_j}[f](x)$ is a tangent vector at x.*

Each vector field $X \in T(\Omega)$ induces a linear mapping of $\mathcal{E}(\Omega)$ to itself. If $f \in \mathcal{E}(\Omega)$ and $X = \sum_{j=1}^{n} a_j \, \partial_{x_J}$, then $X[f] \in \mathcal{E}(\Omega)$ is the infinitely differentiable function given by $X[f](x) = \sum_{j=1}^{n} a_j(x) \, \partial_{x_j}[f](x)$. If $X \in T(\Omega)$ and $x \in \Omega$, the product rule for derivatives shows that

$$X[fg] = X[f] \, g + f \, X[g] \qquad \text{if } f, g \in \mathcal{E}(\Omega), \tag{1.1}$$

$$X_x[fg] = X_x[f] \, g(x) + f(x) \, X_x[g] \qquad \text{if } f, g \in \mathcal{E}_{(\Omega)x}. \tag{1.2}$$

Linear mappings satisfying equations (2.4) and (1.2) are called *derivations*. We shall see below in section 8 that tangent vectors and vector fields can be characterized as linear maps which are derivations.

The space $T(\Omega)$ of smooth vector fields on Ω has the structure of a module over the ring $\mathcal{E}(\Omega)$ of smooth functions. Thus if $X = \sum_j a_j \, \partial_{x_j}$ and $Y = \sum_k b_k \, \partial_{x_k}$ are vector fields and $f, g \in \mathcal{E}(\Omega)$ we set

$$fX + gY = \sum_{j=1}^{n} (f \, a_j + g \, b_j) \, \partial_{x_j}.$$

1.2. Commutators.

In addition to its structure as a module over $\mathcal{E}(\Omega)$, the space $T(\Omega)$ of smooth vector fields on Ω also carries in a natural way the structure of a *Lie algebra* over $\mathbb{R}$. A real Lie algebra is a real vector space $\mathfrak{G}$ equipped with a distributive but non-associative product written $[x, y]$ for $x, y \in \mathfrak{G}$ so that

(1) $\alpha \, [x, y] = [\alpha x, y] = [x, \alpha y]$ for all $\alpha \in \mathbb{R}$ and all $x, y \in \mathfrak{G}$;

(2) $[x + y, z] = [x, z] + [y, z]$ and $[x, y + z] = [x, y] + [x, z]$ for all $x, y, z \in \mathfrak{G}$;

(3) $[x, y] = -[y, x]$ for all $x, y \in \mathfrak{G}$;

(4) $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$ for all $x, y, z \in \mathfrak{G}$.

Property (3) expresses the anti-symmetry of the product, and property (4) is called the *Jacobi identity*. Additional information about Lie algebras can be found in Chapter 7.

To define a Lie algebra structure on $T(\Omega)$, let $X = \sum_j a_j \, \partial_{x_j}$ and $Y = \sum_k b_k \, \partial_{x_k}$ be smooth vector fields on Ω. If $f \in \mathcal{E}(\Omega)$, we can apply X to the smooth function $Y[f]$ or Y to the smooth function $X[f]$. We get smooth functions $XY[f]$ and $YX[f]$ where

$$XY[f](x) = \sum_{j,k=1}^{n} a_j(x)\, b_k(x)\, \frac{\partial^2 f}{\partial x_j \partial x_k}(x) + \sum_{k=1}^{n} \Big[\sum_{j=1}^{n} a_j(x) \frac{\partial b_k}{\partial x_j}(x) \Big] \frac{\partial f}{\partial x_k},$$

$$YX[f](x) = \sum_{j,k=1}^{n} a_j(x)\, b_k(x)\, \frac{\partial^2 f}{\partial x_k \partial x_j}(x) + \sum_{j=1}^{n} \Big[\sum_{k=1}^{n} b_k(x) \frac{\partial a_j}{\partial x_k}(x) \Big] \frac{\partial f}{\partial x_j}$$

Thus XY and YX are second order partial differential operators. Notice moreover that the second order terms in XY are the same as the second order terms in YX since $\partial_{x_j}\partial_{x_k}[f] = \partial_{x_k}\partial_{x_j} f$. Thus $XY - YX$, although formally a second order operator, is actually another first order partial differential operator, and is thus another vector field, called the *commutator* of X and Y.

Definition 1.3. *If $X = \sum_j a_j\, \partial_{x_j}$ and $Y = \sum_k b_k\, \partial_{x_k}$ are two smooth vector fields on Ω, their commutator is the vector field $[X,Y] = XY - YX$ which is given by*

$$[X,Y] = \sum_{j=1}^{n} \left[\sum_{k=1}^{n} \left(a_k \frac{\partial b_j}{\partial x_k} - b_k \frac{\partial a_j}{\partial x_k} \right) \right] \frac{\partial}{\partial x_j}. \tag{1.3}$$

The commutator $[X,Y]$ clearly measures the failure of the two linear mappings $X, Y : \mathcal{E}(\Omega) \to \mathcal{E}(\Omega)$ to commute. This is an algebraic interpretation of $[X,Y]$. We will see later that the commutator also has an important intrinsic geometric meaning as well.

The next proposition records some easily verified properties of commutators. Properties (1), (2) and (3) below express the fact that $T(\Omega)$ forms a Lie algebra under the bracket operation.

Proposition 1.4. *Let X, Y, Z be vector fields on Ω, and let $f, g \in \mathcal{E}(\Omega)$. Then:*

(1) $[X+Y, Z] = [X,Z] + [Y,Z]$;
(2) $[X,Y] + [Y,X] = 0$;
(3) $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$;
(4) $[fX, gY] = fg[X,Y] + f\,X[g]\,Y - g\,Y[f]\,X$.

1.3. Vector fields of finite type.

Given a Lie algebra $\mathfrak{G}$, a *Lie subalgebra* is a linear subspace $\mathfrak{H} \subset \mathfrak{G}$ which is closed under the Lie algebra product. We shall be particularly interested in Lie subalgebras and Lie submodules of $T(\Omega)$ generated by a finite number of vector fields.

Definition 1.5. *Let $X_1, \ldots, X_p \in T(\Omega)$.*

(i) The Lie subalgebra generated by $\{X_1, \ldots, X_p\}$ is the smallest vector subspace of $T(\Omega)$ containing $\{X_1, \ldots, X_p\}$ which is closed under the bracket operation. We denote it by $\mathcal{L}(X_1, \ldots, X_p)$.

(ii) The Lie submodule generated by $\{X_1, \ldots, X_p\}$ is the smallest smallest $\mathcal{E}(\Omega)$ submodule of $T(\Omega)$ containing the vectors $\{X_1, \ldots, X_p\}$ which is closed under the bracket product. We denote this by $\mathcal{L}_{\mathcal{E}(\Omega)}(X_1, \ldots, X_p)$.

In addition to being closed under the bracket operation, $\mathcal{L}_{\mathcal{E}(\Omega)}(X_1, \dots, X_p)$ is closed under multiplication by functions in $\mathcal{E}(\Omega)$. It follows from conclusion (4) in Proposition 1.4 that $\mathcal{L}(X_1, \dots, X_p)$ is just the $\mathcal{E}(\Omega)$ submodule of $T(\Omega)$ generated by $\mathcal{L}(X_1, \dots, X_p)$.

The elements of $\mathcal{L}(X_1, \dots, X_p)$ span a subspace of the tangent space $T_x(\Omega)$ for each $x \in \Omega$. Thus set

$$\mathcal{L}(X_1, \dots, X_p)_x = \left\{ L \in T_x(\Omega) \,\middle|\, \big(\exists Y \in \mathcal{L}(X_1, \dots, X_p)\big)\big(L = Y_x\big) \right\}.$$

One could also consider the corresponding subspace $\mathcal{L}_{\mathcal{E}(\Omega)}(X_1, \dots, X_p)_x$, but it is easy to see from Proposition 1.4 that $\mathcal{L}_{\mathcal{E}(\Omega)}(X_1, \dots, X_p)_x = \mathcal{L}(X_1, \dots, X_p)_x$. We shall be particularly interested in cases where $\mathcal{L}(X_1, \dots, X_p)_x = T_x(\Omega)$ for all $x \in \Omega$, since in this case we can construct a metric from the vector fields $\{X_1, \dots, X_p\}$.

Definition 1.6. *The vector fields* $\{X_1, \dots, X_p\} \subset T(\Omega)$ *are of finite type if* $\mathcal{L}(X_1, \dots, X_p)_x = T_x(\Omega)$ *for every* $x \in \Omega$.

It is important to know which vector fields belong to $\mathcal{L}(X_1, \dots, X_p)$. However, since the bracket product is non-associative, even the enumeration of all Lie products of elements $\{X_1, \dots, X_p\}$ can be quite complicated. For example

$$\begin{array}{lll} [X, [Y, [Z, W]]] & [[X, Y], [Z, W]] & [[X, [Y, Z]], W] \\ [X, [[Y, Z], W]] & & [[[X, Y], Z], W] \end{array}$$

are five distinct possible products of the four elements $\{X, Y, Z, W\}$. Some simplification is possible. It follows from the Jacobi identity, and is proved in Chapter **??**, Corollary **??**, that every element of $\mathcal{L}(X_1, \dots, X_p)$ can be written as a linear combination with real coefficients of *iterated commutators*, which are Lie products of the elements $\{X_1, \dots, X_p\}$ having the special form $[X_{i_k}, [X_{i_{k-1}}, \cdots [X_{i_3}, [X_{i_2} X_{i_1}]] \cdots]]$. However the Jacobi identity also shows that not all of these iterated commutators are linearly independent.

Let us illustrate the concept of finite type with two examples. We shall work in $\mathbb{R}^3$ where we denote the coordinates by (x, y, z).

Example 1: Consider the three vector fields in $T(\mathbb{R}^3)$ given by

$$A = x\,\partial_y - y\,\partial_x, \quad B = y\,\partial_z - z\,\partial_y, \quad C = x\,\partial_z - z\,\partial_x.$$

We then have

$$[A, B] = C, \qquad [B, C] = A, \qquad [C, A] = B.$$

The vector fields $\{A, B, C\}$ are linearly independent over $\mathbb{R}$, so $\mathcal{L}(A, B, C)$ is the three dimensional subspace of $T(\mathbb{R})$ spanned by A, B, and C. In fact, $\mathcal{L}(A, B, C)$ is isomorphic as a Lie algebra to $\mathbb{R}^3$ with the usual cross product, and this is the same as the Lie algebra of the orthogonal group $O(3)$.

However since all the coefficients vanish at the origin we have $\mathcal{L}(A, B, C)_{(0,0,0)} = (0)$, and if $p = (x_0, y_0, z_0) \neq (0, 0, 0)$, it is easy to check that $\mathcal{L}(A, B, C)_p$ is the two dimensional subspace of the tangent space $T_p(\mathbb{R}^3)$ which is perpendicular to the vector (x_0, y_0, z_0). Thus the vectors $\{A, B, C\}$ are not of finite type.

Example 2: Consider the two vector fields in $T(\mathbb{R}^3)$ given by

$$X = \partial_x - \frac{1}{2} y\,\partial_z, \qquad Y = \partial_y + \frac{1}{2} x\,\partial_z.$$

If we set $Z = \partial_z$, then we have $[X, Y] = Z$, while $[X, Z] = 0$ and $[Y, Z] = 0$. The vector fields $\{X, Y, Z\}$ are linearly independent over $\mathbb{R}$, so $\mathcal{L}(X, Y)$ is the three dimensional subspace of $T(\mathbb{R}^3)$ spanned by X, Y, and Z. Moreover, for every $p \in \mathbb{R}^3$ the three tangent vectors X_p, Y_p and Z_p are linearly independent, and so $\mathcal{L}(X, Y)_p = T_p(\mathbb{R}^3)$. Thus the two vector fields $\{X, Y\}$ are of finite type on $\mathbb{R}^3$.

This is the first appearance of the Lie algebra of the Heisenberg group $\mathbb{H}_1$, and we shall frequently return to this example.

1.4. Derivatives of determinants.

Definition 1.7. *Let $X_1, \ldots, X_n \in T(\Omega)$ with $X_j = \sum_{k=1}^n a_{j,k}\partial_{x_k}$. The determinant is the scalar function*

$$\det(X_1, \ldots, X_n)(x) = \det\{a_{j,k}(x)\}.$$

The determinant gives the (signed) volume of the parallelopiped spanned by the vectors $\{X_1(x), \ldots, X_n(x)\}$. In particular the tangent vectors $\{(X_1)_x, \ldots, (X_n)_x\}$ span $T_x(\Omega)$ if and only if $\det(X_1, \ldots, X_n)(x) \neq 0$. We shall need a formula for the derivative of the scalar function $\det(X_1, \ldots, X_n)(x)$, and for this we need one additional definition.

Definition 1.8. *If $X = \sum_{k=1}^N a_k\, \partial_{x_k} \in T(\Omega)$, the divergence of X is the scalar function*

$$(\nabla\cdot X)(x) = \sum_{k=1}^n \frac{\partial a_k}{\partial x_k}(x).$$

We can now state the promised formula.

Lemma 1.9. *Let $X_1, \ldots, X_n, T \in T(\Omega)$. Then*

$$\begin{aligned} T\big(\det(X_1, \ldots, X_n)\big) = \sum_{k=1}^n &\det(X_1, \ldots, X_{k-1}, [T,\, X_k]\,, X_{k+1}, \ldots, X_n) \\ &+ \big(\nabla\cdot T\big)\, \det(X_1, \ldots, X_n). \end{aligned} \tag{1.4}$$

Proof. Let $X_j = \sum_k a_{j,k}\, \partial_{x_k}$ and $T = \sum_l b_l\, \partial_{x_k}$. Then

$$\begin{aligned} [T,\, X_j] &= \sum_{l=1}^n \sum_{k=1}^n \left[b_l \frac{\partial}{\partial x_l},\, a_{j,k} \frac{\partial}{\partial x_k} \right] \\ &= \sum_{k=1}^n \Big(\sum_{l=1}^n b_l \frac{\partial a_{j,k}}{\partial x_l} \Big) \frac{\partial}{\partial x_k} - \sum_{k=1}^n \Big(\sum_{l=1}^n a_{j,l} \frac{\partial b_k}{\partial x_l} \Big) \frac{\partial}{\partial x_k} \\ &= \sum_{k=1}^n T[a_{j,k}] \frac{\partial}{\partial x_k} - \sum_{k=1}^n \Big(\sum_{l=1}^n a_{j,l} \frac{\partial b_k}{\partial x_l} \Big) \frac{\partial}{\partial x_k}. \end{aligned}$$

Hence

$$\begin{aligned}
&\det(X_1, \ldots, X_{j-1}, [T, X_j], X_{j+1}, \ldots, X_n) \\
&\qquad = \det(X_1, \ldots, X_{j-1}, \sum_{k=1}^{n} T[a_{j,k}] \frac{\partial}{\partial x_k}, X_{j+1}, \ldots, X_n) \\
&\qquad\quad - \det(X_1, \ldots, X_{j-1}, \sum_{k=1}^{n} a_{j,k} \frac{\partial b_k}{\partial x_k} \frac{\partial}{\partial x_k}, X_{j+1}, \ldots, X_n) \\
&\qquad\quad - \det(X_1, \ldots, X_{j-1}, \sum_{k=1}^{n} \Big(\sum_{\substack{l=1 \\ l \neq k}}^{n} a_{j,k} \frac{\partial b_k}{\partial x_l} \Big) \frac{\partial}{\partial x_k}, X_{j+1}, \ldots, X_n) \\
&\qquad = A_j - B_j - C_j.
\end{aligned}$$

Now using the product rule we see that

$$T\big(\det(X_1, \ldots, X_n)\big) = \sum_{j=1}^{n} A_j.$$

Expanding B_j by minors of the j^{th} entry, we have

$$B_j = \sum_{k=1}^{n} (-1)^{j+k} a_{j,k} \frac{\partial b_k}{\partial x_k} M_{j,k}$$

where $M_{j,k}$ is the $(j,k)^{th}$ minor of the matrix $\{a_{r,s}\}$. Hence

$$\sum_{j=1}^{n} B_j = \sum_{k=1}^{n} \frac{\partial b_k}{\partial x_k} \sum_{j=1}^{n} (-1)^{j+k} a_{j,k} M_{j,k} = \big(\nabla \cdot T\big) \det(X_1, \ldots, X_n).$$

Similarly

$$\begin{aligned}
\sum_{j=1}^{n} C_j &= \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} \frac{\partial b_k}{\partial x_l} \sum_{j=1}^{n} (-1)^{j+k} a_{j,l} M_{j,k} \\
&= \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} \frac{\partial b_k}{\partial x_l} \det(X_1, \ldots, X_{k-1}, X_l, X_{k+1}, \ldots, X_n) \\
&= 0
\end{aligned}$$

since each determinant has a repeated row. This completes the proof. □

2. Vector fields and smooth mappings

In this section $\Omega_j \subset \mathbb{R}^{n_j}$, $j = 1, 2$, are open sets, and $\Phi : \Omega_1 \to \Omega_2$ is a smooth mapping. This means that there are functions $\{\varphi_1, \ldots \varphi_{n_2}\} \subset \mathcal{E}(\Omega_1)$ so that for $x \in \Omega_1$,

$$\Phi(x) = \big(\varphi_1(x), \ldots, \varphi_{n_2}(x)\big).$$

We shall sometimes assume that Φ is a diffeomorphism, which means that $n_1 = n_2 = n$, that the mapping Φ is one-to-one and onto, and that the inverse mapping $\Phi^{-1} : \Omega_2 \to \Omega_1$ is also a smooth mapping. In this case, if we denote the coordinates in Ω_1 by $x = (x_1, \ldots, x_n)$ and the coordinates in Ω_2 by $y = (y_1, \ldots, y_n)$, then Φ can also be thought of as a change of variables $y = \Phi(x)$.

2.1. The differential acting on tangent vectors, vector fields, and commutators.

If $x \in \Omega_1$, every smooth mapping $\Phi : \Omega_1 \to \Omega_2$ induces a linear mapping $d\Phi_x : T_x \to T_{\Phi(x)}$ called the *differential* of Φ at x.

DEFINITION 2.1. *Let $\Phi : \Omega_1 \to \Omega_2$ be a smooth mapping, let $x \in \Omega_1$, let $L \in T_x$, and let $f \in \mathcal{E}_{\Phi(x)}$. Then $f \circ \Phi \in \mathcal{E}_x$ and we define*

$$d\Phi_x[L][f] = L\big[f \circ \Phi\big].$$

If $L[g] = \sum_{j=1}^{n_1} b_j\, \partial_{x_j}[g](x)$ for $g \in \mathcal{E}(\Omega_1)$, the chain rule shows that

$$d\Phi_x[L][f](x) = \sum_{k=1}^{n_2} \Big[\sum_{j=1}^{n_1} b_j\, \frac{\partial \varphi_k}{\partial x_j}(x)\Big]\, \frac{\partial f}{\partial y_k}\big(\Phi(x)\big). \tag{2.1}$$

This shows that $d\Phi_x[L] \in T_{\Phi(x)}$. Indeed, equation (2.1) shows that if we identify T_x with $\mathbb{R}^{n_1}$ and $T_{\Phi(x)}$ with $\mathbb{R}^{n_2}$ via the correspondences

$$\begin{aligned} \mathbb{R}^{n_1} \ni (b_1, \ldots, b_{n_1}) &\longleftrightarrow \sum_{j=1}^{n_1} b_j\, \partial_{x_j} \in T_x \\ \mathbb{R}^{n_2} \ni (c_1, \ldots, c_{n_2}) &\longleftrightarrow \sum_{k=1}^{n_2} c_k\, \partial_{y_k} \in T_{\Phi(x)}, \end{aligned}$$

then $d\Phi_x$ has the representation as the $n_1 \times n_2$ matrix

$$d\Phi_x = \begin{bmatrix} \frac{\partial \varphi_1}{\partial x_1} & \cdots & \frac{\partial \varphi_{n_2}}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial \varphi_1}{\partial x_{n_1}} & \cdots & \frac{\partial \varphi_{n_2}}{\partial x_{n_1}} \end{bmatrix}.$$

When $\Phi : \Omega_1 \to \Omega_2$ is a diffeomorphism, more is true. The mapping Φ induces a linear mapping $d\Phi : T(\Omega_1) \to T(\Omega_2)$.

DEFINITION 2.2. *Let $\Phi : \Omega_1 \to \Omega_2$ be a diffeomorphism. Let $X \in T(\Omega_1)$ and let $f \in \mathcal{E}(\Omega_2)$. Put*

$$d\Phi[X][f] = X\big[f \circ \Phi\big] \circ \Phi^{-1}.$$

In fact, let $X = \sum_{j=1}^{n_1} a_j\, \partial_{x_j}$ where $a_j \in \mathcal{E}(\Omega_1)$. If we use equation (2.1) and let $y = \Phi(x)$ we have

$$\begin{aligned} d\Phi[X][f](y) &= \sum_{k=1}^{n_2} \Big[\sum_{j=1}^{n_1} a_j\big(\Phi^{-1}(y)\big)\, \frac{\partial \varphi_k}{\partial x_j}\big(\Phi^{-1}(y)\big)\Big]\, \frac{\partial f}{\partial y_k}(y) \\ &= \sum_{k=1}^{n_2} X[\varphi_k]\big(\Phi^{-1}(y)\big)\, \frac{\partial f}{\partial y_k}(y), \end{aligned} \tag{2.2}$$

and this shows that $d\Phi[X]$ is indeed a vector field on Ω_2.

The next proposition shows that commutators behave correctly under changes of coordinates.

PROPOSITION 2.3. *Let $\Phi : \Omega_1 \to \Omega_2$ be a diffeomorphism. If $X, Y \in T(\Omega_1)$, then*

$$d\Phi\big[[X,Y]\big] = \big[d\Phi[X], d\Phi[Y]\big].$$

PROOF. Let $f \in \mathcal{E}(\Omega_2)$. Then $Y\big[f \circ \Phi\big] = d\Phi[Y][f] \circ \Phi$, and so

$$X\Big[Y\big[f \circ \Phi\big]\Big] = X\big[d\Phi[Y][f] \circ \Phi\big] = d\Phi[X]\big[d\Phi[Y][f]\big] \circ \Phi.$$

Similarly, we have

$$Y\Big[X\big[f \circ \Phi\big]\Big] = d\Phi[Y]\big[d\Phi[X][f]\big] \circ \Phi.$$

Subtracting the second equation from the first shows that

$$[X,Y]\big[f \circ \Phi\big] = \big[\Phi_*[X], \Phi_*[Y]\big][f] \circ \Phi,$$

and it follows that $d\Phi\big[[X,Y]\big] = \big[d\Phi[X], d\Phi[Y]\big]$. □

2.2. Action on determinants.

Suppose again that $\Phi : \Omega_1 \to \Omega_2 \subset \mathbb{R}^n$ is a diffeomorphism, and that $\Phi(x) = \big(\varphi_1(x), \ldots, \varphi_n(x)\big)$. The Jacobian determinant of Φ is

$$J\Phi(x) = \det \begin{bmatrix} \frac{\partial \varphi_1}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_n}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_n}{\partial x_n}(x) \end{bmatrix} \tag{2.3}$$

PROPOSITION 2.4. *If $X_1, \ldots, X_n \in T(\Omega_1)$ then*

$$\det\big(d\Phi[X_1], \ldots, d\Phi[X_n]\big)\big(\Phi(x)\big) = J\Phi(x)\,\det(X_1, \ldots, X_n)(x).$$

PROOF. Suppose that $X_l = \sum_{j=1}^n a_{j,l}\,\partial_{x_j}$ and that $d\Phi[X_m] = \sum_{k=1}^n b_{k,m}\,\partial_{y_m}$. Then equation (2.1) shows that

$$\begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix} = \begin{bmatrix} \frac{\partial \varphi_1}{\partial x_1} & \cdots & \frac{\partial \varphi_{n_2}}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial \varphi_1}{\partial x_{n_1}} & \cdots & \frac{\partial \varphi_{n_2}}{\partial x_{n_1}} \end{bmatrix} \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

where the right hand side is the product of two matrices. The proposition follows by taking determinants. □

3. Integral curves and the exponential maps

In this section, Ω is an open subset of $\mathbb{R}^n$ and $X \in T(\Omega)$ is a smooth real vector field on Ω.

3.1. Integral curves.

Returning for a moment to the geometric interpretation of a vector field, we can think of a vector field $X \in T(\Omega)$ as prescribing a velocity X_x at each point $x \in \Omega$. We can then ask for parametric equations $t \to \psi(t) \in \Omega$ of particles whose velocity at time t is $X_{\psi(t)}$. The orbits of such particles are called integral curves of the vector field X.

To make this question precise, we must identify the velocity of a particle with a tangent vector. We do this as follows. If $\psi = (\psi_1, \ldots, \psi_n) : \mathbb{R} \to \mathbb{R}^n$ parameterizes a curve in $\mathbb{R}^n$, then the usual definition of the derivative (*i.e.* velocity) at time t is $\psi'(t) = \big(\psi_1'(t), \ldots, \psi_n'(t)\big)$. On the other hand, let $T = \frac{\partial}{\partial t}$ be the standard vector field on $\mathbb{R}$. Then using Definition 2.2 and equation (2.2) we see that

$$d\psi[T]_{\psi(t)} = \sum_{k=1}^{n} \psi_k'(t) \frac{\partial}{\partial x_k}$$

is a tangent vector at the point $\psi(t)$. Thus it is natural to identify the velocity at time t of a particle whose orbit is parameterized by the mapping ψ with the tangent vector $d\psi[T]_{\psi(t)}$, and we write

$$\psi'(t) = d\psi[T]_{\psi(t)}.$$

DEFINITION 3.1. *Let X be a vector field on $\Omega \subset \mathbb{R}^n$. An* integral curve *for the vector field X is a $\mathcal{C}^1$ mapping $\psi = (\psi_1, \ldots, \psi_n)$ from an interval $(a, b) \subset \mathbb{R}$ to Ω such that for $t \in (a, b)$*

$$\psi'(t) = d\psi[T]_{\psi(t)} = X_{\psi(t)}.$$

If $X = \sum_{j=1}^{n} a_j \, \partial_{x_j}$, the equation $\psi'(t) = X_{\psi(t)}$ is equivalent to the system of ordinary differential equations

$$\begin{aligned} \psi_1'(t) &= a_1\big(\psi_1(t), \ldots, \psi_n(t)\big), \\ &\vdots \\ \psi_n'(t) &= a_n\big(\psi_1(t), \ldots, \psi_n(t)\big). \end{aligned} \tag{3.1}$$

In addition to these equations, we can also require that the solution $\psi(t)$ pass through a given point $x \in \Omega$. The following theorem provides the basic results about existence and uniqueness of such systems.

THEOREM 3.2. *Let $X = \sum_{j=1}^{n} a_j \, \partial_{x_j} \in T(\Omega)$ and let $K \subset \Omega$ be compact. There exists $\epsilon > 0$ and a $\mathcal{C}^\infty$ mapping $E_X = E : (-\epsilon, +\epsilon) \times K \to \Omega$ with the following properties:*

1. *For each $x \in K$, the mapping $t \to E(t, x)$ is an integral curve for X passing through x when $t = 0$. Thus*
$$\partial_t E(t, x) = X_{E(t,x)};$$
$$E(t, 0) = x.$$

2. *This solution is unique. If $\varphi : (-\eta, +\eta) \to \Omega$ is a $\mathcal{C}^1$ mapping such that $\varphi'(t) = X_{\varphi(t)}$ and $\varphi(0) = x$, then $\varphi(t) = E(t, x)$ for $|t| < \min(\epsilon, \eta)$.*

(3) The constant ϵ can be chosen to depend only the supremum of the first derivatives of the coefficients $\{a_j\}$ of X on Ω and on the distance from K to the complement of Ω.

(4) For $(t, x) \in (-\epsilon, +\epsilon) \times K$, we can estimate $\left|\partial_x^\alpha \partial_t^\beta E(t,x)\right|$ in terms of the supremum of finitely many derivatives of the coefficients $\{a_j\}$ on the compact set Ω.

(5) If the coefficients $\{a_j\}$ of the vector field X depend smoothly on additional parameters $\{\lambda_1, \ldots, \lambda_m\}$ then the solution E_X also depends smoothly on these parameters. We can estimate derivatives $\left|\partial_x^\alpha \partial_t^\beta \partial_\lambda^\gamma \Psi(t, x, \lambda)\right|$ in terms of the supremum of derivatives in x and λ of the coefficients $\{a_j\}$ on Ω.

We remark that the solution function $E(t, x)$ actually exists for all $x \in \Omega$, but that as x approaches the boundary of Ω, the interval in t for which the solution exists may shrink to zero. In particular, there is a unique integral curve of X passing through each point of Ω. These curves foliated the open set Ω, and the vector field X is everywhere tangent to this foliation.

COROLLARY 3.3. *Let X be a vector field on Ω and let $K \subset \Omega$ be a compact subset. Let $E(t, x)$ be the solution given in the Theorem 3.2. There exists $\epsilon > 0$ so that*

(1) If $x \in K$ and if $|t_1| + |t_2| < \epsilon$, then
$$E(t_1, E(t_2, x)) = E(t_1 + t_2, x).$$

(2) If $x \in K$, if $\lambda \in \mathbb{R}$ and if $|\lambda\, t| < \epsilon$, then
$$E_X(\lambda\, t, x) = E_{\lambda X}(t, x).$$

(3) Let $|t| < \epsilon$ and let
$$K_t = \{y \in \Omega \,\big|\, y = E(t, x) \text{ for some } x \in K\}.$$
Then $x \to E(t, x)$ is a diffeomorphism of K onto K_t. This mapping is called the flow *associated to the vector field X.*

PROOF. We observe that the mappings $t \to E(t_1, (E(t, x)))$ and $t \to E(t_1 + t, x)$ satisfy the same differential equation and agree when $t = 0$. Thus the first statement in the corollary follows from the uniqueness assertion of the theorem. Similarly, the mappings $t \to E_X(\lambda\, t, x)$ and $t \to E_{\lambda X}(t, x)$ satisfy the same differential equation and agree when $t = 0$. Finally, the first assertion of the corollary shows that $x \to E(-t, x)$ is the inverse of the mapping $x \to E(t, x)$. □

3.2. The exponential mapping.

We now investigate the formal Taylor series expansion of the flow associated to a vector field $X \in T(\Omega)$. We will use the following notation. Let $f \in \mathcal{E}_{x_0}$ be a germ of a smooth function at a point $x_0 \in \mathbb{R}^n$. Then we write
$$f(x) \overset{x}{\sim} \sum_\alpha a_\alpha\, (x - x_0)^\alpha$$
to indicated that the infinite series on the right hand side is the formal Taylor series of the function f about the point x_0. If $X \in T(\Omega)$, we can inductively define the powers $\{X^k\}$ as operators on $\mathcal{E}(\Omega)$. Thus if $f \in \mathcal{E}(\Omega)$, set $X^0[f] = f$,

$X^1[f] = X[f]$ and $X^{k+1}[f] = X\big[X^k[f]\big]$. Note that with this definition X^k is a k^{th} order differential operator. The basic result which allows us to compute formal Taylor series is the following.

LEMMA 3.4. *Let $X \in T(\Omega)$, let $x \in \Omega$, and let $f \in \mathcal{E}(U)$ where $U \subset \Omega$ is an open neighborhood of x. There exists $\epsilon > 0$ so that $E(t,x) \in U$ if $|t| < \epsilon$. Put $F(t) = f\big(E(t,x)\big)$ for $|t| < \epsilon$. Then for every $k \geq 0$,*

$$\frac{d^k F}{dt^k}(t) = F^{(k)}(t) = X^k[f]\big(E(t,x)\big).$$

PROOF. Let $X = \sum_{k=1}^n a_k\,\partial_{x_k}$ with $a_k \in \mathcal{E}(\Omega)$. With x fixed, write $E(t,x) = \big(\psi_1(t), \ldots, \psi_n(t)\big)$. Using the chain rule and equation (3.1) we have

$$\begin{aligned} F'(t) &= \sum_{k=1}^n \frac{\partial f}{\partial x_k}\big(E(t,x)\big)\,\psi_k'(t) \\ &= \sum_{k=1}^n a_k\big(E(t,x)\big)\,\frac{\partial f}{\partial x_k}(E(t,x) = X[f]\big(E(t,x)\big). \end{aligned}$$

The case of general k now follows easily by induction. □

COROLLARY 3.5. *Suppose that f is a smooth function defined in a neighborhood of the point x. Then*

$$f\big(E(t,x)\big) \overset{t}{\sim} \sum_{k=0}^\infty \frac{t^k}{k!} X^k[f](x),$$

where now $\overset{t}{\sim}$ denotes equality as formal Taylor series about $t = 0$.

The form of the sum in Corollary 3.5 suggests the following definition.

DEFINITION 3.6. *The solution $E_X(t,x)$ to the initial value problem $\psi'(t) = X_{\psi(t)}$, $\psi(0) = x$, is called the* exponential map *associated to the vector field X. We write*

$$E(t,x) = e^{tX}(x) = \exp[tX](x).$$

It follows from Corollary 3.3 that for s and t sufficiently small we have

$$\begin{aligned} \exp[(s+t)X](x) &= \exp[sX]\big(\exp[tX](x)\big); \\ \exp[(st)X](x) &= \exp[s(tX)](x). \end{aligned}$$

We can rewrite the case $k = 1$ of Lemma 3.4 in a form that will frequently be useful.

COROLLARY 3.7. *Let $f \in \mathcal{E}(\Omega)$ and $X \in T(\Omega)$. Then for $x \in \Omega$ and $|t|$ sufficiently small we have*

$$\frac{d}{dt} f\big(\exp(tX)(x)\big) = X[f]\big(\exp(tX)(x)\big).$$

In particular

$$X[f](x) = \frac{d}{ds} f\big(\exp((s)X)(x)\big)\Big|_{s=0}.$$

If we use Corollary 3.5 and apply Taylor's theorem, we also have

COROLLARY 3.8. *Let $X \in T(\Omega)$, let $K \subset \Omega$ be compact. Let $x \in K$ and let V be an open neighborhood of x. Let $f \in \mathcal{E}(V)$. There exists $\epsilon > 0$ so that for every integer $M \geq 1$ and every $x \in K$, if $|t| < \epsilon$ then*

$$\left| f\big(\exp[tX](x)\big) - \sum_{k=0}^{M-1} \frac{t^k}{k!} X^k[f](x) \right| \leq C_M\, t^M. \tag{3.2}$$

Here C_M depends on the supremum of derivatives of f up to order M on V and on the supremum of derivatives of the coefficients of the vector field X up to order $M-1$ on Ω.

3.3. Exponential maps of several vector fields.

Now suppose that $X_1, \dots, X_p \in T(\Omega)$. Fix $\delta > 0$, and let $\{d_1, \dots, d_p\}$ be positive real numbers[1]. If $u = (u_1, \dots, u_p) \in \mathbb{R}^p$, then $\sum_{j=1}^p u_j\, \delta^{d_j} X_j$ is again a smooth vector field on Ω. Let $\mathbb{B}(\epsilon)$ be the Euclidean ball of radius ϵ centered at the origin in $\mathbb{R}^p$. It follows from Theorem 3.2 that if $K \subset \Omega$ is compact, there exists $\epsilon > 0$ so that the mapping

$$E_\delta(u; x) = E\big((\delta^{d_1} u_1, \dots, \delta^{d_p} u_p); x\big) = \exp\Big[\sum_{j=1}^p u_j\, \delta^{d_j} X_j\Big](x)$$

is defined and infinitely differentiable on the set $\mathbb{B}(\epsilon) \times K$ and takes values in Ω. We investigate the Taylor expansion of smooth functions composed with the mapping $E_\delta(u; x)$. We have the following analogue of Corollaries 3.5 and 3.8.

LEMMA 3.9. *Suppose that g is a smooth function defined in a neighborhood V of $x \in K$, let $0 < \delta \leq 1$, and put $G(u_1, \dots, u_p) = g\Big(\exp\Big[\sum_{j=1}^p u_j\, \delta^{d_j} X_j\Big](x)\Big)$. Then G is a smooth function defined in a neighborhood of the origin in $\mathbb{R}^p$, and its formal Taylor series at the origin is given by*

$$G(u_1, \dots, u_p) \overset{u}{\sim} \sum_{k=0}^{\infty} \frac{1}{k!} \Big(\sum_{j=1}^p u_j\, \delta^{d_j} X_j\Big)^k [g](x).$$

Let $d = \min_{1 \leq j \leq p} d_j$. For every integer $M \geq 1$ and every $x \in K$, if $|u| < 1$ then

$$\left| G(u_1, \dots, u_p) - \sum_{k=0}^{M-1} \frac{1}{k!} \Big(\sum_{j=1}^p u_j\, \delta^{d_j} X_j\Big)^k [g](x) \right| \leq C_M\, \delta^{Md} |u|^M.$$

Here C_M is independent of δ and depends on the supremum of derivatives of g up to order M on V and on the supremum of derivatives of the coefficients of the vector fields $\{X_1, \dots, X_p\}$ up to order $M-1$ on Ω.

PROOF. Set

$$\widetilde{G}(t, u_1, \dots, u_p) = G(tu_1, \dots, tu_p) = g\Big(\exp\Big[t \sum_{j=1}^p u_j\, \delta^{d_j} X_j\Big](x)\Big).$$

If we write $Y_u = \sum_{j=1}^p u_j\, \delta^{d_j} X_j$, then Y_u is a smooth vector field and

$$\widetilde{G}(t, u_1, \dots, u_p) = g\big(\exp(t\, Y_u)(x)\big).$$

[1] This extra parameter δ will be needed in Chapter 3, Section 3, when we introduce the mapping $\Theta_{x,\delta}$ in equation (**??**).

Thus according to the chain rule and Corollary 3.5, we have

$$\sum_{|\alpha|=n} \frac{n!}{\alpha!}\, u^\alpha\, \frac{\partial^\alpha G}{\partial u^\alpha}(0) = \frac{\partial^n \widetilde{G}}{\partial t^n}(0, u_1, \dots, u_p) = \Big(\sum_{j=1}^p u_j\, \delta^{d_j}\, X_j\Big)^n [g](x).$$

Consequently, as formal power series in $(u_1, \dots, u_p)$, we have the identity of formal series

$$\sum_\alpha \frac{u^\alpha}{\alpha!} \frac{\partial^\alpha G}{\partial u^\alpha}(0) = \sum_{n=0}^\infty \frac{1}{n!}\Big(\sum_{j=1}^p u_j\, \delta^{d_j}\, X_j\Big)^n [g](x).$$

The left hand side is the formal Taylor expansion of G, which gives the first assertion of the lemma. The second assertion about approximation follows as before from Taylor's theorem. This completes the proof. □

Note that $\Big(\sum_{j=1}^p u_j\, X_j\Big)^n$ is a differential operator of order n. If the vector fields $\{X_1, \dots, X_p\}$ commute, we can write $\Big(\sum_{j=1}^p u_j\, X_j\Big)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!}\, u^\alpha\, X^\alpha$ where $X^\alpha = X_1^{\alpha_1} \cdots X_p^{\alpha_p}$. In this case the Taylor expansion of G can be written

$$G(u_1, \dots, u_p) \overset{u}{\sim} \sum_\alpha \frac{u^\alpha}{\alpha!}\, X^\alpha[g](x),$$

and we can conclude that $\dfrac{\partial^\alpha G}{\partial u^\alpha}(0) = X^\alpha[g](x)$. However, if the vector fields do not commute, the expression for $\dfrac{\partial^\alpha G}{\partial u^\alpha}(0)$ is a non-commuting polynomial of degree $|\alpha|$ in the vector fields $\{X_1, \dots, X_p\}$ applied to g and evaluated at 0.

We shall need one further generalization of Lemma 3.9. Suppose that

$$P(u; X) = \sum_{j=1}^p p_j(u)\, X_j,$$

where each p_j is a polynomial in the variables $(u_1, \dots, u_p)$ and has no constant term.

Corollary 3.10. *Suppose that $V \subset \Omega$ is an open neighborhood of $x \in K$ and that $g \in \mathcal{E}(V)$. Put $G(u) = g\Big(\exp\big(P(u;X)\big)(x)\Big)$. Then G is a smooth function defined in a neighborhood of the origin in $\mathbb{R}^p$, and*

(1) *The formal Taylor series of G_x at the origin in $\mathbb{R}^p$ is given by*

$$G(u) \overset{u}{\sim} \sum_{k=0}^\infty \frac{1}{k!}\big(P(u;X)\big)^k[g](x); \tag{3.3}$$

(2) *There exists $\epsilon > 0$ so that for all $x \in K$ and all integers $M \geq 1$, if $|u| < \epsilon$ then*

$$\left| G_x(u) - \sum_{k=0}^{M-1} \frac{1}{k!}\big(P(u;X)\big)^k[g](x) \right| \leq C_M |u|^M. \tag{3.4}$$

The constant C_M depends on the supremum of the derivatives of g up to order M on V and on the supremum of the derivatives of the coefficients of the vector fields $\{X_1, \dots, X_p\}$ up to order $M-1$ on the set Ω.

4. Composition of flows

Let X and Y be two smooth real vector fields on an open subset $\Omega \subset \mathbb{R}^n$. Each vector field defines a flow, $\exp(sX)$ and $\exp(tY)$, which are defined on compact subsets $K \subset \Omega$ if $|s|$ and $|t|$ are sufficiently small. In this section we study the composition $\exp(sX) \circ \exp(tY)$ when it is defined.

We begin by showing that the flows $\exp(sX)$ and $\exp(tY)$ commute if and only if the commutator $[X, Y] = 0$. Moreover, we show that in this case we have the usual law for products of exponents $\exp(sX) \circ \exp(tY) = \exp(sX + tY)$. If the two flows do not commute, the situation is more complicated. We show, using the Campbell-Baker-Hausdorff formula, that there are smooth vector fields $\{Z_N(s,t)\}$ so that $\exp(sX) \circ \exp(tY) - \exp\big(Z_N(s,t)\big)$ vanishes to high order.

4.1. Commuting vector fields.

The goal of this subsection is to establish the following result which gives a first indication of the relationship between flows and commutators.

THEOREM 4.1. *Let $X, Y \in T(\Omega)$ be smooth real vector fields. The the following two conditions are equivalent.*

(1) For every $x \in \Omega$ we have $\exp(sX)\big(\exp(tY)(x)\big) = \exp(tY)\big(\exp(sX)(x)\big)$ for all (s,t) in an open neighborhood of the origin in $\mathbb{R}^2$, where the neighborhood may depend on x.

(2) The commutator $[X, Y] = 0$ on Ω.

We begin by studying the differential of the flow associated to a smooth real vector field $X \in T(\Omega)$. If $K \subset \Omega$ is compact, Corollary 3.3 guarantees the existence of a constant $\epsilon > 0$ so that for $|t| < \epsilon$, the mapping $\exp(tX) : K \to K_t$ is a diffeomorphism. The differential of the flow $x \to \exp(tX)(x)$ gives an isomorphism between tangent spaces $d\exp(tX) : T_x \to T_{\exp(tX)(x)}$. In particular, the vector field X defines a tangent vector $X_x \in T_x$, and it follows from Corollary 3.7 that

$$d\exp(tX)[X_x] = X_{\exp(tX)(x)}. \tag{4.1}$$

Thus the vector field X is invariant under the flow generated by X.

We next observe that if $Y \in T(\Omega)$ is a second vector field, we can use the flow associated to X to generate a local[2] one parameter family of vector fields $t \to Y_t$ with $Y_0 = Y$. More precisely, for each compact subset $K \subset \Omega$, let ϵ be the constant from Corollary 3.3. We define Y_t by giving the corresponding tangent vector $(Y_t)_x \in T_x$ at each point $x \in K$. For $|t| < \frac{\epsilon}{2}$ the vector field Y defines a tangent vector $Y_{\exp(tX)(x)}$ at the point $\exp(tX)(x)$. Since $x = \exp(-tX)\big(\exp(tX)(x)\big)$, the differential $d\exp(-tX)$ maps $T_{\exp(tX)(x)}$ to T_x, and we define a tangent vector $(Y_t)_x \in T_x$ by

$$(Y_t)_x = d\exp(-tX)[Y_{\exp(tX)(x)}]. \tag{4.2}$$

This means that if f is a smooth function defined in a neighborhood of K, then for $x \in K$ and $|t|$ sufficiently small

$$(Y_t)_x[f] = Y_t[f](x) = Y[f \circ \exp(-tX)]\big(\exp(tX)(x)\big). \tag{4.3}$$

[2]The term *local* here means that the family $\{Y_t\}$ is defined on compact subsets for $|t|$ sufficiently small.

Let us see what this means in a simple example. Suppose that $X = \partial_{x_1}$ is differentiation with respect to the first coordinate. Then $\exp(tX)(x_1, x_2, \dots, x_n) = (x_1 + t, x_2, \dots, x_n)$, and $f \circ \exp(-tX)(x) = f(x_1 - t, x_2, \dots, x_n)$. Suppose $Y = \sum_{j=1}^{n} b_j \, \partial_{x_j}$. Then

$$Y[f \circ \exp(tX)](x_1, \dots, x_n) = \sum_{j=1}^{n} b_j(x_1, \dots, x_n) \frac{\partial f}{\partial x_j}(x_1 - t, x_2, \dots, x_n)$$

and hence

$$Y_t[f](x_1, x_2, \dots, x_n) = \sum_{j=1}^{n} b_j(x_1 + t, x_2, \dots, x_n) \frac{\partial f}{\partial x_j}(x_1, x_2, \dots, x_n).$$

Thus $(Y_t)_x$ is just the value of the vector field Y at points along the integral curve of X through x.

The next Lemma provides a connection between this family of tangent vectors and the commutator of X and Y. (See Narasimhan [**Nar85**]).

LEMMA 4.2. *Let X and Y be smooth real vector fields on Ω, and let $x \in \Omega$. Let $(Y_t)_x \in T_x$ be the local one-parameter family of tangent vectors defined in equations (4.2). Then*

$$\frac{d}{dt}(Y_t)_x = [X, Y_t]_x \,.$$

(We remark that if $X = \partial_{x_1}$ as in the above example, then

$$\frac{d}{dt} Y_t[f](x_1, x_2, \dots, x_n) = \sum_{j=1}^{n} \frac{\partial b_j}{\partial x_1}(x_1 + t, x_2, \dots, x_n) \frac{\partial f}{\partial x_j}(x_1, x_2, \dots, x_n),$$

and this is indeed the commutator $[X, Y_t]$.)

PROOF OF LEMMA 4.2. We have

$$\begin{aligned}
\frac{d}{dt}\big((Y_t)_x[f]\big)\Big|_{t=0} &= \lim_{t \to 0^+} t^{-1} \Big[Y[f \circ \exp(-tX)] \circ \exp(tX)(x) - Y[f](x)\Big] \\
&= \lim_{t \to 0} t^{-1} \Big[Y[f \circ \exp(-tX)] - Y[f]\Big] \circ \exp(tX)(x) \\
&\qquad - \lim_{t \to 0^+} t^{-1} \Big[Y[f] \circ \exp(-tX) - Y[f]\Big] \circ \exp(tX)(x) \\
&= \lim_{t \to 0} t^{-1} \Big[Y[f \circ \exp(-tX)] - Y[f]\Big](x) \\
&\qquad - \lim_{t \to 0^+} t^{-1} \Big[Y[f] \circ \exp(-tX) - Y[f]\Big](x)
\end{aligned}$$

since $\lim_{t \to 0} \exp(tX)(x) = x$. Now

$$\lim_{t \to 0} t^{-1} \Big[Y[f] \circ \exp(-tX) - Y[f]\Big](x) = -X\big[Y[f]\big](x)$$

by the definition of the exponential map. On the other hand

$$\begin{aligned}
\lim_{t \to 0} t^{-1} \Big[Y[f \circ \exp(-tX)] - Y[f]\Big](x) &= -\lim_{t \to 0} Y\Big[\frac{f \circ \exp(tX)] - f}{t}\Big](x) \\
&= -Y\big[X[f]\big](x)
\end{aligned}$$

This establishes the lemma when $t = 0$. But for $|t|$ small we have

$$\begin{aligned} d\exp(tX)\big[\frac{d}{dt}\big(Y_t[f]\big)\big] &= d\exp(tX)\big[\frac{d}{dt}\big(Y_t[f]\big)\Big|_{t=0}\big] \\ &= d\exp(tX)\,[Y_0,\, X]\,[f] \\ &= [d\exp(tX)Y_0,\, d\exp(tX)X]\,[f] \\ &= [Y_t,\, X]\,[f], \end{aligned}$$

which completes the proof of the lemma. □

Proof of Theorem 4.1. Suppose first that the flows generated by X and Y commute. That is, assume that for all $x \in \Omega$ we have

$$\exp(sX)\big(\exp(tY)(x)\big) = \exp(tY)\big(\exp(sX)(x)\big)$$

for all (s,t) in an open neighborhood of the origin in $\mathbb{R}^2$. Let f be a smooth function defined a neighborhood of x, and put

$$F(s,t) = f\big(\exp(sX)\big(\exp(tY)(x)\big).$$

Then by Corollary 3.7, we have

$$\begin{aligned} \frac{\partial^2 F}{\partial t \partial s}(s,t) &= \frac{\partial}{\partial t}\left[\frac{\partial F}{\partial s}(s,t)\right] \\ &= \frac{\partial}{\partial t}\left[X[f]\big(\exp(sX)\big(\exp(tY)(x)\big)\right] \\ &= \frac{\partial}{\partial t}\left[X[f]\big(\exp(tY)\big(\exp(sX)(x)\big)\right] \\ &= Y\big[X[f]\big]\big(\exp(tY)\big(\exp(sX)(x)\big) \\ &= YX[f]\big(\exp(sX)\big(\exp(tY)(x)\big). \end{aligned}$$

Since $\frac{\partial^2 F}{\partial t \partial s} = \frac{\partial^2 F}{\partial s \partial t}$, we see that $YX[f] = XY[f]$, and hence $[X,\, Y] = 0$.

Next suppose that $[X,\, Y] = 0$. Let $\{Y_t\}$ be the local family of vector fields defined in equation (4.2. Then since $0 = d\exp(tX)[0]$, we have

$$\begin{aligned} \frac{d}{dt}\,(Y_t) &= [X,\, Y_t] \\ &= [d\exp(tX)[X],\, d\exp(tX)[Y]] \\ &= d\exp(tX)\big[\,[X,\, Y]\,\big] \\ &= d\exp(tX)[0] = 0. \end{aligned}$$

Thus Y_t is constant, and since $Y_0 = Y$, it follows that $Y_t = Y$ for all small t. It follows from equation (4.3) that if $f \in \mathcal{E}(\Omega)$ and if $x \in \Omega$, then

$$Y[f]\big(\exp(tX)(x)\big) = Y[f \circ \exp(tX)](x)$$

for $|t|$ small. Then for $x \in \Omega$ and $|s| + |t|$ small we have

$$\begin{aligned} \frac{d}{ds}\Big[f\big(\exp(tX) \circ \exp(sY)(x)\big)\Big] &= \frac{d}{ds}\Big[f \circ \exp(tX)\big(\exp(sY)(x)\big)\Big] \\ &= Y\big[f \circ \exp(tX)\big]\big(\exp(sY)(x)\big) \\ &= Y[f]\Big(\exp(tX)\big(\exp(sY)(x)\big)\Big) \\ &= Y[f]\big(\exp(tX) \circ \exp(sY)(x)\big), \end{aligned}$$

or equivalently

$$\frac{d}{ds}\Big[f\big(\exp(tX)\circ\exp(sY)\circ(-tX)(x)\big)\Big] = Y[f]\big(\exp(tX)\circ\exp(sY)\circ\exp(-tX)(x)\big).$$

But then the curves

$$\begin{aligned}\gamma_1(s) &= \exp(sY)(x),\\ \gamma_2(s) &= \exp(tX)\circ\exp(sY)\circ\exp(-tX)(x),\end{aligned}$$

satisfy the differential equation

$$\frac{d}{ds}\big[f\big(\gamma_j(s)\big)\big] = Y[f]\big(\gamma_j(s)\big)$$

and satisfy the initial condition $\gamma_j(0) = x$. By the uniqueness statement in Theorem 3.2, it follows that $\gamma_1(s) = \gamma_2(s)$, and hence

$$\exp(sY)\circ\exp(tX)(x) = \exp(tX)\circ\exp(sY)(x)$$

for $|s| + |t|$ sufficiently small. This completes the proof. □

We can now show that the usual rule for multiplication of exponentials holds for $\exp(sX)\circ\exp(tY)$ is the vector fields X and Y commute.

PROPOSITION 4.3. *Suppose that $X, Y \in T(\Omega)$ and that $[X, Y] = 0$. Let $K \subset \Omega$ be compact. Then there exists $\epsilon > 0$ so that*

$$\exp(sX)\circ\exp(tY)(x) = \exp(sX + tY)(x)$$

for all $x \in K$ and all $|s| + |t| < \epsilon$.

PROOF. Let V be an open neighborhood of K, and let $f \in \mathcal{E}(V)$. For $x \in K$ and $|t| + |s|$ sufficiently small, let

$$H(s,t) = f\big(\exp(sX)\circ\exp(tY)(x)\big) = f\big(\exp(tY)\circ\exp(sX)(x)\big).$$

Then it follows from Lemma 3.4 that

$$\begin{aligned}\frac{\partial H}{\partial s}(s,t) &= X[f]\big(\exp(sX)\circ\exp(tY)(x)\big),\\ \frac{\partial H}{\partial t}(s,t) &= Y[f]\big(\exp(sX)\circ\exp(tY)(x)\big).\end{aligned}$$

Hence

$$\frac{d}{dt}f(\exp(tX)\circ\exp(tY)(x) = (X+Y)[f](\exp(tX)\circ\exp(tY)(x).$$

But since we also have

$$\frac{d}{dt}f(\exp\big(t(X+Y)\big)(x) = (X+Y)[f](\exp\big(t(X+Y)\big)(x)$$

the uniqueness of solutions implies that whenever X and Y commute, we have

$$\exp(tX)\circ\exp(tY)(x) = \exp\big(t(X+Y)\big)(x) \tag{4.4}$$

for all small $|t|$.

To show that $\exp(sX)\circ\exp(tY)(x) = \exp(sX+tY)(x)$, we may assume that $t \neq 0$. Then $sX = t(st^{-1})X$, and since $(st^{-1})X$ and Y commute, we can apply equation (4.4) to get

$$\begin{aligned}\exp(sX)\circ\exp(tY)(x) &= \exp\big(t\,(st^{-1})X\big)\circ\exp(tY)(x)\\ &= \exp\big(t((st^{-1}X+Y)\big)\\ &= \exp(sX+tY)(x).\end{aligned}$$

This completes the proof. □

4.2. Non-commuting vector fields.

Next we turn to the description of a composition $\exp(sX)\circ\exp(tY)$ when X and Y do not commute. Ideally we would like to find a vector field $Z(s,t)$, depending smoothly on s and t, so that $\exp(sX)\circ\exp(tY) = \exp\big(Z(s,t)\big)$. In general this is not possible, but we can find smooth vector fields $\{Z_N(s,t)\}$ so that $\exp\big(Z_N(s,t)\big)$ agrees to high order with $\exp(sX)\circ\exp(tY)$. Moreover, we can take $Z_N(s,t)$ to be the N^{th} partial sum of a *formal* infinite series in s and t with coefficients which are iterated commutators of X and Y. This construction depends on the Campbell-Baker-Hausdorff formula.

We shall, in fact, work somewhat more generally. Let $K \subset \Omega$ be compact and let $X_1,\ldots,X_p,Y_1,\ldots,Y_q \in T(\Omega)$ be smooth vector fields. If $u=(u_1,\ldots,u_p)\in\mathbb{R}^p$ and $v=(v_1,\ldots,v_q)\in\mathbb{R}^q$, write $u\cdot X=\sum_{j=1}^p u_jX_j$ and $v\cdot Y=\sum_{k=1}^q v_kY_k$. Then there exists $\epsilon>0$ so that if $|u|<\epsilon$ and $|v|<\epsilon$, the mapping

$$x \to \big[\exp(u\cdot X)\circ\exp(v\cdot Y)\big](x) = \exp(u\cdot X)\big(\exp(v\cdot Y)(x)\big)$$

is defined on K and maps K diffeomorphically onto its image.

Let $x\in K$, let V be an open neighborhood of x, and let $f\in\mathcal{E}(V)$. Then

$$F(u,u) = f\Big(\exp(u\cdot X)\big(\exp(v\cdot Y)(x)\big)\Big)$$

is infinitely differentiable in a neighborhood of the origin in $\mathbb{R}^p\times\mathbb{R}^q$. Using Lemma 3.9, we have

$$F(u,v) \overset{u}{\sim} \sum_{m=0}^{\infty}\frac{1}{m!}(u\cdot X)^m[f]\Big(\exp(tY)(x)\Big),$$

or equivalently, for each $m\geq 0$,

$$\sum_{|\alpha|=m}\frac{u^\alpha}{\alpha!}\frac{\partial^{|\alpha|}F}{\partial u^\alpha}(0,v) = \frac{1}{m!}\,(u\cdot X)^m[f]\Big(\exp(v\cdot Y)(x)\Big).$$

We can apply Lemma 3.9 again and obtain

$$\sum_{|\alpha|=m}\frac{u^\alpha}{\alpha!}\frac{\partial^{|\alpha|}F}{\partial u^\alpha}(0,v) \overset{v}{\sim} \frac{1}{m!}\sum_{n=0}^{\infty}\frac{1}{n!}(t\cdot Y)^n\big[X^m[f]\big](x),$$

or equivalently, for each $m,n\geq 0$.

$$\sum_{|\alpha|=m}\sum_{|\beta|=n}\frac{v^\beta}{\beta!}\frac{u^\alpha}{\alpha!}\frac{\partial^{|\alpha|+|\beta|}F}{\partial^\beta v\partial u^\alpha}(0,0) = \frac{1}{n!m!}(v\cdot Y)^n(u\cdot X)^m[f](x).$$

Thus the Taylor expansion of F in u and v about the point $(0,0)$ is given by

$$F(u,v) \overset{u,v}{\sim} \sum_{m,n=0}^{\infty} \frac{1}{m!n!}\,(v\cdot Y)^n (u\cdot X)^m [f](x).$$

However, as a formal power series in the non-commuting variables $(u\cdot X)$ and $(v\cdot Y)$, we have

$$\sum_{m,n=0}^{\infty} \frac{1}{m!n!}\,(v\cdot Y)^n (u\cdot X)^m = \exp(v\cdot Y)\,\exp(u\cdot X).$$

Thus we have a product of two formal power series given as exponentials.

The Campbell-Baker-Hausdorff formula expresses a product of exponentials $\exp(A)\exp(B)$ as the exponential of a formal series which is given in terms of commutators of A and B. Explicitly,

$$\exp(A)\,\exp(B) = \exp\big(Z(A,B)\big)$$

where

$$Z(A,B) = A + B + \sum_{n=2}^{\infty} P_N(A,B)$$

and

$$P_N(A,B) = \frac{1}{N}\sum_{k=1}^{N} \frac{(-1)^{k+1}}{k} \sum_{\substack{\sum(m_j+n_j)=N\\ m_j+n_j\geq 1}} \frac{1}{\prod m_j!n_j!}\ \mathrm{ad}_A^{m_1}\mathrm{ad}_B^{n_1}\cdots\mathrm{ad}_A^{m_k}\mathrm{ad}_B^{n_k}. \tag{4.5}$$

The proof of the Campbell-Baker-Hausdorff formula is given in Chapter 7, Section 5.

Recall that $\mathrm{ad}_x[y] = [x,y]$, and the meaning of equation (4.5) is

$$\mathrm{ad}_A^{m_1}\mathrm{ad}_B^{n_1}\cdots\mathrm{ad}_A^{m_k}\mathrm{ad}_B^{n_k} = \begin{cases} \mathrm{ad}_A^{m_1}\mathrm{ad}_B^{n_1}\cdots\mathrm{ad}_A^{m_k}\mathrm{ad}_B^{n_k-1}[B] & \text{if } n_k\geq 1,\\ \\ \mathrm{ad}_A^{m_1}\mathrm{ad}_B^{n_1}\cdots\mathrm{ad}_B^{n_{k-1}}\mathrm{ad}_A^{m_k-1}[A] & \text{if } n_k = 0.\end{cases} \tag{4.6}$$

Thus each P_N is a linear combination of iterated commutators of A and B of total length N. In particular,

$$\begin{aligned} P_2(A,B) &= \frac{1}{2}[A,B],\\ P_3(A,B) &= \frac{1}{12}[A,[A,B]] - \frac{1}{12}[B,[A,B]],\\ P_4(A,B) &= -\frac{1}{48}[B,[A,[A,B]]] - \frac{1}{48}[A,[B,[A,B]]]. \end{aligned}$$

We now apply this formula to the case when $A = \sum_{j=1}^{p} u_j\,X_j = u\cdot X$ and $B = \sum_{k=1}^{q} v_k\,Y_k$. It follows that

$$F(u,v) \overset{u,v}{\sim} \exp\big(Z(u,v)\big)[f](x)$$

where as a formal series

$$Z(u,v) = u\cdot X + v\cdot Y + \sum_{N=2}^{\infty} \sum_{\substack{|\alpha|+|\beta|=N\\ |\alpha|,|\beta|\geq 1}} u^\alpha v^\beta P_{\alpha,\beta}(X,Y)$$

and $P_{\alpha,\beta}(X,Y)$ is a linear combination of iterated commutators of the vectors $\{X_1,\ldots,X_p,Y_1,\ldots,Y_q\}$ of length $|\alpha|+|\beta|$. Thus for example

$$\sum_{\substack{|\alpha|+|\beta|=2\\|\alpha|,|\beta|\geq 1}} u^\alpha v^\beta P_{\alpha,\beta}(X,Y) = \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{q} u_j v_k [X_j, Y_k]$$

and

$$\sum_{\substack{|\alpha|+|\beta|=3\\|\alpha|,|\beta|\geq 1}} u^\alpha v^\beta P_{\alpha,\beta}(X,Y) = \frac{1}{12}\sum_{j,k=1}^{p}\sum_{l=1}^{q} u_j u_k v_l [X_j,[X_k,Y_l]]$$
$$- \frac{1}{12}\sum_{k=1}^{p}\sum_{j,l=1}^{q} v_j u_k v_l [Y_j,[X_k,Y_l]]$$

For any integer $M \geq 2$ let

$$Z_M(u,v) = u\cdot X + v\cdot Y + \sum_{\substack{|\alpha|,|\beta|\geq 1\\|\alpha|+|\beta|\leq M}} u^\alpha\, v^\beta P_{\alpha,\beta}(X,Y).$$

$Z_M(u,v)$ is a polynomial of degree M in u and v with coefficients which are linear combinations of vector fields, each of which is an iterated commutators of the vector fields $\{X_1,\ldots,X_p,Y_1,\ldots,Y_q\}$. In particular, $Z_M(u,v)$ is a smooth vector field for each fixed (u,v).

LEMMA 4.4. *Let $K\subset\Omega$ be compact. There exists $\epsilon>0$ so that if $x\in K$ and if f is a smooth function defined in a neighborhood of x, then for all $|u|+|v|<\epsilon$*

$$\Big|f\Big(\exp\Big[\sum_{j=1}^{p} u_j\, X_j\Big]\circ\exp\Big[\sum_{k=1}^{q} v_j\, Y_j\Big](x)\Big) - f\Big(\exp\big[Z_M(u,v)\big](x)\Big)\Big|$$
$$\leq C_M\,\big(|u|+|v|\big)^{M+1}$$

where the constant C_M depends on estimates for derivatives of f up to order $M+1$ in a neighborhood of x and on the supremum of the derivatives of the coefficients of the vector fields $\{X_1,\ldots,X_p,Y_1,\ldots,Y_q\}$. In particular, if we take for f any coordinate function on Ω, it follows that

$$\Big|\exp\Big[\sum_{j=1}^{p} u_j\, X_j\Big]\circ\exp\Big[\sum_{j=1}^{p} v_j\, X_j\Big](x) - \exp\big[Z_M(u,v)\big](x)\Big| \leq C_M\,\big(|u|+|v|\big)^{M+1}$$

Suppose again that $\{X_1,\ldots,X_p\}$ are smooth vector fields on Ω. For $x\in\Omega$ and $|s|+|t|$ sufficiently small, we can consider the mapping

$$\Gamma(s,t)(x) = \exp\left[\sum_{j=1}^{p}(s_j+t_j)\,X_j\right]\circ\exp\left[-\sum_{j=1}^{p} t_j\,X_j\right](x)$$
$$\equiv \exp\left[(s+t)\cdot X\right]\circ\exp\left[-t\cdot X\right](x).$$

We list some elementary properties of the mappings $\Gamma(s,t)$:

(1) For $t=0$, $\Gamma(s,0)(x)=\exp[sX](x)$.

(2) For $s=0$, $\Gamma(0,t)(x)=x$.

(3) If the vector fields $\{X_j\}$ commute, then for all t, $\Gamma(s,0)(x) = \exp[sX](x)$.

Thus for each $t \in \mathbb{R}^p$, $s \to \Gamma(s,t)(x)$ is a smooth mapping of a neighborhood of the origin in $\mathbb{R}^p$ to Ω which takes 0 to x. When $t = 0$, this mapping is just the exponential mapping, but as t varies, we get a family of such mappings.

LEMMA 4.5. *For $1 \leq j \leq p$, and for each multi-index α with $|\alpha| \geq 2$ there is a smooth vector fields $Z_{j,\alpha}$ on Ω so that for any integer M, if we let $Z_{j,M}(t)$ denote the smooth vector field*

$$Z_{j,M} = X_j + \sum_{2\leq|\alpha|\leq M} t^\alpha\, Z_{j,\alpha},$$

then if f is smooth in a neighborhood of x, we have

$$\big|f\big(\Gamma(s,t)(x)\big) - f\big(\exp[s\cdot Z_M](x)\big)\big| \leq C\,\left[s^2 + s\,t^{M+1}\right].$$

Proof?

5. Normal coordinates

Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $X_1, \ldots, X_n \in T(\Omega)$ be smooth vector fields. The exponential mapping allows us to introduce special coordinates near each point in Ω which are particularly adapted to the n vector fields.

5.1. The definition of Θ_x.

Let $K \subset \Omega$ be compact. Then using Theorem 3.2, it follows that there exists $\epsilon > 0$ so that if $x \in K$ and if $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ with $|u| < \epsilon$, then

$$\Theta_x(u_1, \ldots, u_n) = \exp\Big[\sum_{j=1}^n u_j X_j\Big](x) \tag{5.1}$$

is defined and belongs to Ω. Moreover, if $\mathbb{B}(\epsilon)$ is the ball of radius ϵ centered at the origin in $\mathbb{R}^n$, $\Theta_x : \mathbb{B}(\epsilon) \to \Omega$ is an infinitely differentiable mapping with $\Theta_x(0) = 0$.

LEMMA 5.1. *Let $X_1, \ldots, X_n \in T(\Omega)$. For every $x \in \Omega$,*

$$J\Theta_x(0) = \det(X_1, \ldots, X_n)(x),$$

where $J\Theta_x$ is the Jacobian determinant of the mapping Θ_x. If $K \subset \Omega$ is compact and if $\det(X_1, \ldots, X_n)(x) \neq 0$ for all $x \in K$ there exists $\eta > 0$ so that for all $x \in K$

(1) the mapping Θ_x is a diffeomorphism of $\mathbb{B}(\eta)$ onto its image, denoted by N_x;

(2) for all $y \in N_x$, $\det(X_1, \ldots, X_n)(y) \neq 0$.

PROOF. Let $x \in \Omega$, and let f be a smooth function defined in an open neighborhood of x. Put

$$F(t) = f\big(\Theta_x(0, \ldots, 0, t, 0, \ldots, 0)\big) = f\big(\exp(tX_j)(x)\big),$$

where t appears in the j^{th} entry of Θ_x. Then F is a smooth function defined in a neighborhood of the origin in $\mathbb{R}$, and Lemma 3.4 shows that $F'(0) = X_j[f](x)$. According to Definition 2.2, $d\Theta_x[\partial_{u_j}]_x = (X_j)_x$, and this shows that $J\Theta_x(0) = \det(X_1, \ldots, X_n)(x)$. It then follows from the open mapping theorem that there is an $\eta > 0$ so that Θ_x is a diffeomorphism on $\mathbb{B}(\eta)$, and η can be chosen uniformly for all $x \in K$. Shrinking η if necessary, we can insure that $\det(X_1, \ldots, X_n)$ does not vanish on N_x. This completes the proof. □

We shall denote the inverse of the mapping Θ_x by Ξ_x. Thus N_x is an open neighborhood of $x \in K$ and $\Xi_x : N_x \to \mathbb{B}(\eta)$ is a diffeomorphism. We can compose the coordinate functions $\{u_1, \ldots, u_n\}$ on $\mathbb{B}(\eta)$ with the mapping Ξ_x, and this gives a new coordinate system on N_x called the *normal coordinates* relative to the vector fields $\{X_1, \ldots, X_n\}$.

5.2. The action of $d\Theta_x$ on tangent vectors.

On $\mathbb{B}(\eta) \subset \mathbb{R}^n$ we have the standard coordinate vector fields $\{\partial_{u_1}, \ldots, \partial_{u_n}\}$, and we can use the diffeomorphism Θ_x to push these vector fields forward and obtain n vector fields $\{d\Theta_x[\partial_{u_1}], \ldots, d\Theta_x[\partial_{u_n}]\}$ on N_x. We also have the original n vector fields $\{X_1, \ldots, X_n\}$ on N_x. We observed in the proof of Lemma 5.1 that these two sets of vector fields determine the same tangent vectors at the point x. However, this is *not* true in general for $y \in N_x$ with $y \neq x$. To understand how the two sets of vector fields are related in general, we need to use the Campbell-Baker-Hausdorff formula.

LEMMA 5.2. *Let $K \subset \Omega$ be compact and let N be a positive integer. Shrinking η if necessary, there is a constant C_N so that if $x \in K$ and $|u| < eta$*

$$\Big| d\Theta_x[\partial_{u_j}] - \Big[X_j + \sum_{k=2}^{N} \alpha_k \overbrace{[u\cdot X,\, [u\cdot X,\, \cdots [u\cdot X,\, X_j]]]}^{k\text{-fold iterate}} \Big] \Big| \leq C_N\, |u|^{N+1}$$

where it is understood that all vector fields are evaluated at the point $\Theta_x(u) \in N_x$. Here $\{\alpha_2, \alpha_3, \ldots\}$ are constants (rational numbers) which arise from the Campbell-Baker Hausdorff formula.

PROOF. If we could find a family of vector fields $W_j(u)$ on N_x, depending smoothly on $u \in \mathbb{B}(\eta)$, such that

$$\exp(u\cdot X + tX_j)(x) = \exp\big(tW_j(u)\big)\big(\exp\big(u\cdot X\big)(x)\big) \tag{5.2}$$

it would follow from Lemma 3.4 that for any smooth function f defined in a neighborhood of x we have

$$\frac{d}{dt}\Big[f\big(\exp\big(u\cdot X + tX_j\big)(x)\big)\Big]_{t=0} = W_j(u)[f](\Theta_x(u)).$$

The definition of the mapping Θ_x would then show that

$$\big(d\Theta_x[\partial_{u_j}]\big)_{\Theta_x(u)} = W_j(u).$$

There may not be any such vector fields $W_j(u)$, but we can use the Campbell-Baker-Hausdorff formula to find a family of vector fields $W_{j,N}(u)$ for which equation (5.2) holds up to order N in u. To do this, note that equation (5.2) is equivalent to the equation

$$\exp\big(tW_j\big) = \exp\big(u\cdot X + tX_j\big) \circ \exp\big(-u\cdot X\big).$$

We use Lemma 4.4 to approximate the right hand side. If $\hat{e}_j$ denotes the unit vector in $\mathbb{R}^n$ with $+1$ in the j^{th} entry and zeros elsewhere, we have $u\cdot X + tX_j = (u + t\hat{e}_j)\cdot X$. Then if we put

$$Z_{j,N}(u,t) = (u\cdot X + tX_j) + (-u\cdot X) + \sum_{\substack{|\alpha|,|\beta|\geq 1\\ |\alpha|+|\beta|\leq N}} (u + t\hat{e}_j)^{\alpha}(-u)^{\beta} P_{\alpha,\beta}(X,X),$$

it follows from Lemma 4.4 that

$$\Big|\exp\big(u\cdot X + tX_j\big)\circ\exp\big(-u\cdot X\big) - \exp\big(Z_{j,N}(u,t)\big)\Big| \leq C_N\big(|u|^{N+1}+|t|^{N+1}\big).$$

Note that $Z_{j,N}(u,0)=0$, so $Z_{j,N}(u,t)$ is divisible by t. We only want the part that is linear in t, and we can write

$$\begin{aligned} Z_{j,N}(u,t) &= t\Big[X_j + \sum_{k=2}^{N}\alpha_k\overbrace{[u\cdot X,\,[u\cdot X,\,\cdots[u\cdot X,\,X_j]]]}^{k\text{-fold iterate}}\Big] + O(t^2)\\ &= t\,W_{j,N}(u) + O(t^2). \end{aligned}$$

Then if f is a smooth function defined in a neighborhood of x we have

$$\left|\frac{d}{dt}\Big[f\big(\exp\big(u\cdot X+tX_j\big)(x)\big)\Big]_{t=0} - W_{j,N}(u)[f]\big(\exp(u\cdot X)(x)\big)\right| \leq C_N\,|u|^{N+1}$$

Since

$$\big(d\Theta_x[\partial_{u_j}]\big) = \frac{d}{dt}\Big[f\big(\exp\big(u\cdot X+tX_j\big)(x)\big)\Big]_{t=0},$$

this completes the proof. □

6. Submanifolds and tangent vectors

Let $\Omega\subset\mathbb{R}^n$ be an open set, and let $M\subset\Omega$ be a k-dimensional submanifold. If $X\in T(\Omega)$, we investigate what it means to say that X is *tangent* to M at a point $p\in M$.

6.1. Submanifolds.

We begin with the definition of a submanifold. Thus

Definition 6.1. *Let $\Omega\subset\mathbb{R}^n$ be an open set. A subset $M\subset\Omega$ is a k-dimensional submanifold of Ω if M is relatively closed in Ω and for every point $x_0\in M$ there is a neighborhood U of x_0 in Ω and functions $\{\rho_{k+1},\ldots,\rho_n\}\subset\mathcal{E}(U)$ so that*

1. *$M\cap U=\big\{x\in U\,\big|\,\rho_{k+1}(x)=\cdots=\rho_n(x)=0\big\}$.*
2. *The matrix $\Big\{\frac{\partial\rho_k}{\partial x_j}(x)\Big\}$ has maximal rank $n-k$ for $x\in U$.*

We list some local properties of submanifolds which follow easily from the open mapping theorem and implicit function theorem.

Theorem 6.2. *Suppose that $M\subset\Omega$ is a k-dimensional submanifold of an open subset of $\mathbb{R}^n$.*

1. *For each $p\in M$, there is a neighborhood U of p in Ω and a diffeomorphism $\rho=(\rho_1,\ldots,\rho_n)$ from U onto an open neighborhood V of the origin in $\mathbb{R}^n$ so that*
$$\rho(M\cap U)=\Big\{(y_1,\ldots y_n)\in V\,\Big|\,y_{k+1}=\cdots=y_n=0\Big\}.$$
If ρ and $\widetilde{\rho}$ are two diffeomorphisms with this property, then $\widetilde{\rho}\circ\rho^{-1}$ is infinitely differentiable on its domain.
2. *For each $p\in M$ there is an neighborhood U of p in Ω and an infinitely differentiable mapping Φ from a neighborhood W of the origin in $\mathbb{R}^k$ to $\mathbb{R}^n$ so that $\Phi(0)=p$, and $\Phi(W)=M\cap U$. If Φ and $\widetilde{\Phi}$ are two such mappings, then $\Phi^{-1}\circ\widetilde{\Phi}$ is infinitely differentiable on its domain.*

(3) *Let $p \in M$, and let U and $\{\rho_{k+1}, \ldots, \rho_n\} \subset \mathcal{E}(U)$ be as in Definition 6.1 so that $M \cap U = \{x \in U \mid \rho_{k+1}(x) = \cdots = \rho_n(x) = 0\}$. If $f \in \mathcal{E}(U)$ with $f(x) = 0$ for all $x \in M \cap U$, then $f = \sum_{j=k+1}^{n} f_j\, \rho_j$ with $f_j \in \mathcal{E}(U)$.*

Given Theorem 6.2, we say that a real-valued function f defined on M is infinitely differentiable near a point $p \in M$ if $f \circ \Phi$ is an infinitely differentiable function defined in a neighborhood of the origin in $\mathbb{R}^k$, where Φ is given in part (2) of the theorem.

6.2. Tangential vector fields.

Now let $M \subset \Omega$ be a k-dimensional submanifold, let $X \in T(\Omega)$ and let f be a infinitely differentiable function function defined on M. We ask whether it makes sense to apply X to f. Tentatively, let us try to do this as follows. Near any point $p \in M$, we can certainly extend f to an infinitely differentiable function F defined in a neighborhood U of p in Ω (so that $F(x) = f(x)$ for all $x \in M \cap U$). We can then apply X to F, and we can restrict the result back to M. Of course, this will provide a good definition of $X[f]$ provided we can show the result does not depend on the particular choice of the extension F of the function f.

Thus suppose F_1 and F_2 both extend the function f to some open neighborhood U of $p \in M$. Then $F_1 - F_2 = 0$ on $M \cap U$. Suppose that $M \cap U = \{x \in U \mid \rho_{k+1}(x) = \cdots = \rho_n(x) = 0\}$ as in part (3) of Theorem 6.2. Then $F_1 - F_2 = \sum_{j=k+1}^{n} f_j\, \rho_j$, and hence for $x \in M \cap U$

$$\begin{aligned} X[F_1](x) - X[F_2](x) &= \sum_{j=k+1}^{n} X[f_j](x)\, \rho_j(x) + \sum_{j=k+1}^{n} f_j(x)\, X[\rho_j](x) \\ &= \sum_{j=k+1}^{n} f_j(x)\, X[\rho_j](x) \end{aligned}$$

It is now clear that for $x \in M$, the value of $X[F](x)$ is independent of the choice of extension F of f if and only if $X[\rho_j](x) = 0$ for $k+1 \leq j \leq n$.

DEFINITION 6.3. *Let $M \subset \Omega$ be a k-dimensional submanifold. A vector field $X \in T(\Omega)$ is tangential to M if $X[g](x) = 0$ for every $x \in M$ and every function $g \in \mathcal{E}(\Omega)$ with $g(y) = 0$ for all $y \in M$.*

6.3. The Frobenius theorem.

THEOREM 6.4. *Let $\{X_1, \ldots, X_p\}$ be smooth real vector fields on an open subset $\Omega \subset \mathbb{R}^n$. Suppose there is a positive integer $m < n$ so that for each $x \in \Omega$ the tangent vectors $\{(X_1)_x, \ldots, (X_p)_x\}$ span a subspace of dimension m of the tangent space T_x. Then the following two conditions are equivalent.*

(A) *For every $x \in \Omega$ there is an open neighborhood $x \in U_x \subset \Omega$ and a m-dimensional submanifold $M_x \subset U_x$ such that $x \in M_x$ and for each $y \in M_x$ the tangent vectors $\{(X_1)_y, \ldots, (X_p)_y\}$ span the tangent space $T_y(M_x)$ to M_x at y.*

(B) *For every $x \in \Omega$ there is an open neighborhood $x \in V_x \subset \Omega$ and functions $\alpha^l_{j,k} \in \mathcal{E}(V_x)$ so that on V_x we have*

$$[X_j,\, X_k] = \sum_{l=1}^{p} \alpha^l_{j,k}\, X_l. \tag{6.1}$$

Proof that (A) implies (B). Let $x \in \Omega$, and let $x \in M_x \subset U_x$ be as in (A). Since the tangent vectors $\{(X_1)_x, \dots, (X_p)_x\}$ span a subspace of dimension m, we can assume, after renumbering if necessary, that the tangent vectors $\{(X_1)_y, \dots, (X_m)_y\}$ are linearly independent at every point y in an open neighborhood $x \in V_x \subset U_x$.

Let $y \in V_x$. We can apply the hypothesis (A) again and conclude that there exists an open neighborhood $y \in W_y \subset V_x$ and a m-dimensional submanifold $y \in M_y \subset W_y$ so that the tangent vectors $\{(X_1)_z, \dots, (X_p)_z\}$ span the tangent space $T_z(M_y)$ for every $z \in M_y$. The vector fields $\{X_1, \dots, X_p\}$ are tangent to M_y, and hence every commutator $[X_j,\, X_k]$ is also tangent to M_y. In particular, the tangent vector $[X_j,\, X_k]_y \in T_y M_x$. But the tangent vectors $\{(X_1)_y, \dots, (X_m)_y\}$ are linearly independent, and hence span the m-dimensional space $T_y M_x$. It follows that for every $y \in V_x$ we can find real numbers $\{\alpha^l_{j,k}(y)\}$ so that

$$[X_j,\, X_k]_y = \sum_{l=1}^{m} \alpha^l_{j,k}(y)\, (X_l)_y.$$

It remains to show that the functions $y \to \alpha^l_{j,k}(y)$ are infinitely differentiable. However, the vector fields $\{X_1, \dots, X_m\}$ and $[X_j,\, X_k]$ are smooth. Since we now know that $[X_j,\, X_k]$ is in the linear span of $\{X_1, \dots, X_m\}$ at each point $y \in V_x$, Cramer's rule shows that the coefficients $\{\alpha^i_{j,k}\} \subset \mathcal{E}(V_x)$. □

Proof that (B) implies (A). Let $x \in \Omega$, and assume that equation (6.1) holds in a neighborhood $x \in V_x \subset \Omega$. Let us write $X_j = \sum_{k=1}^{n} a_{j,k}\, \partial_{x_k}$ where $\{a_{j,k}\} \subset \mathcal{E}(\Omega)$. Since the tangent vectors $\{(X_1)_x, \dots, (X_p)_x\}$ span a subspace of dimension m, the rank of the matrix $\{a_{j,k}(x)\}$, $1 \le j \le p$, $1 \le k \le n$ is m. Thus after renumbering the vector fields and relabeling the coordinates if necessary, we can assume that the tangent vectors $\{(X_1)_x, \dots, (X_m)_x\}$ are linearly independent and

$$\det \begin{bmatrix} a_{1,1}(x) & \cdots & a_{1,m}(x) \\ \vdots & \ddots & \vdots \\ a_{m,1}(x) & \cdots & a_{m,m}(x) \end{bmatrix} \neq 0.$$

It follows that there is an open neighborhood $x \in U_x \subset V_x$ so that if $y \in U_x$, the tangent vectors $\{(X_1)_y, \dots, (X_m)_y\}$ are linearly independent and the $m \times m$ matrix $\big[a_{j,k}(y)\big]_{j,k=1}^{m}$ is invertible. Let $\big[b_{j,k}(y)\big]_{j,k=1}^{m}$ be the inverse matrix. Then $b_{j,k} \in \mathcal{E}(U_x)$. Note that since the tangent vectors $\{(X_1)_y, \dots, (X_p)_y\}$ span a subspace of dimension m and $\{(X_1)_y, \dots, (X_m)_y\}$ are linearly independent, we can write $(X_l)_y = \sum_{i=1}^{m} c_{l,i}(y)\, (X_i)_y$ for $m+1 \le l \le p$. As before, it follows from Cramer's rule that the functions $c_{l,i} \in \mathcal{E}(U_x)$. Thus we have

$$X_l = \sum_{i=1}^{m} c_{l,i}\, X_i \qquad m+1 \le l \le p. \tag{6.2}$$

We now construct new vector fields $\{Y_1, \ldots, Y_m\}$ on U_x by setting

$$\begin{aligned} Y_l &= \sum_{j=1}^{m} b_{l,j} X_j \qquad (6.3)\\ &= \sum_{k=1}^{n} \Big[\sum_{j=1}^{m} b_{l,j}\, a_{j,k} \Big] \partial_{x_k} \end{aligned}$$

It follows that

$$Y_l = \partial_{x_l} + \sum_{k=m+1}^{n} c_{l,k}\, \partial_{x_k} \tag{6.4}$$

where $c_{l,k} = \sum_{j=1}^{m} b_{l,j}\, a_{j,k}$ for $m+1 \leq k \leq n$.

The two sets of vector fields $\{X_1, \ldots, X_p\}$ and $\{Y_1, \ldots, Y_m\}$ span the same subset of T_y for any $y \in U_x$. Equation (6.3) shows that $(Y_l)_y$ is in the span of $\{X_1, \ldots, X_p\}$. Next, using equation (6.3) and the fact that $[a_{j,k}]$ and $[b_{j,k}]$ are inverse matrices, we can write

$$X_j = \sum_{k=1}^{m} a_{j,k}\, Y_k, \qquad 1 \leq j \leq m, \tag{6.5}$$

Also, it follows from equation (6.2) that

$$X_j = \sum_{k=1}^{m} \Big[\sum_{l=1}^{m} c_{j,l}\, a_{l,k} \Big] Y_k, \qquad m+1 \leq j \leq p. \tag{6.6}$$

Thus equations (6.5) and (6.6) show that $(X_j)_y$ is in the span of $\{Y_1, \ldots, Y_m\}$.

Now if we use the hypothesis given in (6.1), equations (6.3), (6.5), and (6.6), and Proposition 1.4, it follows that we can write

$$[Y_j,\, Y_k] = \sum_{l=1}^{m} \beta_{j,k}^{l}\, Y_l$$

for $1 \leq j, k \leq m$. However, it follows from equation (6.4) that for $1 \leq l \leq m$, the coefficient of ∂_{x_l} in $[Y_j,\, Y_k]$ is zero, and hence $\beta_{j,k}^{l} \equiv 0$. Hence $[Y_j,\, Y_k] = 0$, and the vector fields $\{Y_1, \ldots, Y_m\}$ all commute on the set U_x.

Now define a mapping Φ from a neighborhood of the origin in $\mathbb{R}^m$ to Ω by setting

$$\Phi(y_1, \ldots, y_m) = \exp(y_1 Y_1) \circ \cdots \circ \exp(y_n Y_n)(x).$$

Note that $\Phi(0) = x$. The Jacobian determinant of this mapping at the origin of $\mathbb{R}^m$ is $\det(Y_1, \ldots, Y_m)(x) \neq 0$, and so there is a neighborhood $x \in U_x \subset W_x$ so that the image of some neighborhood of the origin in $\mathbb{R}^m$ is an m-dimensional submanifold $M_x \subset U_x$.

We claim that $d\Phi[\partial_{y_j}] = Y_j$. Since $[Y_j,\, Y_k] = 0$, all of the flows

$$\{\exp(y_1 Y_1), \ldots, \exp(y_n Y_n)\}$$

commute. Thus for any j we can write

$$\Phi(y_1, \ldots, y_m) = \exp(y_j Y_j) \circ \exp(y_1 Y_1) \circ \cdots \circ \exp(y_n Y_n)(x)$$

where the flow $\exp(y_jY_j)$ has been moved to the left. It follows from Corollary 3.7 that

$$\begin{aligned}d\Phi[\partial_{y_j}][f]\big(\Phi(y)\big) &= d\Phi[\partial_{y_j}][f]\big(\exp(y_jY_j)\circ\exp(y_1Y_1)\circ\cdots\circ\exp(y_nY_n)(x)\big)\\&= \frac{\partial}{\partial y_j}\Big[f\big(\exp(y_jY_j)\circ\exp(y_1Y_1)\circ\cdots\circ\exp(y_nY_n)(x)\big)\Big]\\&= [Y_jf]\big(\exp(y_jY_j)\circ\exp(y_1Y_1)\circ\cdots\circ\exp(y_nY_n)(x)\big)\\&= [Y_jf]\big(\Phi(y)\big).\end{aligned}$$

It follows that the vector fields $\{Y_1,\dots,Y_m\}$ are tangent to M_x, and consequently the vector fields $\{X_1,\dots,X_p\}$ are tangent to M_x, which completes the proof. □

7. Normal forms for vector fields

The discussion in Section 2 shows that the coefficients of a vector field change if one makes a change of coordinate. In this section we study changes of variables which allow us to write one or two vector fields in particularly simple form.

7.1. The case of a single vector field.

A single non-vanishing vector field is, after a change of coordinates, just the derivative with respect to one of the coordinates.

PROPOSITION 7.1. *Let $X\in T(\Omega)$ and let $x\in\Omega$. Suppose that $X_x\neq 0$. Then there is an open neighborhood U of the point x and a diffeomorphism $\Phi: U\to V$ where V is an open neighborhood of the origin in $\mathbb{R}^n$ so that $\Phi(x)=0$, and so that if $(y_1,\dots,y_n)$ are coordinates in V, then*

$$d\Phi[X] = \frac{\partial}{\partial y_n}. \tag{7.1}$$

Before starting the proof, let us investigate the meaning of (7.1). Suppose that $\Phi=(\varphi_1,\dots,\varphi_n)$ is a diffeomorphism satisfying Proposition 7.1. It is easy to check that

(i) the functions $\{\varphi_1,\dots,\varphi_{n-1}\}$ are constant along each integral curve of X;

(ii) the function $t\to\varphi_n\big(\exp(tX)(x)\big)$ must equal a constant plus t.

Conversely, any diffeomorphism Φ satisfying (i) and (ii) will satisfy (7.1). With this in mind, we now turn to the proof of the proposition.

PROOF. To construct the required diffeomorphism, it is easier to first construct its inverse. Let $\widetilde{\Psi}:\mathbb{R}^{n-1}\to\mathbb{R}^n$ be a smooth mapping such that $\widetilde{\Psi}(0)=x$. Let $\widetilde{V}$ be a sufficiently small neighborhood of the origin such that $\widetilde{\Psi}(\widetilde{V})$ has compact closure $K\subset\Omega$. According to Theorem 3.2 there exists $\epsilon>0$ so that $\exp\big(tX)(y)$ is defined for $y\in K$ and $|t|<\epsilon$.

In particular $\exp\big(tX)\big(\widetilde{\Psi}(t_1,\dots,t_{n-1})\big)$ is defined for $(t_1,\dots,t_{n-1})\in\widetilde{V}$, and $|t_n|<\epsilon$. Define $\Psi:\widetilde{V}\times(-\epsilon,+\epsilon)\to\mathbb{R}^n$ by

$$\Psi(t_1,\dots,t_n) = \exp(tX)\big(\widetilde{\Psi}(t_1,\dots,t_{n-1}),t_n\big),$$

It follows from Theorem 3.2 that Ψ is a smooth mapping. Moreover, it follows from Lemma 3.4 that

$$d\Psi_x\Big(\frac{\partial}{\partial t_k}\Big) = \begin{cases} d\widetilde{\Psi}_x\big(\frac{\partial}{\partial t_k}\big) & \text{for } 1 \leq k \leq n-1 \\ \\ X_x & \text{for } k = n. \end{cases}$$

If we choose any $\widetilde{\Psi}$ so that the Jacobian matrix $J\widetilde{\Psi}(x)$ at x has rank $n-1$, and so that X_x is not in the range of $J\widetilde{\Psi}(x)$, then the inverse function theorem implies that Ψ is a diffeomorphism on a sufficiently small neighborhood of 0. The inverse mapping satisfies (i) and (ii), and this proves the proposition. □

7.2. The case of two vector fields.

After a change of variables, any two non-vanishing vector fields are locally identical. The situation is more complicated if we try to find a standard form for a pair of vector fields.

PROPOSITION 7.2. *Let $X, Y \in T(\Omega)$ and let $x_0 \in \Omega$. Suppose that $X(x_0)$ and $Y(x_0)$ are linearly independent. Then there is an open neighborhood U of the point x_0 and a diffeomorphism $\Phi : U \to V$ where V is an open neighborhood of the origin in $\mathbb{R}^n$ so that $\Phi(x_0) = 0$, and so that if $(y_1, \ldots, y_n)$ are coordinates in V, then*

$$\begin{aligned} \Phi_*[X] &= b_n(y)\,\frac{\partial}{\partial y_n}, \\ \Phi_*[Y] &= \sum_{j=1}^{n-1} b_j(y)\,\frac{\partial}{\partial y_j}. \end{aligned} \tag{7.2}$$

Moreover $b_n(0) \neq 0$.

PROOF. We must have $X(x_0) \neq 0$. According to Proposition **??** we can assume that $x_0 = 0$ and that $X = \frac{\partial}{\partial x_n}$. Let $Y = \sum_{j=1}^{n} a_j \frac{\partial}{\partial x_j}$. Let $\Phi(x_1, \ldots, x_n) = \big(x_1, \ldots, x_{n-1}, \varphi(x_1, \ldots, x_n)\big)$. Then at $y = \Phi(x)$ we have

$$\Phi^*[X] = \frac{\partial \varphi}{\partial x_n}(x)\,\frac{\partial}{\partial y_n} \quad \text{and} \quad \Phi^*[Y] = \sum_{j=1}^{n-1} a_j(x)\,\frac{\partial}{\partial y_j} + Y[\varphi](x)\,\frac{\partial}{\partial y_n}$$

Thus we only need to choose φ so that it is constant along the integral curves of Y, and such that $\frac{\partial \varphi}{\partial x_n}(0) \neq 0$. But this is possible because the integral curves of X and of Y are not tangent anywhere near 0. □

8. Vector fields and Derivations

We return now to the question of finding an algebraic characterizing of tangent vectors and vector fields. As promised in Section 1, we show that this can be done using the concept of derivation. If A_1 and A_2 are two algebras over $\mathbb{R}$, a linear mapping $D : A_1 \to A_2$ is a derivation if $D(xy) = D(x)y + xD(y)$ for all $x, y \in A_1$.

8.1. Factorization.

The key to the desired characterization is the following simple but important result on factorization of smooth functions.

PROPOSITION 8.1. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $y = (y_1, \ldots, y_n) \in \Omega$. Let $f \in \mathcal{E}(U)$. Then for $1 \leq k, l \leq n$ there exist functions $f_{k,l} \in \mathcal{E}(\Omega)$ so that for $x \in \Omega$ we have*

$$f(x) = f(y) + \sum_{j=1}^{n}(x_j - y_j)\,\frac{\partial f}{\partial x_j}(y) + \sum_{k,l=1}^{n}(x_k - y_k)(x_l - y_l)\,f_{k,l}(y).$$

PROOF. Suppose that $B(y,r) = \big\{x \in \mathbb{R}^n \,\big|\, |x-y| < r\big\} \subset \Omega$. Let $x \in B(y,r)$, and for $0 \leq t \leq 1$ put $\psi(t) = f\big(y + t(x-y)\big)$. Then

$$f(x) = \psi(1) = \psi(0) + \int_0^1 \psi'(s)\,ds = \psi(0) + \psi'(0) + \int_0^1 (1-u)\,\psi''(u)\,du.$$

Now

$$\begin{aligned}\psi(0) &= f(y)\\ \psi'(0) &= \sum_{j=1}^{n}(x_j - y_j)\,\frac{\partial f}{\partial x_j}(y)\\ \psi''(u) &= \sum_{k,l=1}^{n}(x_k - y_k)(x_l - y_l)\,\frac{\partial f}{\partial x_k \partial x_l}\big(y + u(x-y)\big).\end{aligned}$$

If we put

$$\widetilde{f}_{k,l}(x) = \int_0^1 (1-u)\,\frac{\partial f}{\partial x_k \partial x_l}\big(y + u(x-y)\big)\,du,$$

then the functions $\{\widetilde{f}_{k,l}\}$ are infinitely differentiable on $B(y,r)$, and for $|x-y| < r$ we have

$$f(x) = f(y) + \sum_{j=1}^{n}(x_j - y_j)\,\frac{\partial f}{\partial x_j}(y) + \sum_{k,l=1}^{n}(x_k - y_k)(x_l - y_l)\,\widetilde{f}_{k,l}(x).$$

Choose $\chi \in \mathcal{E}(\Omega)$ with $\chi(x) \equiv 1$ for $|x-y| < \frac{1}{2}r$ and $\chi(x) \equiv 0$ for $|x-y| \geq \frac{3}{4}4$. The functions $\{\chi\widetilde{f}_{k,l}\}$ are identically zero for $|x-y| \geq \frac{3}{4}r$ and hence extend to infinitely differentiable functions on Ω which we continue to write as $\chi\widetilde{f}_{k,l}$. Then

$$\widetilde{F}(x) = f(x) - f(y) - \sum_{j=1}^{n}(x_j - y_j)\,\frac{\partial f}{\partial x_j}(y) - \sum_{k,l=1}^{n}(x_k - y_k)(x_l - y_l)\,\chi(x)\,\widetilde{f}_{k,l}(x)$$

is an infinitely differentiable function on Ω and is identically zero for $|x-y| \leq \frac{1}{2}r$. Hence the function $F(x) = \widetilde{F}(x)\,|x-y|^{-2}$ is also infinitely differentiable on Ω. Thus we can write

$$\begin{aligned}f(x) = f(y) &+ \sum_{j=1}^{n}(x_j - y_j)\,\frac{\partial f}{\partial x_j}(y)\\ &+ \sum_{k,l=1}^{n}(x_k - y_k)(x_l - y_l)\,(\chi\,\widetilde{f}_{k,l})(x) + \sum_{j=1}^{n}(x_j - y_j)^2\,F(x).\end{aligned}$$

If we now put

$$f_{k,l}(x) = \begin{cases} \chi(x)\, \widetilde{f}_{k,l}(x) & \text{if } k \neq l \\ \chi(x)\, \widetilde{f}_{j,j}(x) + F(x) & \text{if } k = l = j \end{cases}$$

then we obtain the decomposition asserted in the proposition. □

8.2. Derivations.

LEMMA 8.2. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let $D : \mathcal{E}(\Omega) \to \mathcal{E}(\Omega)$ be a derivation. Then there exists a unique n-tuple $(a_1, \ldots, a_n) \in \mathcal{E}(\Omega)^n$ so that for all $f \in \mathcal{E}(\Omega)$ we have $D[f](x) = \sum_{j=1}^n a_j(x) \frac{\partial f}{\partial x_j}(x)$. In particular, D is a vector field.*

PROOF. Since D is a derivation, $D[1] = D[1^2] = D[1]\,1 + 1\,D[1] = 2D[1]$, so that $D[1] = 0$. By linearity, $D[c] = 0$ for any function which is constant. Next, if f, g, $h \in \mathcal{E}(\Omega)$, then $D[fgh](y) = D[f]\,g(y)\,h(y) + f(y)\,D[g]\,h(y) + f(y)\,g(y)\,D[h]$. Hence if $y \in \Omega$ and if $f(y) = g(y) = 0$, then $D[fgh](y) = 0$.

Now put $a_j = D[x_j]$. Then $a_j \in \mathcal{E}(\Omega)$, and it follows from Proposition 8.1 and the above remarks that

$$D[f](y) = \sum_{j=1}^n a_j(y)\, \frac{\partial f}{\partial x_j}(y).$$

This completes the proof. □

Exactly the same argument gives the following result.

LEMMA 8.3. *Let $x \in \mathbb{R}^n$ and let $D : \mathcal{E}_x \to \mathbb{R}$ be a derivation. Then there exists a unique $(c_1, \ldots, c_n) \in \mathbb{R}^n$ so that $D[f] = \sum_{j=1}^n c_j \frac{\partial f}{\partial x_j}(x)$. In particular, D is a tangent vector.*

CHAPTER 3

CARNOT-CARATHÉODORY METRICS

A Carnot-Carathéodory space is a smooth real manifold M equipped with a metric induced by a distinguished family of vector fields. There are several essentially equivalent definitions of this metric, and a typical approach is the following. Let $X_1, \ldots, X_p \in T(M)$ be smooth real vector fields on M. If $[a, b] \subset \mathbb{R}$ is an interval and if $\gamma : [a, b] \to M$ is absolutely continuous, then γ is said to be a *sub-unit* curve joining $\gamma(a)$ and $\gamma(b)$ if for almost all $t \in [a, b]$ one can write $\gamma'(t) = \sum_{j=1}^{p} a_j(t)\,(X_j)_{\gamma(t)}$ with $\sum_{j=1}^{p} |a_j(t)|^2 \leq 1$. The Carnot-Carathéodory distance between two points $p, q \in M$ is then defined to be the infimum of those $\delta > 0$ such that there is a sub-unit mapping joining p and q defined on an interval of length δ.

In general, if the vector fields $\{X_1, \ldots, X_p\}$ do not span the tangent space at each point of M, it may be happen that the Carnot-Carathéodory distance between two points is infinite. There may be *no* path $\gamma : [a, b] \to M$ joining p and q whose derivative $\gamma'(t)$ lies in the space spanned by $\{(X_1)_{\gamma(t)}, \ldots, (X_p)_{\gamma(t)}\}$ for almost every $t \in [a, b]$. However, if the vector fields $\{X_1, \ldots, X_p\}$ together with all their iterated commutators span the tangent space to M at each point, then the distance between any two points is finite, and this distance is a metric. This result goes back at least to the work of Chow [**Cho40**].

The following (imprecise) argument gives a rough idea of why this is so. The crucial point is that if X and Y are two vector fields on M, then motion along the integral curves of X and Y also allows permits motion in the direction of the commutator $[X, Y]$. Thus it follows from the Campbell-Baker-Hausdorff formula, (see Chapter 2, Lemma 4.5) that

$$\exp[-tX]\,\exp[-tY]\,\exp[tX]\,\exp[tY] = \exp[t^2[X, Y] + O(t^3)].$$

Thus we can flow (approximately) along the integral curve of $[X, Y]$ by flowing along Y, then flowing along X, then flowing along $-Y$, and finally flowing along $-X$. We can flow (approximately) along integral curves of higher order commutators by a more complicated composition[1] of flows along Y and X. Since the iterated commutators span the tangent space, we can flow in any direction.

In this Chapter we develop a theory of metrics constructed from vector fields, but we shall start with a different definition in which it is immediately clear that

[1]For example, the Campbell-Baker-Hausdorff formula gives

$$\exp[-tX]\,\exp[-t^2\,[X, Y]]\,\exp[tX]\,\exp[t^2\,[X, Y]] = \exp[t^3[X, [X, Y]] + O(t^4)].$$

Using again the Campbell-Baker-Hausdorff formula for $\exp t^2\,[X, Y]$, we have

$$\exp[t^3[X, [X, Y]] = \exp[tY]\,\exp[-tX]\,\exp[-2tY]\,\exp[tX]\,\exp[tY] + O(t^4).$$

Thus we can flow (approximately) along the integral curve of the vector field $[X, [X, Y]]$ by a composition of five elementary flows along integral curves of Y and X.

the distance between any two points is finite. We will later prove Chow's theorem, and show that that the different metrics are equivalent. In brief, our construction of Carnot-Carathéodory metrics proceeds as follows. Let $\{Y_1, \ldots, Y_q\}$ be vector fields on M which span the tangent space at each point. Then

- The vector fields $\{Y_1, \ldots, Y_q\}$ induce a 'length' function on each tangent space T_xM. In general, this length will not be a norm.
- If $\gamma : [0,1] \to M$ is an absolutely continuous mapping, then for almost all $t \in [0,1]$, we can calculate the length of the tangent vector $\gamma'(t) \in T_{\gamma(t)}M$. We then define the 'length' of γ to the the essential supremum of the lengths of the derivatives $\gamma'(t)$.
- The distance between two points in M is then the infimum over the lengths of all absolutely continuous curves $\gamma : [0,1] \to M$ joining them.

Note that this construction involves a very large class of curves. This has the advantage that it is then relatively easy to show that the resulting distance function is actually a metric. However, the corresponding disadvantage is that it is often difficult to calculate the metric exactly, or to understand the geometry of the corresponding family of balls. Thus the general theory develops in several stages. We first deal with a large class of curves in order to establish the global metric properties of the distance function. Later we show that an equivalent metric is generated by a smaller classes of curves, and this allows a precise description of local geometry of the metric.

Since we are primarily interested in local questions we shall take the underlying manifold to be a connected open subset of Euclidean space. However there is no difficulty in extending the global theory to the manifold setting.

1. Construction of Carnot-Carathéodory metrics

1.1. Homogeneous norms.

A *non-isotropic family of dilations* on $\mathbb{R}^q$ is a family of mappings $D_\delta : \mathbb{R}^q \to \mathbb{R}^q$ of the form

$$D_\delta(y) = D_\delta\big((y_1, \ldots, y_q)\big) = (\delta^{d_1} y_1, \ldots, \delta^{d_q} y_q). \tag{1.1}$$

Note that $D_{\delta_1} \circ D_{\delta_2} = D_{\delta_1+\delta_2}$. We shall assume that the *exponents* $\{d_j\}$ are all strictly positive. A $\{D_\delta\}$*-homogeneous norm* is then a continuous function on $\mathbb{R}^q$, written $x \to \big|\big|\, x \,\big|\big|$, such that

(i) for all $y \in \mathbb{R}^q$, $\big|\big|\, y \,\big|\big| \geq 0$ and $\big|\big|\, y \,\big|\big| = 0$ if and only if $y = 0$;

(ii) for all $y \in \mathbb{R}^q$, $\big|\big|\, y \,\big|\big| = \big|\big| - y \,\big|\big|$;

(iii) for all $y \in \mathbb{R}^q$ and all $\delta > 0$, $\big|\big|\, D_\delta(y) \,\big|\big| = \delta \,\big|\big|\, y \,\big|\big|$.

If the family of dilations $\{D_\delta\}$ is understood, we shall simply call $||\cdot||$ a *homogeneous norm.*

Given the family $\{D_\delta\}$, there are many $\{D_\delta\}$-homogeneous norms. For example, for $1 \leq p \leq \infty$, we can set

$$||y||_p = \begin{cases} \Big[\sum_{j=1}^q |y_j|^{\frac{p}{d_j}}\Big]^{\frac{1}{p}} & \text{if } p < \infty, \\ \\ \sup_{1\leq j\leq q} |y_j|^{\frac{1}{d_j}} & \text{if } p = \infty. \end{cases} \tag{1.2}$$

When all the exponents $d_j = 1$, these are the standard L^p-norms on $\mathbb{R}^q$. We will write

$$|y| = \left(y_1^2 + \cdots + y_q^2\right)^{\frac{1}{2}}$$

for the standard Euclidean norm, which is the case $d_j = 1$ and $p = 2$.

As with an ordinary norm, a homogeneous norm on $\mathbb{R}^q$ is determined by its restriction to the unit (Euclidean) sphere $S^{q-1} = \left\{y \in \mathbb{R}^q \,\middle|\, |y| = 1\right\}$. Since S^{q-1} is compact, it follows that any two homogeneous norms are equivalent.

Proposition 1.1. *Let $\{D_\delta\}$ be a family of dilations on $\mathbb{R}^q$, and let $\big|\big| \cdot \big|\big|_a$ and $\big|\big| \cdot \big|\big|_b$ be two D_δ-homogeneous norms. Then there exists a constant $C \geq 1$ so that for all $y \in \mathbb{R}^q$,*

$$C^{-1} \big|\big| \, y \, \big|\big|_b \leq \big|\big| \, y \, \big|\big|_a \leq C \big|\big| \, y \, \big|\big|_b.$$

We can also compare a homogeneous norm to the standard Euclidean norm.

Proposition 1.2. *Let $\{D_\delta\}$ be a family of dilations on $\mathbb{R}^q$ with exponents $\{d_1, \ldots, d_q\}$. Let $d = \min\{d_1, \ldots, d_q\}$ and $D = \max\{d_1, \ldots, d_q\}$. Let $\big|\big| \cdot \big|\big|$ be a D-homogeneous norm. Then there is a constant $C \geq 1$ so that for all $y \in \mathbb{R}^q$,*

$$C^{-1} \min\left\{|y|^{1/d}, |y|^{1/D}\right\} \leq \big|\big| \, y \, \big|\big| \leq C \max\left\{|y|^{1/d}, |y|^{1/D}\right\}.$$

Proof. Since $y \to ||y||$ is continuous, there is a constant $C \geq 1$ so that $C^{-1} \leq ||y|| \leq C$ if $|y| = 1$. Given an arbitrary $y \in \mathbb{R}^q$ with $y \neq 0$, there is a unique $\delta_y > 0$ so that $|D_{\delta_y}(y)| = 1$, and hence $C^{-1}\,\delta_y^{-1} \leq ||y|| \leq C\,\delta_y^{-1}$. Now if $y = (y_1, \ldots y_q)$, δ_y is the solution of the equation

$$\delta_y^{2d_1}|y_1|^2 + \cdots + \delta_y^{2d_q}|y_q|^2 = 1.$$

It follows that

$$\begin{cases} \delta_y^{2D}|y|^2 \leq 1 \leq \delta_y^{2d}|y|^2 & \text{if } \delta_y \leq 1, \\ \\ \delta_y^{2d}|y|^2 \leq 1 \leq \delta_y^{2D}|y|^2 & \text{if } \delta_y \geq 1. \end{cases}$$

But these are equivalent to the inequalities

$$\begin{cases} |y|^{1/D} \leq \delta_y^{-1} \leq |y|^{1/d} & \text{if } |y| \geq 1, \\ \\ |y|^{1/d} \leq \delta_y^{-1} \leq |y|^{1/D} & \text{if } |y| \leq 1, \end{cases}$$

and this gives the desired estimate. □

1.2. Control systems.

Let $\Omega \subset \mathbb{R}^n$ be a connected open set. We shall always assume that there is a constant $M < +\infty$ so that for all $x, y \in \Omega$ there is a continuously differentiable mapping $\varphi : [0, 1] \to \Omega$ with $\varphi(0) = x$, $\varphi(1) = y$, and for all $t \in [0, 1]$

$$|\varphi'(t)| \leq M\,|y - x|, \tag{1.3}$$

This is a smoothness assumption on the boundary $\partial\Omega$ of Ω, and it is always satisfied if the closure $\overline{\Omega}$ of Ω is compact and $\partial\Omega$ is a smooth hypersurface.

DEFINITION 1.3. *A control system on* Ω *is a set of smooth real vector fields* $Y_1, \ldots, Y_q \in T(\Omega)$ *and positive integers* $\{d_1, \ldots, d_q\}$ *called the* formal degrees *of the vector fields such that the following conditions are satisfied:*

(H-1) *For every* $x \in \Omega$, *the tangent vectors* $\{(Y_1)_x, \ldots, (Y_q)_x\}$ *span the tangent space* T_x.

(H-2) *For* $1 \leq k, l, m \leq q$ *there are functions* $c_{j,k}^m \in \mathcal{E}(\Omega)$ *so that*

$$[Y_j,\, Y_k] = \sum_{\substack{1 \leq m \leq q \\ d_m \leq d_j + d_k}} c_{j,k}^m\, Y_m. \tag{1.4}$$

Thus $c_{k,l}^m(x) \equiv 0$ *for all* $x \in \Omega$ *unless* $d_m \leq d_k + d_l$. *We sometimes refer to the functions* $\{c_{j,k}^m\}$ *as the* structure functions *for* $\mathcal{Y}$..

Because issues of uniformity will be very important, we need a method for measuring the 'size' of a control system $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$. We measure the size of the formal degrees by using notation already introduced in Proposition 1.2:

$$\begin{aligned} d_{\mathcal{Y}} = d &= \min_{1 \leq j \leq q} d_j, \\ D_{\mathcal{Y}} = D &= \max_{1 \leq j \leq q} d_j. \end{aligned} \tag{1.5}$$

Next let us write

$$Y_j = \sum_{k=1}^n b_{j,k} \frac{\partial}{\partial x_k} = \sum_{k=1}^n b_{j,k}\, \partial_{x_k} \tag{1.6}$$

where $b_{j,k} \in \mathcal{E}(\Omega)$. Then there are three additional quantities we need to control. We need estimates on the size of derivatives of the coefficient functions $\{b_{j,k}\}$ from equation (1.6), we need estimates on the size of derivatives of the structure functions $\{c_{j,k}^m\}$ from equation (1.4), and we need to quantify the hypothesis (H-1) that the vectors $\{(Y_j)_x\}$ span the tangent space T_x for each $x \in \Omega$.

DEFINITION 1.4.

(1) *For any set* $E \subset \Omega$ *and every positive integer* N, *set*

$$\big|\big|\, \mathcal{Y} \,\big|\big|_{E,N} = \sup_{x \in E} \sum_{|\alpha| \leq N} \sum_{j,k} |\partial^\alpha b_{j,k}(x)| + \sup_{x \in E} \sum_{|\alpha| \leq N} \sum_{k,l,m} |\partial^\alpha c_{k,l}^m(x)|.$$

Note that if $K \subset \Omega$ *is compact, then* $\big|\big|\, \mathcal{Y} \,\big|\big|_{K,N}$ *is finite.*

(2) *For any set* $E \subset \Omega$, *put*

$$\nu_{\mathcal{Y}}(E) = \nu(E) = \inf_{x \in E} \sup_{1 \leq i_i < \cdot < i_n \leq q} \big| \det(Y_{i_1}, \ldots, Y_{i_n})(x) \big|.$$

Note that hypothesis (H-1) is equivalent to the statement that for every compact subset $K \subset \Omega$ *we have* $\nu(K) > 0$.

1.3. Carnot-Carathéodory metrics.

Let $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ be a control system on a connected open set $\Omega \subset \mathbb{R}^n$. Let $\{D_\delta\}$ be the family of non-isotropic dilations on $\mathbb{R}^q$ with exponents $\{d_1, \ldots, d_q\}$, and let $\big|\big| \cdot \big|\big|$ be a D-homogeneous norm on $\mathbb{R}^q$. In this section we define the Carnot-Carathéodory metric associated to control system $\mathcal{Y}$ and the norm $\big|\big| \cdot \big|\big|$.

To do this we first introduce a class of allowable paths joining two point of Ω.

DEFINITION 1.5. *For $x, y \in \Omega$ and $\delta > 0$, $AC(x, y; \delta)$ denotes the set of all absolutely continuous mappings $\phi : [0, 1] \to \Omega$ with $\phi(0) = x$ and $\phi(1) = y$ such that for almost every $t \in [0, 1]$ we can write*

$$\phi'(t) = \sum_{j=1}^{q} a_j(t)\,(Y_j)_{\phi(t)} \qquad \textit{with} \qquad \big|\big| \big(a_1(t), \ldots, a_q(t)\big) \big|\big| < \delta.$$

Next we define the Carnot-Carathéodory distance ρ and the corresponding family of balls $B_\rho(x; \delta)$.

DEFINITION 1.6. *The Carnot-Carathéodory distance between two points $x, y \in \Omega$ induced by $\mathcal{Y}$ and the homogeneous norm $\big|\big| \cdot \big|\big|$ is given by*

$$\rho(x, y) = \inf \Big\{ \delta > 0 \,\Big|\, AC(x, y; \delta) \text{ is not empty} \Big\}.$$

For $x \in \Omega$ and $\delta > 0$, the ball with center x and radius $\delta > 0$ is

$$B_\rho(x; \delta) = \Big\{ y \in \Omega \,\Big|\, \rho(x, y) < \delta \Big\}.$$

PROPOSITION 1.7. *Let $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots d_q\}$ be a control system on an open set $\Omega \subset \mathbb{R}^n$, let $\big|\big| \cdot \big|\big|$ be a homogeneous metric on $\mathbb{R}^q$. Suppose that $\big|\big| \mathcal{Y} \big|\big|_{\Omega,1} < +\infty$ and $\nu_{\mathcal{Y}}(\Omega) > 0$. Then:*

(1) There is a constant $C \geq 1$ so that for all $x, y \in \Omega$,
$$C^{-1} \min \big\{ |x - y|^{\frac{1}{d}},\, |x - y|^{\frac{1}{D}} \big\} \leq \rho(x, y) \leq C \max \big\{ |x - y|^{\frac{1}{d}},\, |x - y|^{\frac{1}{D}} \big\}.$$

(2) ρ is a metric. Explicitly:
- *(a) For all $x, y \in \Omega$, $0 \leq \rho(x, y) < +\infty$, and $\rho(x, y) = 0$ if and only if $x = y$;*
- *(b) For all $x, y \in \Omega$, $\rho(x, y) = \rho(y, x)$;*
- *(c) For all $x, y, z \in \Omega$, $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$.*

(3) The function ρ is jointly continuous on $\Omega \times \Omega$. In particular, the balls $\{B_\rho(x, \delta)\}$ for $x \in \Omega$ and $\delta > 0$ are open subsets of $\mathbb{R}^n$.

PROOF. We first compare $\rho(x, y)$ with the Euclidean distance $|x - y|$. It follows from the hypothesis in equation (1.3) that if $x, y \in \Omega$ there is a continuously differentiable function $\phi : [0, 1] \to \Omega$ such that $\phi(0) = x$, $\phi(1) = y$, and $|\phi'(t)| \leq M|y - x|$. Since the vectors $\{(Y_1)_{\phi(t)}, \ldots, (Y_q)_{\phi(t)}\}$ span $\mathbb{R}^n$, we can write $\phi'(t) = \sum_{k=1}^{q} a_k(t)\,(Y_k)_{\phi(t)}$. Moreover, Cramer's rule implies that we can choose the coefficients $\{a_k(t)\}$ so that $|a_k(t)| \leq C|\varphi'(t)| \leq C_1 M|x - y|$ where C_1 depends on the measures $||\mathcal{Y}||_{\Omega,1}$ and $\nu_{\mathcal{Y}}(\Omega)$ given in Definition 1.4. It follows that

$$\big|\big(a_1(t), \ldots, a_q(t)\big)\big| \leq \sqrt{q}\, C_1 M |x - y|,$$

and hence by Proposition 1.2 there is a constant C_2 depending on C_1, M, and the homogeneous norm, so that

$$\big|\big|\big(a_1(t),\dots,a_q(t)\big)\big|\big| \leq C_2 \max\{|x-y|^{1/d}, |x-y|^{1/D}\}.$$

It follows that $AC(x,y,C_2\max\{|x-y|^{1/d},|x-y|^{1/D}\})$ is not empty, and hence

$$\rho(x,y) \leq C_2 \max\big\{|x-y|^{1/d}, |x-y|^{1/D}\big\}. \tag{A}$$

In particular $\rho(x,y) < \infty$ for any two points $x,y \in \Omega$.

Suppose next that $\rho(x,y) = \delta < +\infty$. Then given any $\epsilon > 0$, there exists $\psi \in AC(x,y,\delta+\epsilon)$. Since ψ is absolutely continuous, it follows that

$$\begin{aligned}
|y-x| = \left|\int_0^1 \psi'(t)\,dt\right| &= \left|\int_0^1 \sum_{j=1}^{q} a_j(t)(Y_j)_{\psi(t)}\,dt\right| \\
&\leq C_3 \sup_{0\leq t\leq 1} \big|\big(a_1(t),\dots,a_q(t)\big)\big| \\
&\leq C_3 \sup_{0\leq t\leq 1} \max\{||\big(a_1(t),\dots,a_q(t)\big)||^d, ||\big(a_1(t),\dots,a_q(t)\big)||^D\} \\
&< C_3 \max\{(\delta+\epsilon)^d, (\delta+\epsilon)^D\}
\end{aligned}$$

where $C_3 \geq 1$ depends on $\big|\big|\,\mathcal{Y}\,\big|\big|_1$. Since $\epsilon > 0$ is arbitrary, it follows that

$$|x-y| \leq C_3 \max\{\rho(x,y)^d, \rho(x,y)^D\},$$

which implies

$$C_3^{-1/d} \min\{|x-y|^{1/d}, |x-y|^{1/D}\} \leq \rho(x,y). \tag{B}$$

The two inequalities (A) and (B) give the desired comparison asserted in part (1) of the Proposition.

Next we show that the function ρ is indeed a metric. It follows from the comparison with the Euclidean distance that if $x,y \in \Omega$, then $\rho(x,y) = 0$ if and only if $x = y$. Also, if $\varphi \in AC(x,y,\delta)$, then $\psi(t) = \varphi(1-t)$ has the property that $\psi \in AC_\Omega(y,x,\delta)$. It follows that $\rho(x,y) = \rho(y,x)$.

It remains to check the triangle inequality (c). Let $x,y,z \in \Omega$. Choose any δ_1, δ_2 so that $\rho(x,y) < \delta_1$ and $\rho(y,z) < \delta_2$. Then we can choose $\varphi_1 \in AC(x,y,\delta_1)$ and $\varphi_2 \in AC(y,z,\delta_2)$. For $i = 1,2$ we have

$$\varphi_i'(t) = \sum_{k=1}^{q} a_{i,k}(t)\,(Y_k)_{\varphi_i(t)}$$

where for almost every t we have $\big|\big|\big(a_{i,1}(t),\dots,a_{i,q}(t)\big)\big|\big| < \delta_i$. Put $\lambda = \delta_1(\delta_1+\delta_2)^{-1}$ so that $0 < \lambda < 1$, and then define

$$\psi(t) = \begin{cases} \varphi_1\left(\lambda^{-1}t\right) & \text{if} \quad 0 \leq t \leq \lambda, \\ \\ \varphi_2\left((1-\lambda)^{-1}(t-\lambda)\right) & \text{if} \quad \lambda \leq t \leq 1. \end{cases}$$

Clearly $\psi : [0,1] \to \Omega$ and $\psi(0) = x$, $\psi(1) = z$. Moreover, since $\varphi_1(1) = y = \varphi_2(0)$, the function ψ is continuous at the point $t = \lambda$, and so ψ is absolutely continuous on the interval $[0,1]$. If $0 \leq t \leq \lambda$ we have for almost all such t

$$\psi'(t) = \lambda^{-1}\varphi_1'\left(\lambda^{-1}t\right) = \sum_{k=1}^{q} \lambda^{-1} a_{1,k}\left(\lambda^{-1}t\right) Y_{\psi(t)}^k.$$

But

$$\begin{aligned}
&\big|\big|\big(\lambda^{-1}a_{1,1}(\lambda^{-1}t), \ldots, \lambda^{-1}a_{1,q}(\lambda^{-1}t)\big)\big|\big| \\
&\qquad\qquad \leq \lambda^{-1}\big|\big|\left(a_{1,1}\left(\lambda^{-1}t\right), \ldots, a_{1,q}\left(\lambda^{-1}t\right)\right)\big|\big| \\
&\qquad\qquad \leq \lambda^{-1}\,\delta_1 = (\delta_1 + \delta_2)
\end{aligned}$$

by property (iv) of homogeneous norms. Similarly, if $\lambda \leq t \leq 1$ we have

$$\begin{aligned}
\psi'(t) &= (1-\lambda)^{-1}\varphi_2\left((1-\lambda)^{-1}(t-\lambda)\right) \\
&= \sum_{k=1}^{q}(1-\lambda)^{-1}a_{2,k}\left((1-\lambda)^{-1}(t-\lambda)\right) Y_{\psi(t)}^k,
\end{aligned}$$

and

$$\begin{aligned}
&\big|\big|\big((1-\lambda)^{-1}a_{2,1}\big((1-\lambda)^{-1}(t-\lambda)\big), \ldots, (1-\lambda)^{-1}a_{2,1}\big((1-\lambda)^{-1}(t-\lambda)\big)\big)\big|\big| \\
&\qquad \leq (1-\lambda)^{-1}\big|\big|\big(a_{2,1}\big((1-\lambda)^{-1}(t-\lambda)\big), \ldots a_{2,1}\big((1-\lambda)^{-1}(t-\lambda)\big)\big)\big|\big| \\
&\qquad \leq (1-\lambda)^{-1}\delta_2 = (\delta_1 + \delta_2).
\end{aligned}$$

It follows that $\psi \in AC(x, z, \delta_1 + \delta_2)$, and so $\rho(x,z) \leq \delta_1 + \delta_2$. But since δ_1 was any number larger than $\rho(x,y)$ and δ_2 was any number larger than $\rho(y,z)$, it follows that $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$ so ρ is indeed a metric. This completes the proof of assertion (2).

Finally we show that ρ is jointly continuous. Let $x_1, x_2, y_1, y_2 \in \Omega$. Then

$$\begin{aligned}
\rho(x_2, y_2) &\leq \rho(x_2, x_1) + \rho(x_1, y_1) + \rho(y_1, y_2) \\
\rho(x_1, y_1) &\leq \rho(x_1, x_2) + \rho(x_2, y_2) + \rho(y_2, y_1),
\end{aligned}$$

and hence

$$\begin{aligned}
&\big|\rho(x_1, y_1) - \rho(x_2, y_2)\big| \leq \rho(x_1, x_2) + \rho(y_1, y_2) \\
&\qquad\qquad \leq C\,\max\left\{|x_1 - x_2|^{1/d},\, |x_1 - x_2|^{1/D},\, |y_1 - y_2|^{1/d},\, |y_1 - y_2|^{1/D}\right\}.
\end{aligned}$$

This last inequality establishes joint continuity, and completes the proof. □

The metric ρ we have constructed depends both on the control system $\mathcal{Y}$ and on the choice of homogeneous norm $\big|\big| \cdot \big|\big|$. We remark that for many applications, what is important is not one particular choice of metric, but rather a corresponding equivalence class of metrics, and it turns out that the equivalence class does not depend on the choice of homogeneous norm. We make this precise as follows.

Definition 1.8. *Let $\Omega \subset \mathbb{R}^n$ be an open set.*

(i) Two metrics ρ_1 and ρ_2 are globally equivalent on Ω if there is a constant $C \geq 1$ so that for all $x, y \in \Omega$,

$$C^{-1}\,\rho_2(x,y) \leq \rho_1(x,y) \leq C\,\rho_2(x,y).$$

(ii) *Two metrics ρ_1 and ρ_2 are locally equivalent on Ω if for every compact subset $K \subset \Omega$ there is a constant $C(K) \geq 1$ so that for all $x, y \in K$,*

$$C(K)^{-1}\,\rho_2(x,y) \leq \rho_1(x,y) \leq C(K)\,\rho_2(x,y).$$

The notion of local equivalence is appropriate when one is concerned only with local behavior, and hence small distances.

We now observe that although different choices of homogeneous norm can lead to different metrics, these metrics are globally equivalent.

PROPOSITION 1.9. *Let $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ be a control structure on a connected open set $\Omega \subset \mathbb{R}^n$. Let $\big|\big| \cdot \big|\big|_a$ and $\big|\big| \cdot \big|\big|_b$ be two homogeneous norms. Then the corresponding Carnot-Carathéodory metrics ρ_a and ρ_b are globally equivalent.*

PROOF. There is a constant $C \geq 1$ so that $C^{-1}\big|\big| y \big|\big|_b \leq \big|\big| y \big|\big|_a \leq C \big|\big| y \big|\big|_b$ for all $y \in \mathbb{R}^q$. Suppose that $\psi : [0,1] \to \Omega$ and $\psi'(t) = \sum_{k=1}^q a_k(t)\, Y^k_{\psi(t)}$. Then $||(a_1(t), \ldots, a_q(t))||_b < \delta$ implies $||(a_1(t), \ldots, a_q(t))||_a < C\,\delta$. It follows that the space $AC_b(x, y; \delta)$ of allowable paths for the homogeneous norm $|| \cdot ||_b$ is contained in the space $AC_a(x, y; C\delta)$. This together with the corresponding inclusion $AC_a(x, y; \delta) \subset AC_b(x, y; C\delta)$ show that the two metrics are globally equivalent. □

There are other modifications of a control system which do not change the equivalence class of the corresponding Carnot-Carathéodory metric. The following is an illustration.

PROPOSITION 1.10. *Let $\Omega \subset \mathbb{R}^n$ be a connected open set, and suppose $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ and $\mathcal{Z} = \{Z_1, \ldots, Z_r; e_1, \ldots e_r\}$ are two control systems on Ω. Suppose $\big|\big| \mathcal{Y} \big|\big|_{\Omega,1} < +\infty$, $\big|\big| \mathcal{Z} \big|\big|_{\Omega,1} < \infty$, $\nu_{\mathcal{Y}}(\Omega) > 0$, and $\nu_{\mathcal{Z}}(\Omega) > 0$. Suppose that for each $1 \leq j \leq q$ and $1 \leq k \leq r$ we can write*

$$Y_j = \sum_{l=1}^{r} \alpha_{j,l}\, Z_l \qquad \text{and} \qquad Z_k = \sum_{m=1}^{q} \beta_{k,m} Y_m,$$

where the coefficients $\{\alpha_{j,l}\}$ and $\{\beta_{k,m}\}$ are bounded on Ω.

(1) *If $\alpha^j_l = 0$ unless $e_l \leq d_j$ and $\beta^k_m = 0$ unless $d_m \leq e_k$, the metrics $\rho_{\mathcal{Y}}$ and $\rho_{\mathcal{Z}}$ are locally equivalent.*
(2) *If $\alpha^j_l = 0$ unless $e_l = d_j$ and $\beta^k_m = 0$ unless $d_m = e_k$, the metrics $\rho_{\mathcal{Y}}$ and $\rho_{\mathcal{Z}}$ are globally equivalent.*

PROOF. Since different homogeneous norms result in globally equivalent metrics, we can take the homogeneous norms both metrics to be $|| \cdot ||_\infty$ as defined in equation (1.2).

Let $K \subset \Omega$ be compact, and let $\delta_K = \sup_{x,y \in K} \rho_{\mathcal{Y}}(x, y)$. Let $x, y \in K$ and let $\delta \leq 2\delta_K$. Suppose that $\phi \in AC_{\mathcal{Y}}(x, y; \delta)$ so that for almost all $t \in [0, 1]$ we can write $\phi'(t) = \sum_{j=1}^q a_j(t)(Y_j)_{\phi(t)}$ with $|a_j(t)| < \delta^{d_j}$. It follows that

$$\phi'(t) = \sum_{l=1}^{r} \Big[\sum_{j=1}^{q} a_j(t)\,\alpha_{j,l}(\phi(t)) \Big] (Z_l)_{\phi(t)}.$$

Under the hypotheses in (1), since $\delta \leq 2\delta_K$ we have

$$\Big| \sum_{j=1}^{q} a_j(t)\,\alpha_{j,l}(\phi(t)) \Big| \leq \sum_{\substack{j=1 \\ e_l \leq d_j}}^{q} \delta^{d_j} \sup_{x \in \overline{\Omega}} |\alpha_{j,l}(x)| \leq (C(K)\,\delta)^{e_l} \tag{1.7}$$

where $C(K)$ depends on δ_K and the supremum of the coefficients $\{\alpha_{j,k}\}$ on Ω. It follows that $\phi \in AC_{\mathcal{Z}}(x, y; C(K)\delta)$ and hence $\rho_{\mathcal{Y}}(x, y) \leq C(K)\, \rho_{\mathcal{Z}}(x, y)$. The opposite inequality is proved the same way, so the metrics $\rho_{\mathcal{Y}}$ and $\rho_{\mathcal{Z}}$ are locally equivalent.

If the hypotheses in (2) hold, we can repeat the argument for any pair of points $x, y \in \Omega$. If $\phi \in AC_{\mathcal{Y}}(x, y; \delta)$, write $\phi'(t) = \sum_{j=1}^{q} a_j(t)(Y_j)_{\phi(t)}$ with $|a_j(t)| < \delta^{d_j}$. Then inequality (1.7) is replaced by

$$\Big|\sum_{j=1}^{q} a_j(t)\, \alpha_{j,l}(\phi(t))\Big| \leq \sum_{\substack{j=1 \\ e_l = d_j}}^{q} \delta^{d_j} \sup_{x \in \overline{\Omega}} |\alpha_{j,l}(x)| \leq (C\,\delta)^{e_l} \tag{1.7'}$$

where C depends only on the supremum of the coefficients $\{\alpha_{j,k}\}$ on Ω. Thus the two metrics are globally equivalent. This completes the proof. □

It follows that adding additional vector fields with maximal formal degree does not change the local behavior of the metric associated to a control system.

COROLLARY 1.11. *Let $\Omega \subset \mathbb{R}^n$ be a connected open set with compact closure, and suppose $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ is a control system on Ω. Let $d = \sup\{d_1, \ldots, d_q\}$. Let $\{Y_{q+1}, \ldots, Y_r\}$ be vector fields on Ω and suppose that we can write*

$$Y_s = \sum_{j=1}^{q} \gamma_{s,j}\, Y_j, \qquad q+1 \leq s \leq r,$$

where the functions $\gamma_{s,j}$ are bounded on Ω. If we set $d_{q+1} = \cdots = d_r = d$, then $\mathcal{Y}_1 = \{Y_1, \ldots, Y^r; d_1, \ldots, d_r\}$ is a control system on Ω which is locally equivalent to $\mathcal{Y}$.

1.4. Vector fields of finite type.

In this section we describe a standard method for constructing a control system from a collection of vector fields $S = \{X_1, \ldots, X_p\}$ satisfying a certain spanning hypothesis. Recall that if $\Omega \subset \mathbb{R}^n$ is an open set, then $T(\Omega)$ denotes the space of real-valued infinitely differentiable vector fields on Ω. If $x \in \Omega$ and if $Y \in T(\Omega)$, the value of Y at x is a tangent vector $Y_x \in T_x$ and . Also recall from Chapter 2, Definition 1.5, that if $X_1, \ldots, X_p \in \mathcal{D}(\Omega)$, then $\mathcal{L}(X_1, \ldots, X_p)$ is the Lie subalgebra of $T(\Omega)$ generated by $X_1, \ldots, X_p$. As a real vector space, $\mathcal{L}(X_1, \ldots, X_p)$ is spanned by the vector fields $\{X_1, \ldots, X_p\}$ and all iterated commutators $\{[X_1, X_2], \ldots, [X_{p-1}, X_p], \ldots, [X_j, [X_k, X_l]], \ldots\}$.

DEFINITION 1.12. *Let $\Omega \subset \mathbb{R}^n$ be a connected open set. Let $X_1, \ldots, X_p \in T(\Omega)$. Then $\{X_1, \ldots, X_p\}$ are of finite type[2] on Ω if for every $x \in \Omega$,*

$$T_x = \Big\{ Y_x \,\Big|\, Y \in \mathfrak{L}(X_1, \ldots, X_p) \Big\}.$$

If for every $x \in \Omega$, T_x is spanned by the values at x of iterated commutators of $\{X_1, \ldots, X_p\}$ of length at most m, then we say that $\{X_1, \ldots, X_p\}$ are of finite type m on Ω.

[2]This condition is sometimes referred to as *Hörmander's condition.* Hörmander made crucial use of this condition in his study of partial differential operators of the form $\mathcal{L} = X_1^2 + \cdots + X_p^2 + X_0$ in [**Hö67**]. However, this condition was previously used by other mathematicians such as Carathéodory and Chow.

Suppose that $\Omega \subset \mathbb{R}^n$ is a connected open set, and vector fields $\{X_1, \ldots, X_p\}$ are of finite type m on Ω. We shall write down a list of all possible iterated commutators of $\{X_1, \ldots, X_p\}$ of length at most m. Thus for $1 \leq r \leq m$ let $\mathbb{I}_r$ denote the set of all ordered r-tuples $J = (j_1, \ldots, j_r)$ of positive integers with $1 \leq j_k \leq p$. For each $J \in \mathbb{I}_r$ let

$$Y_{[J]} = \left[X_{j_r},\, \left[X_{j_{r-1}},\, \cdots,\, [X_{j_2},\, X_{j_1}]\cdots\right]\right].$$

It is important to note that in general the vector fields $\{Y_{[J]}\}_{J\in\mathbb{I}_r}$ are not all distinct. For example, if $r \geq 2$ and if $j_1 = j_2$, then $Y_{[J]} = 0$ since already $[X_{j_2}, X_{j_1}] = 0$.. However, the collection $\left\{Y_{[J]},\, J \in \mathbb{I}_r\right\}$ does give us a collection of p^r different symbols, each of which represents a vector field.

Now let $\{Y_1, \ldots, Y_q\}$ be a list of all these symbols $\{Y_{[J]}\}$ where $J \in \mathbb{I}_r$ and $1 \leq r \leq m$. We have $q = p + p^2 + \cdots + p^m$, and this list contains all iterated commutators of $\{X_1, \ldots, X_p\}$ of length at most m. If $Y_j = Y_{[J]}$ with $J \in \mathbb{I}_r$, we set the formal degree of the symbol Y_j to be $d_j = r$. Thus for example, X_1 and X_2 have formal degree 1, the vector fields $[X_1, X_1]$ and $[X_1, X_2]$ have formal degree 2, and the vector field $[X_1, [X_1, X_2]]$ has formal degree 3.

Some of the vector fields Y_j may be zero (as in the example $[X_1, X_1]$ above), and it can happen that the same vector field is represented more than once in the list $\{Y_1, \ldots, Y_q\}$. Thus it is important to note that the formal degree d_j is assigned to a particular representation of the vector field in this list, and not necessarily to the underlying vector field itself.

PROPOSITION 1.13. *The data $\mathcal{Y}_0$ of vector fields $\{Y_1, \ldots, Y_q\}$ and formal degrees $\{d_1, \ldots, d_q\}$ described above are a control system on Ω.*

PROOF. We need to verify condiitions (H-1) and (H-2) of Definition 1.3. But the vectors $\{(Y_1)_x, \ldots, (Y_q)_x\}$ span the tangent space T_x for every $x \in \Omega$ since the vectors $\{X_1, \ldots, X_p\}$ are assumed to be of finite type m on Ω. Thus (H-1) is satisfied.

Next one observes that the Jacobi identity in a Lie algebra implies that the commutator of two iterated commutators, one of length k and one of length l, can be written as a linear combination of iterated commutators of length $k + l$. (See Chapter 7, Proposition 2.1 and Corollary 2.2). Thus since we have defined the formal degree d_k of a vector field Y_k to be the length of the commutator in the given representation for Y^k, it follows that (H-2) is also satisfied. □

There are variants of this construction which are also used in applications. Rather than formulate the most general result, we provide one illustration which is important in the study of heat equations $\triangle_x + \partial_t$ where formally the derivative with respect to time acts like a second order operator, or more generally, operators $\mathcal{L} = X_1^2 + \cdots + X_p^2 + X_0$ studied by Hörmander in which X_0 should be counted as a second order operator.[3]. Thus suppose $X_0, X_1, \ldots X_p \in T(\Omega)$ are of finite type m. We again let $\{Y_1, \ldots, Y_q\}$ be a list of all iterated commutators of length at most m. However, this time we assign formal degree 1 to each of the vector fields $\{X_1, \ldots, X_p\}$, and assign formal degree 2 to the vector field X_0. For any iterated commutator $Y_j = Y_J$ we assign formal degree d_j which is the sum of the formal degrees of the vectors $\{X_k\}$ that are present in the commutator. For example,

[3]See the discussion in Chapter 5.

$[X_1, X_2]$ has formal degree 2, $[X_0, X_2]$ has formal degree 3, and $[X_0, [X_1, X_2]]$ has formal degree 4. Let $\mathcal{Y}_1$ denote the collection of vector fields $\{Y_1, \ldots, Y_q\}$ together with the (new) formal degrees $\{d_1, \ldots, d_q\}$. Exactly as in Proposition 1.13 we have:

COROLLARY 1.14. *The data $\mathcal{Y}_1$ of vector fields $\{Y_1, \ldots, Y_q\}$ and formal degrees $\{d_1, \ldots, d_q\}$ are a control system on Ω.*

2. Examples of metrics and operators

In this section we look at several examples of control systems that arise in the study of partial differential equations.

2.1. Isotropic Euclidean space.

On the space $\mathbb{R}^n$ consider the vector fields

$$\{Y_1, \ldots, Y_n\} = \Big\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\Big\}.$$

Since $[Y_j, Y_k] = 0$ for all $1 \leq j \leq k$, it follows that if we take $d_1 = \cdots = d_n = 1$, then $\mathcal{Y}_1 = \{Y_1, \ldots, Y^n; d_1, \ldots, d_n\}$ is a control system. Let $|| \cdot ||$ be any norm on $\mathbb{R}^n$. It follows from Proposition 1.7 that the corresponding Carnot-Carathéodory metric $\rho_{\mathcal{Y}}$ is comparable to the Euclidean metric. As a warm-up exercise, we show that in fact $\rho_{\mathcal{Y}}(x, y) = ||x - y||$.

PROOF. Suppose that $AC(x, y; \delta)$ is not empty. Then there exists an absolutely continuous mapping $\psi : [0, 1] \to \mathbb{R}^n$ with $\psi(0) = x$, $\psi(1) = y$, $\psi'(t) = \sum_{j=1}^n a_j(t)\, Y^j_{\psi(t)}$ and $\big|\big|\big(a_1(t), \ldots, a_n(t)\big)\big|\big| < \delta$ for almost all t. Hence

$$||y - x|| = \Big|\Big| \int_0^1 \psi_j'(t)\, dt \Big|\Big| \leq \int_0^1 \big|\big|\big(a_1(t), \ldots, a_n(t)\big)\big|\big|\, dt \leq \delta.$$

Thus $||y - x|| \leq \rho_{\mathcal{Y}}(x, y)$. On the other hand, let $x, y \in \mathbb{R}^n$ and suppose that $||y - x|| = \delta > 0$. Put $\psi_j(t) = x_j + t(y_j - x_j)$, and $\psi(t) = \big(\psi_1(t), \ldots, \psi_n(t)\big)$. Then $||\psi'(t)|| = \delta$. If $\epsilon > 0$, it follows that $\psi \in AC_{\mathcal{Y}}(x, y; \delta + \epsilon)$, and hence $\rho_{\mathcal{Y}}(x, y) \leq \delta + \epsilon$. Since ϵ is arbitrary, it follows that $\rho_Y(x, y) \leq ||x - y||$, and this completes the proof. □

The Euclidean metric (or any equivalent metric) plays a fundamental role in the analysis of elliptic operators on $\mathbb{R}^n$. Here we wish to provide a very simple example. Consider the Laplace operator on $\mathbb{R}^n$, which we can write

$$\triangle = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2} = (Y_1)^2 + \cdots + (Y^n)^2.$$

It is well known that when $n \geq 3$, the Newtonian potential

$$N(x, y) = \frac{\Gamma\left(\frac{n}{2}\right)}{2(2 - n)\pi^{\frac{n}{2}}}\, |x - y|^{2-n}$$

is a fundamental solution for $\triangle$. This means that, in the sense of distributions, $\triangle_x[N(x, y)] = \delta_x(y)$, where δ_x is the delta function at x. From a more classical point of view, if $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and if

$$u(x) = \int_{\mathbb{R}^n} \varphi(y)\, N(x, y)\, dy$$

then $\triangle u(x) = \varphi(x)$.

We can relate this fundamental solution for $\triangle$ to the Carnot-Carathéodory metric $\rho_{\mathcal{Y}}$ that we have constructed in the following way. Let $\{B_{\mathcal{Y}}(x;\delta)\}$ be the corresponding family of balls. Then it is easy to check that for all multi-indices α, β, there is a constant $C(\alpha,\beta)$ so that

$$\left|\partial_x^\alpha\,\partial_y^\beta\,N(x,y)\right| \le C(\alpha,\beta)\,\frac{\rho_{\mathcal{Y}}(x,y)^{2-|\alpha|-|\beta|}}{\left|B\big(x;\rho_{\mathcal{Y}}(x,y)\big)\right|}.$$

As we shall see, this is a very special case of a general theorem dealing with fundamental solutions for operators of the form $X^0 + (X_1)^2 + \cdots + (X_p)^2$ where the vector fields $\{X^0, X_1, \dots, X_p\}$ are of finite type. Give reference

2.2. Non-isotropic Euclidean space.

Next consider the space $\mathbb{R}^n$ with the same set of vector fields

$$\big\{Y_1, \dots, Y^n\big\} = \Big\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\Big\}.$$

This time, however, assign formal degree d_j to the vector field Y^j where $1 \le d_1 \le d_2 \le \cdots \le d_n$. Then $\mathcal{Y}_2 = \{Y_1, \dots, Y^n; d_1, \dots, d_n\}$ is also a control structure. As a homogeneous norm, we choose

$$||(y_1, \dots, y_n)|| = \sup_{1\le j\le n} |y_j|^{1/d_j}.$$

PROPOSITION 2.1. *The Carnot-Carathéodory metric is given by*

$$\rho(x,y) = \sup_{1\le j\le n} |x_j - y_j|^{1/d_j},$$

and the corresponding family of balls is given by

$$B(x,\delta) = \left\{y \in \mathbb{R}^n \,\middle|\, |y_j - x_j| < \delta^{d_j}\right\}.$$

The proof is a minor modification of the argument given in the isotropic case. If $AC(x,y;\delta)$ is not empty there exists an absolutely continuous mapping $\psi : [0,1] \to \mathbb{R}^n$ with $\psi(0) = x$, $\psi(1) = y$, $\psi'(t) = \sum_{j=1}^n a_j(t)\,Y^j_{\psi(t)}$ and $|a_j(t)|^{1/d_j} < \delta$ for almost all t. Hence $|y_j - x_j| = \left|\int_0^1 a_j(t)\,dt\right| \le \delta^{d_j}$. Thus $\sup_j |y_j - x_j|^{1/d_j} \le \rho_{\mathcal{Y}}(x,y)$. On the other hand, let $x,y \in \mathbb{R}^n$ and suppose that $\sup_j |y_j - x_j|^{1/d_j} = \delta > 0$. Again put $\psi_j(t) = x_j + t(y_j - x_j)$ and $\psi(t) = \big(\psi_1(t), \dots, \psi_n(t)\big)$. If $\epsilon > 0$, it follows that $\psi \in AC_{\mathcal{Y}}(x,y;\delta+\epsilon)$, and hence $\rho_Y(x,y) \le \sup_j |y_j - x_j|^{1/d_j}$.

As an example of the role of such metrics in analysis, consider the heat operator on the space $\mathbb{R}^{n+1}$ with coordinates $(t, x_1, \dots, x_n)$. This time we choose vector fields $Y^0 = \frac{\partial}{\partial t}$ and $Y^j = \frac{\partial}{\partial x_j}$ for $1 \le j \le n$. We assign formal degrees $d_0 = 2$ and $d_j = 1$ for $1 \le j \le n$. The heat operator is then

$$\mathcal{H} = \frac{\partial}{\partial t} - \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2} = Y^0 - (Y_1)^2 - \cdots (Y^n)^2.$$

Again there is a well-know fundamental solution for $\mathcal{H}$ given by Give reference

$$H\big((t,x),(s,y)\big) = \begin{cases} (4\pi(t-s))^{-\frac{n}{2}}\exp\left[\frac{|x-y|^2}{4(t-s)}\right] & \text{if } t-s > 0,\\ 0 & \text{if } t-s \le 0.\end{cases}$$

The size of this heat kernel is somewhat difficult to describe using the standard isotropic Euclidean metric on $\mathbb{R}^{n+1}$. However if we use a non-isotropic Carnot-Carathéodory metric , then

$$\begin{aligned}\big|\partial_x^{\alpha_1}\partial_y^{\alpha_2}\partial_t^{\beta_1}\partial_s^{\beta_2}H\big((t,x),(s,y)\big)\\ \leq C(\alpha_1,\alpha_2,\beta_1,\beta_2)\frac{\rho\big((x,t),(y,s)\big)^{2-|\alpha_1|-|\alpha_2|-2|\beta_1|-2|\beta_2|}}{V\big((x,t),(y,s)\big)}\end{aligned}$$

Here $V\big((x,t),(y,s)\big)$ is the volume of the non-isotropic ball centered at (x,t) and radius $\rho\big((x,t),(y,s)\big)$. Note that differentiation with respect to t or s has the same effect as two derivatives with respect to x or y. This is one of the reasons for assigning the formal degree 2 to the vector field Y^0.

2.3. The Heisenberg group.

In the two examples we have considered so far, the Carnot-Carathéodory metrics have been invariant under Euclidean translations, reflecting the fact that the basic vector fields have constant coefficients. In the next example the vector fields have variable coefficients. Although the resulting metric is no longer translation invariant, it does have an invariance under a different group of transformations, reflecting the fact that the basic vector fields still form a finite dimensional nilpotent Lie group.

The underlying space in this example is $\mathbb{R}^{2n+1} = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$, and it is traditional to label the coordinates $(x,y,t) = (x_1,\ldots,x_n,y_1,\ldots,y_n,t)$. Consider the $(2n+1)$ vector fields $\{X_1,\ldots,X^n,Y_1,\ldots,Y^n,T\}$ where

$$\begin{aligned}X^j &= \frac{\partial}{\partial x_j} + 2y_j\,\frac{\partial}{\partial t}, \quad 1 \leq j \leq n;\\ Y^k &= \frac{\partial}{\partial y_k} - 2x_k\,\frac{\partial}{\partial t}, \quad 1 \leq k \leq n;\\ T &= \frac{\partial}{\partial t}.\end{aligned} \tag{2.1}$$

We assign formal degree 1 to the vector fields $\{X_1,\ldots,X^n\}$ and $\{Y_1,\ldots,Y^n\}$, but we assign degree 2 to the vector field T.

It is clear that the vector fields $\{X_1,\ldots,X^n,Y_1,\ldots,Y^n,T\}$ span the tangent space at each point of $\mathbb{R}^{2n+1}$ so that hypothesis (H-1) of Definition 1.3 is satisfied. Moreover, for $1 \leq j,k \leq n$ we have the commutation relationships

$$\begin{aligned}\big[X^j,\,T\big] = \big[Y^k,\,T\big] &= 0,\\ \big[Y^k,\,X^j\big] &= \begin{cases}4T & \text{if } j=k,\\ 0 & \text{if } j\neq k.\end{cases}\end{aligned} \tag{2.2}$$

It follows that hypothesis (H-2) is satisfied as well. Thus the collection of vector fields and formal degrees $\mathcal{Y}_3 = \{X_1,\ldots,X^n,Y_1,\ldots,Y^n,T;1,\ldots,1,1,\ldots 1,2\}$ is a control structure.[4] It is also clear that the $(2n+1)$-dimensional real vector subspace of $\mathcal{D}(\mathbb{R}^{2n+1})$ spanned by $\{X_1, \ldots, X^n, Y_1, \ldots, Y^n, T\}$ is closed under the Lie bracket, and hence forms a Lie algebra, which we denote hy $\mathfrak{h}_n$

Before investigating the Carnot-Carathéodory metrics induced by $\mathcal{Y}_3$, we define the Lie group corresponding to the Lie algebra $\mathfrak{h}_n$. Define[5] a product on $\mathbb{R}^{2n+1}$ by setting

$$(x, y, t) \cdot (u, v, s) = \big(x + u, y + v, t + s + 2\sum_{j=1}^{n}(y_j u_j - x_j v_j)\big). \tag{2.3}$$

It is a straightforward calculation to check that this product defines a group structure on $\mathbb{R}^{2n+1}$. This is called the *Heisenberg group* and is denoted by $\mathbb{H}_n$. The identity element is $(0, 0, 0)$, the product is associative, and if $(x, y, t) \in \mathbb{H}_n$, then $(x, y, t)^{-1} = (-x, -y, -t)$. Moreover $(x, y, t) \cdot (u, v, s) = (u, v, s) \cdot (x, y, t)$ f and only if $\sum_{j=1}^{n}(y_j u_j - x_j v_j) = 0$. Thus $\mathbb{H}_n$ is not an Abelian group.

Define the left translation operator $T_{x,y,t} : \mathbb{H}_n \to \mathbb{H}_n$ by setting

$$T_{(x,y,t)}(u, v, s) = (x, y, t)^{-1} \cdot (u, v, s), \tag{2.4}$$

and define the corresponding operator $L_{(x,y,t)}$ on functions by

$$L_{(x,y,t)} f(u, v, s) = f\big((x, y, t)^{-1} \cdot (u, v, s)\big). \tag{2.5}$$

Thus $L_{(x,y,t)}[f] = f \circ T_{(x,y,t)}$. As is customary with non-Abelian groups, the inverse is used on the right hand sides so that

$$\begin{aligned} T_{(x_1,y_1,t_1)} \circ T_{(x_2,y_2,t_2)} &= T_{(x_1,y_1,t_1)\cdot(x_2,y_2,t_2)}; \\ L_{(x_1,y_1,t_1)} \circ L_{(x_2,y_2,t_2)} &= L_{(x_1,y_1,t_1)\cdot(x_2,y_2,t_2)}. \end{aligned}$$

As we have already indicated, there is an intimate connection between the group structure $\mathbb{H}_n$ and the vector fields $\{X_1, \ldots, X^n, Y_1, \ldots, Y^n, T\}$. The key to this connection is invariance under translation.

Definition 2.2. *A vector field $Z \in \mathcal{D}(\mathbb{R}^{2n+1})$ is left-invariant on $\mathbb{H}_n$ if it commutes with every left translation operator. This means that for every differentiable function f and all $(x, y, t) \in \mathbb{H}_n$,*

$$Z\big[L_{(x,y,t)}[f]\big] = L_{(x,y,t)}\big[Z[f]\big]. \tag{2.6}$$

The following result then explains the meaning of equation (2.6) and provides the connection between the group $\mathbb{H}_n$ and the Lie algebra $\mathfrak{h}_n$.

Proposition 2.3. *A vector field Z on $\mathbb{R}^{2n+1}$ is left invariant if and only if it is a real linear combination of the (2n+1) vector fields $\{X^j, Y^k, T\}$. In particular, the vector space of left-invariant vector fields on $\mathbb{H}_n$ is the same as the Lie algebra $\mathfrak{h}_n$.*

---

[4]Note that equation (**??**) shows that we would still have a control structure if we assigned T the formal degree 1. In this case, Lemma 1.7 implies that the resulting metric is equivalent to the Euclidean metric on any compact set. However, if T is assigned the formal degree 3 then hypothesis (H-2) no longer holds.

[5]Using the Campbell-Baker-Hausdorff formula, one can derive the product structure of the Lie group from the Lie algebra structure. Here however we first define the group multiplication, and then later show the relationship with $\mathfrak{h}_n$¿

PROOF. Let $Z = \sum_{j=1}^{n} \Big[a_j \frac{\partial}{\partial x_j} + b_j \frac{\partial}{\partial y_j}\Big] + c\frac{\partial}{\partial t}$ where $\{a_j,\, b_j,\, c\} \in \mathcal{E}(\mathbb{R}^{2n+1})$. Then $L_{(x,y,t)}f(u,v,s) = f\big(x+u, y+v, t+s+2[\langle y,u\rangle - \langle x,v\rangle]\big)$ and hence

$$
\begin{aligned}
&Z[L_{(x,y,t)}f](u,v,s) \\
&\quad= \sum_{j=1}^{n} a_j(u,v,s)\frac{\partial f}{\partial u_j}\big((x,y,t)\cdot(u,v,s)\big) + \sum_{j=1}^{n} b_j(u,v,s)\frac{\partial f}{\partial v_j}\big((x,y,t)\cdot(u,v,s)\big) \\
&\qquad + \Big[c(u,v,s) - 2\sum_{j=1}^{n}\big[x_j\, b_j(u,v,s) - y_j\, a_j(u,v,s)\big]\Big]\frac{\partial f}{\partial s}\big((x,y,t)\cdot(u,v,s)\big).
\end{aligned}
$$

On the other hand

$$
\begin{aligned}
L_{(x,y,t)}[Zf](u,v,s) = \sum_{j=1}^{n} a_j\big((x,y,t)\cdot(u,v,s)\big)\frac{\partial f}{\partial u_j}\big((x,y,t)\cdot(u,v,s)\big) \\
+ \sum_{j=1}^{n} b_j\big((x,y,t)\cdot(u,v,s)\big)\frac{\partial f}{\partial v_j}\big((x,y,t)\cdot(u,v,s)\big) \\
+ c\big((x,y,t)\cdot(u,v,s)\big)\frac{\partial f}{\partial s}\big((x,y,t)\cdot(u,v,s)\big)
\end{aligned}
$$

Thus if equation (2.6) is to hold for all functions f, we must have

(i) $a_j(u,v,s) = a_j\big((x,y,t)\cdot(u,v,s)\big)$;
(ii) $b_j(u,v,s) = b_j\big((x,y,t)\cdot(u,v,s)\big)$;
(iii) $c(u,v,s) - \sum_{j=1}^{n} 2x_j b_j(u,v,s) + \sum_{j=1}^{n} 2y_j a_j(u,v,s) = c\big((x,y,t)\cdot(u,v,s)\big)$.

Since these compatibility conditions must hold for all (x,y,t) and in particular for $(u,v,s)^{-1}$, the equations (i) and (ii) show that $a_j(u,v,s) = a_j(0,0,0) = A_j$ and $b_j(u,v,s) = b_j(0,0,0) = B_j$. Then setting $(u,v,s) = (0,0,0)$ and $c(0,0,0) = C$ in equation (iii), we obtain

$$
c(x,y,t) = C - \sum_{j=1}^{n} 2B_j\, x_j + \sum_{j=1}^{n} 2A_j\, y_j.
$$

Substituting back in the original definition for Z, we see that a vector field Z is left invariant if and only if

$$
\begin{aligned}
Z &= \sum_{j=1}^{n} A_j\Big[\frac{\partial}{\partial x_j} + 2y_j\frac{\partial}{\partial t}\Big] + B_j\Big[\frac{\partial}{\partial y_j} - 2x_j\frac{\partial}{\partial t}\Big] + C\frac{\partial}{\partial t} \\
&= \sum_{j=1}^{n} A_j\, X^j + B_j\, Y^j + C\, T.
\end{aligned}
$$

This completes the proof. □

We now describe the Carnot-Carathéodory metrics ρ associated to the control system $\mathcal{Y}_3$. The associated family of dilations on $\mathbb{R}^{2n+1}$ is given by $D_\delta(x,y,t) = (\delta x, \delta y, \delta^2 t)$, and the metric depends on the choice of a D-homogeneous norm on $\mathbb{R}^{2n+1}$. However, for any choice of norm, we have the following invariance.

PROPOSITION 2.4. *Let ρ be a Carnot-Carathéodory metric on $\mathbb{H}_n$ defined using $\mathcal{Y}_3$ and any D-homogeneous norm. Let p_0, p_1, and p_2 be any three points of $\mathbb{H}_n$. Then*

$$\rho(p_1, p_2) = \rho\big(T_{p_0}(p_1), T_{p_0}(p_2)\big).$$

PROOF. This is a simple consequence of the identity

$$AC(p_1, p_2; \delta) = AC(T_{p_0}(p_1), T_{p_0}(p_2)\big); \delta),$$

which in turn follows directly from the left invariance of the basic vector fields $\{X_1, \ldots, X^n, Y_1, \ldots, Y^n, T\}$. □

Next, in order to investigate the actual nature of the metric ρ we choose the homogeneous norm

$$||(x_1, \ldots, x_n, y_1, \ldots, y_n, t)|| = \max\big\{|x_1|, \ldots |x_n|, |y_1|, \ldots, |y_n|, |t|^{\frac{1}{2}}\big\}.$$

Suppose that $p_0 = (x, y, t)$ and $p_1 = (u, v, s)$ are two points in $\mathbb{H}_n$. Let $\psi : [0, 1] \to \mathbb{H}_n$ be an absolutely continuous mapping with $\psi(0) = p_0$ and $\psi(1) = p_1$. We begin by unraveling what it means that $\psi \in AC(p_0, p_1; \delta)$. Thus write $\psi(r) = \big(\phi_1(r), \ldots, \phi_n(r), \eta_1(r), \ldots, \eta_n(r), \tau(r)\big)$. The key point, according to Definition 1.5, is that we must control the sizes of the coefficients of $\psi'(r)$ written as a linear combination of the vectors $\{(X_1)_{\psi(r)}, \ldots, X^n_{\psi(r)}, (Y_1)_{\psi(r)}, \ldots, Y^n_{\psi(r)}, T_{\psi(r)}\}$, rather than the sizes of the derivatives $(\phi_1'(r), \ldots, \phi_n'(r), \eta_1'(r), \ldots, \eta_n'(r), \tau'(r))$. If we do the algebra we obtain

$$\psi'(t) = \sum_{j=1}^{n} \phi_j'(t) X^j_{\psi(t)} + \sum_{j=1}^{n} \eta_j'(t) Y^j_{\psi(t)} + \Big[\tau'(t) - 2\sum_{j=1}^{n}\Big(\phi_j'(t)\eta_j(t) - \eta_j'(t)\phi_j(t)\Big)\Big] T_{\psi(t)}.$$

Thus $\psi \in AC(p_0, p_1; \delta)$ if and only if $\psi(j) = p_j$ for $j = 0$, 1, and if for almost every $t \in [0, 1]$,

$$\begin{aligned} |\phi_j'(t)| &< \delta, \qquad 1 \le j \le n, \\ |\eta_j'(t)| &< \delta, \qquad 1 \le j \le n, \\ \Big|\tau'(t) - 2\sum_{j=1}^{n}\Big(\phi_j'(t)\eta_j(t) - \eta_j'(t)\phi_j(t)\Big)\Big| &< \delta^2. \end{aligned}$$

LEMMA 2.5. *Let ρ be the Carnot-Carathéodory metric associated to the control system $\mathcal{Y}_3$ and the homogeneous norm. If $p_0 = (x, y, t)$ and $p_1 = (u, v, s)$ then*

$$\begin{aligned} \max\Big\{|x_j - u_j|, |y_j - v_j|, \frac{1}{\sqrt{4n+1}}\big|t - s + 2\sum_{k=1}^{n}(y_k u_k - x_k v_k)\big|^{\frac{1}{2}}\Big\} &\le \rho(p_0, p_1) \\ \le \max\Big\{|x_j - u_j|, |y_j - v_j|, \big|t - s + 2\sum_{k=1}^{n}(y_k u_k - x_k v_k)\big|^{\frac{1}{2}}\Big\}&. \end{aligned}$$

PROOF. Suppose first that

$$\max\Big\{|x_j - u_j|, |y_j - v_j|, \big|t - s + 2\sum_{k=1}^{n}(y_k u_k - x_k v_k)\big|^{\frac{1}{2}}\Big\} < \delta.$$

Define $\psi = (\phi_1, \ldots, \phi_n, \eta_1, \ldots, \eta_n, \tau) : [0, 1] \to \mathbb{H}_n$ by setting $\phi_j(r) = u_j + r(x_j - u_j)$, $\eta_j(r) = v_j + r(y_j - v_j)$, and $\tau(r) = s + r(t - s)$. Then $\psi(0) = (u, v, s)$, $\psi(1) = (x, y, t)$, and we have

$$
\begin{aligned}
|\phi_j'(r)| &= |x_j - u_j| < \delta, \quad 1 \le j \le n\\
|\eta_j'(r)| &= |y_j - v_j| < \delta, \quad 1 \le j \le n
\end{aligned}
$$

$$
\Big|\tau'(r) - 2\sum_{j=1}^{n}\Big(\phi_j'(r)\eta_j(r) - \eta_j'(r)\phi_j(r)\Big)\Big| = \Big|(t-s) - 2\sum_{j=1}^{n}(x_jv_j - y_ju_j)\Big| < \delta^2.
$$

Thus $\psi \in AC(p_0, p_1; \delta)$, and it follows that

$$
\rho(p_0,p_1) \le max\Big\{\big|x_j-u_j\big|, \big|y_j-v_j\big|, \big|t-s+2\sum_{k=1}^{n}(y_ku_k - x_kv_k)\big|^{\frac12}\Big\}.
$$

Conversely, suppose that $\rho(p_0, p_1) < \delta$. Then there exists a mapping $\psi \in AC(p_0, p_1; \delta)$. Write $\psi = (\phi_1, \ldots, \phi_n, \eta_1, \ldots, \eta_n, \tau)$. For $1 \le j \le n$ we have

$$
|x_j - u_j| = \Big|\int_0^1 \phi_j'(t)\,dt\Big| < \delta \qquad \text{and} \qquad |y_j - v_j| = \Big|\int_0^1 \eta_j'(t)\,dt\Big| < \delta.
$$

In addition, we have

$$
\begin{aligned}
\big|t-s+2\sum_{k=1}^{n}(y_ku_k - x_kv_k)\big| &= \big|t-s-2\sum_{k=1}^{n}(y_k(x_k-u_k) - x_k(y_k-v_k))\big|\\
&= \Big|\int_0^1 \Big[\tau'(t) - 2\sum_{k=1}^{n}\big(y_k\phi_k'(t) - x_k\eta_k'(t)\big)\Big]\,dt\Big|\\
&\le \Big|\int_0^1 \Big[\tau'(t) - 2\sum_{k=1}^{n}\big(\eta_k(t)\phi_k'(t) - \phi_k(t)\eta_k'(t)\big)\Big]\,dt\Big|\\
&\qquad + 2\Big|\int_0^1 \sum_{k=1}^{n}\big((y_k-\eta_k(t))\phi_k'(t) - (x_k-\phi_k(t))\eta_k'(t)\big)\Big|\\
&< (4n+1)\,\delta^2.
\end{aligned}
$$

It follows that

$$
\max\Big\{\big|x_j-u_j\big|, \big|y_j-v_j\big|, \frac{1}{\sqrt{4n+1}}\big|t-s+2\sum_{k=1}^{n}(y_ku_k - x_kv_k)\big|^{\frac12}\Big\} \le \rho(p_0,p_1)
$$

which completes the proof. □

As an indication of the role of this metric, consider the second order left-invariant non-elliptic operator on $\mathbb{H}_n$

$$
\mathcal{L}_\alpha = -\frac14\sum_{j=1}^{n}\big[(X^j)^2 + (Y^j)^2\big] + i\alpha T.
$$

THEOREM 2.6 (Folland, Stein[**FS74**]). *Set*

$$
\begin{aligned}
\phi_\alpha(x,y,t) &= (|x|^2 + |y|^2 - it)^{-(n+\alpha)/2)}\,(|x|^2 + |y|^2 + it)^{-(n-\alpha)/2}\\
\gamma_\alpha &= \frac{2^{2-n}\,\pi^{n+1}}{\Gamma\big((n+\alpha)/2\big)\,\Gamma\big((n-\alpha)/2\big)}
\end{aligned}
$$

Then $\mathcal{L}_\alpha = \gamma_\alpha\,\delta_0$. *Moreover*

$$
|\phi_\alpha(x,y,t)| \lesssim \rho((x,y,t),(0,0,0))^2\,|B(0,\rho((x,y,t),(0,0,0)))|^{-1}
$$

with corresponding estimates on the derivatives with respect to $\{X_j\}$ *and* $\{Y_k\}$.

2.4. The Grushin plane.

In our first two examples, the Carnot-Carathéodory metric was invariant under linear transformations, and in the third, the metric was invariant under the more complicated Heisenberg group of transformations. In this last example, the Carnot-Carathéodory balls 'twist' as one moves from point to point. Nevertheless, in all cases, the basic geometry of the Carnot-Carathéodory balls is the same at all points. This is a reflection of the fact that the linear relationships between the basic vector fields and their commutators remains the same at all points.

We now consider a class of examples in which the nature of the balls varies from point to point. We begin with the simplest case in which the underlying space is $\mathbb{R}^2$. Consider the vector fields

$$X = \frac{\partial}{\partial x}, \qquad Y = x\,\frac{\partial}{\partial y}, \qquad T = \frac{\partial}{\partial y}.$$

The vector fields X and T by themselves span the tangent space at each point, so the same is true for the triple $\{X, Y, T\}$. We assign degree 1 to the vector fields X and Y, and degree 2 to the vector field T. The commutation relationships between the vector fields is given by

$$[X, Y] = T, \qquad [X, T] = 0, \qquad [Y, T] = 0.$$

Thus $\mathcal{Y}_4 = \{X, Y, T; 1, 1, 2\}$ is a control system on $\mathbb{R}^2$.

There is one clear difference between this example and the three earlier ones. In the earlier cases, there were the same number of basic vector fields as the dimension of the underlying space, and these vector fields were linearly independent at each point. In this case, the number of vector fields is larger than the dimension. Moreover, the linear relationships between them change from point to point. The pair $\{X, Y\}$ is linearly dependent along the y-axis where $x = 0$ since $Y = 0$ on that locus. However at all other points, the pair $\{X, Y\}$ are linearly dependent.

The Carnot-Carathéodory ball centered at the point $(0, 0)$ of radius δ is essentially

$$\Big\{(x, y)\,\Big|\,|x| < \delta,\ |y| < \delta^2\Big\}.$$

The CC ball centered at the point $(0, a)$ of radius δ is essentially

$$\Big\{(x, y)\,\Big|\,|x| < \delta,\ |y| < \delta^2 + a\delta\Big\}.$$

3. Local structure of Carnot-Carathéodory metrics

We now return to the general theory of Carnot-Carathéodory metrics. Fix a connected open set $\Omega \subset \mathbb{R}^n$ and a control system $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ on Ω. Choose the compatible homogeneous norm

$$||(b_1, \ldots, b_q)|| = \sup_{1 \le j \le q} |b_j|^{\frac{1}{d_j}} \tag{3.1}$$

on $\mathbb{R}^q$. Let ρ be the corresponding Carnot-Carathéodory metric, and let $\{B_\rho(x, \delta)\}$ be the associated family of balls. Our objective in this section is to give an alternate characterization of these balls in terms of another family of sets $\big\{B_\eta^I(x, \delta)\big\}$, called *exponential balls*, which are defined using appropriately chosen exponential maps and canonical coordinates. This is primarily based on the paper [**NSW85**].

Because these new sets are defined explicitly as images of exponential mappings, it is easier to understand the local geometry of exponential balls than Carnot-Carathéodory balls. On the other hand, the main structure theorem for Carnot-Carathéodory balls says, in particular, that for $x \in \Omega$ and $\delta > 0$ sufficiently small, there is a choice of I and η so that the balls $B_\rho(x,\delta)$ and $B^I_\eta(x,\delta)$ are comparable.

In subsection 3.1 below, introduce some important notation. In subsection 3.2, we give the definition of exponential balls, and we show that we always have an inclusion $B_\rho(x,\delta) \subset B^I_\eta(x,\delta)$. In order to establish the opposite inclusion, we will need to make a careful choice of I and η for each x and δ. The key to this choice is the notion that an n-tuple $I = (i_1,\dots,i_n)$ is (x,δ,κ)-dominant, and we give a precise definition of this concept in subsection 3.3. In subsection 3.4 we state the main result, Theorem 3.7, about the equivalence of exponential and Carnot-Carathéodory balls. The proof of this theorem, which is quite long and technical, is deferred to Section 4.

3.1. Notation.

We begin by reviewing and establishing some notation.

Definition 3.1. *Let $\mathcal{Y} = \{Y_1,\dots,Y_q; d_1,\dots,d_q\}$ be a control system on a connected open set $\Omega \subset \mathbb{R}^n$.*

(1) For each positive integer r, let $\mathbb{I}_r$ denote the set of all ordered r-tuples of positive integers $I = (i_1,\dots,i_r)$ such that $1 \le i_j \le q$ for $1 \le j \le r$.

(2) If $I \in \mathbb{I}_r$ set

$$d(I) = d_{i_1} + \cdots + d_{i_r}. \tag{3.2}$$

(3) Put

$$\begin{aligned} d_{\mathcal{Y}} = d &= \min_{1\le j\le q} d_j,\\ D_{\mathcal{Y}} = D &= \max_{1\le j\le q} d_j,\\ N_0 &= \max_{I,J\in\mathbb{I}_n} |d(I)-d(J)|,\\ N_1 &= \inf\left\{N \in \mathbb{Z}^+ \,\middle|\, Nd \ge N_0 + nD\right\}. \end{aligned} \tag{3.3}$$

(4) For every $I = (i_1,\dots,i_n) \in \mathbb{I}_n$ let

$$\lambda_I(x) = \det\big(Y_{i_1},\dots,Y_{i_n}\big)(x) \tag{3.4}$$

be the determinant of the vector fields $\{Y_{i_1},\dots,Y_{i_n}\}$ at the point x as defined in Chapter 2, Definition 1.7.

(5) Set

$$\Omega_I = \left\{x \in \Omega \,\middle|\, \lambda_I(x) \ne 0\right\}. \tag{3.5}$$

Note that since $\mathcal{Y}$ is a control system, for each $x \in \Omega$ there exists $I \in \mathbb{I}_n$ so that $\lambda_I(x) \ne 0$. For such x the vectors $\{(Y_{i_1})_x,\dots,(Y_{i_n})_x\}$ are linearly independent and hence are a basis for the tangent space T_x.

(6) *For $x \in \Omega$ and $\delta > 0$, set*

$$\Lambda(x,\delta) = \sum_{I \in \mathbb{I}_n} |\lambda_I(x)|\, \delta^{d(I)}. \tag{3.6}$$

There are q^n elements in $\mathbb{I}_n$, so this is a finite sum.

Note that $\Lambda(x,\delta)$ is a non-zero polynomial in δ of degree at most nD. For each $x \in \Omega$, each non-vanishing term in $\Lambda(x,\delta)$ is of degree at least nd Moreover

$$\sup_{J \in \mathbb{I}_n} \big|\lambda_J(x)\big|\, \delta^{d(J)} \leq \Lambda(x,\delta) \leq q^n \sup_{J \in \mathbb{I}_n} \big|\lambda_J(x)\big|\, \delta^{d(J)}. \tag{3.7}$$

3.2. Exponential Balls.

Recall from Chapter 2, equation (5.1), that for each $I \in \mathbb{I}_n$ and each $x \in \Omega$ we have a mapping

$$\Theta^I_x(u) = \Theta^I_x(u_1,\ldots,u_n) = \exp\Big[\sum_{j=1}^n u_j\, Y_{i_j}\Big](x) \tag{3.8}$$

defined for $u = (u_1,\ldots,u_n)$ belonging to some neighborhood of the origin in $\mathbb{R}^n$. It follows from Chapter 2, Lemma 5.1, that if $x \in \Omega_I$ then $d\Theta^I_x(0) = \big|\lambda_I(x)\big| \neq 0$, and hence Θ^I_x is a diffeomoprhism from a neighborhood of the origin in $\mathbb{R}^n$ to a neighborhood of x in Ω.

One of our objectives is to obtain a more precise description of these neighborhoods when we make an appropriate choice of I. In order to do this, we introduce a variant of the mapping Θ^I_x.

DEFINITION 3.2. *Let $I \in \mathbb{I}_n$, let $x \in \Omega$, and let δ and η be positive parameters. Set*

$$\Theta^I_{x,\delta,\eta}(u) = \Theta^I_{x,\delta,\eta}(u_1,\ldots,u_n) = \exp\Big[\eta \sum_{j=1}^n u_j\, \delta^{d_{i_j}}\, Y_{i_j}\Big](x).$$

We have

$$\begin{aligned} \Theta^I_{x,\delta,\eta}(u) &= \Theta^I_x(\eta\, \delta^{d_{i_1}}\, u_1,\ldots,\eta\, \delta^{d_{i_n}}\, u_n),\\ \Theta^I_{x,\delta,\eta_1\eta_2}(u) &= \Theta^I_{x,\delta,\eta_1}(\eta_2 u), \end{aligned}$$

whenever these expressions are defined. Thus $\Theta^I_{x,\delta,\eta}$ is just a re-scalled version of the mapping Θ^I_x. Let us try to give an explanation of the purpose of the various extra parameters that appear in $\Theta^I_{x,\delta,\eta}$.

(1) Clearly $\Theta^I_{x,\delta,\eta}(0) = x$, and thus x is the 'center' of the image of $\mathbb{B}(0;1)$.
(2) When considering a Carnot-Carathéodory ball $B_\rho(x,\delta)$, the parameter δ is the radius. For the mapping $\Theta^I_{x,\delta,\eta}$, the parameter δ is the corresponding 'scale' at which we are working. Note that the parameter δ appears with powers reflecting the formal degrees of the vector fields $\{Y_{i_1},\ldots,Y_{i_n}\}$.
(3) For each $x \in \Omega$ and $\delta > 0$ we need to choose a coordinate system determined by the canonical coordinates associated to particular 'dominant' set of vector fields $\{Y_{i_1},\ldots,Y_{i_n}\}$. This choice is described by the index $I = (i_1,\ldots,i_n)$.

(4) If we were to eliminate the parameter η in the definition of $\Theta^I_{x,\delta,\eta}$ by setting it equal to 1, then in order to establish good properties of the mapping $u \to \Theta^I_{x,\delta,1}(u)$ we would need to restrict the mapping to the a ball in $\mathbb{R}^n$ of small radius. The parameter η is introduced in order to keep the domain of the mapping the unit ball $\mathbb{B}(0;1) \subset \mathbb{R}^n$. Thus the parameter η is just ordinary Euclidean scaling, unlike the non-isotropic scaling introduced by δ. As we proceed, we shall need to shrink η several times.

The following result follows from the discussion of existence, uniqueness, and smooth dependence on parameters of solutions of differential equations in Chapter 2, Section 3.3.

Proposition 3.3. *Let $K \subset \widetilde{K} \subset \Omega$ and suppose that K and $\widetilde{K}$ are compact and that K is a subset of the interior of $\widetilde{K}$. Then there exists $0 < \eta_0 \leq 1$ so that if $x \in K$, $I \in \mathbb{I}_n$, $0 < \delta \leq 1$, and $0 < \eta \leq \eta_0$ then the mapping $u \to \Theta^I_{x,\delta,\eta}(u)$ is defined and infinitely differentiable on the open unit ball $\mathbb{B}(0;1) \subset \mathbb{R}^n$. Moreover if $u \in \mathbb{B}(0;1)$ then $\Theta^I_{x,\delta,\eta}(u) \in \widetilde{K}$.*

Now fix compact sets $K \subset \widetilde{K} \subset \Omega$ and the corresponding constant η_0 given in Proposition 3.3. We use the mappings $\{\Theta^I_{x,\delta,\eta}\}$ from Definition 3.2 to define *exponential balls* centered at $x \in K$ of radius δ.

Definition 3.4. *Let $x \in K$, let $I \in \mathbb{I}_n$, let $0 < \delta \leq 1$ and let $0 < \eta \leq \eta_0$. Then*

$$B^I_\eta(x,\delta) = \Big\{ y \in \Omega \,\Big|\, \Big(\exists u \in \mathbb{B}(0;1)\Big)\, \Big(y = \Theta^I_{x,\delta,\eta}(u)\Big)\Big\} \subset \widetilde{K}.$$

The sets $\big\{B^I_\eta(x,\delta)\big\}$ are called exponential balls.

Recall that we have two different families of dilations on $\mathbb{R}^n$. Suppose that $\sigma > 0$ and that $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$. Then we put

$$\begin{aligned} \sigma\, u &= \big(\sigma u_1, \ldots, \sigma u_n\big) \\ D_\sigma[u] &= \big(\sigma^{d_{i_1}} u_1, \ldots, \sigma^{d_{i_n}} u_n\big). \end{aligned}$$

Then we have

$$\begin{aligned} \Theta^I_{x,\delta,\sigma\eta}(u) &= \Theta^I_{x,\delta,\eta}(\sigma u) \\ \Theta^I_{x,\sigma\delta,\eta}(u) &= \Theta^I_{x,\delta,\eta}(D_\sigma[u]). \end{aligned}$$

We have the following easy but important relationship between the Carnot-Carathéodory balls $\big\{B_\rho(x,\delta)\big\}$ and the exponential balls $\big\{B^I_\eta(x,\delta)\big\}$.

Proposition 3.5. *Let $x \in K$ and $0 < \delta \leq 1$. Then for every $I \in \mathbb{I}_n$ and $0 < \eta \leq \eta_0$ we have*

$$B^I_\eta(x,\delta) \subset B_\rho(x,\delta).$$

Proof. Let $u \in \mathbb{B}(0;1)$, and define $\varphi : [0,1] \to \Omega$ by setting

$$\varphi(t) = \exp\Big[t\,\eta \sum_{j=1}^n u_j\, \delta^{d_{i_j}}\, Y_{i_j}\Big](x).$$

Then $\varphi(0) = x$, $\varphi(1) = \Theta^I_{x,\delta,\eta}(u)$, and since $\eta \sum_{j=1}^n u_r\, \delta^{d_{i_j}}\, Y_{i_j}$ is a smooth vector field, it follows from Corollary 3.7 in Chapter 2 that

$$\varphi'(t) = \Big(\eta \sum_{j=1}^n u_r\, \delta^{d_{i_j}}\, Y_{i_j}\Big)_{\varphi(t)}.$$

Since each $\eta\, u \in \mathbb{B}(0;1)$, it follows from our choice of the homogeneous norm in equation (3.1) that

$$\big|\big|\,(\delta^{d_1}\,\eta\, u_1, \ldots, \delta^{d_n}\,\eta\, u_n)\,\big|\big| = \delta \sup_{1\le j\le q} (\eta|u_j|)^{\frac{1}{d_j}} < \delta.$$

Thus according to Definition 1.5, this means that $\varphi \in AC\big(x, \Theta^I_{x,\delta,\eta}(u); \delta\big)$, and so $AC\big(x, \Theta^I_{x,\delta,\eta}(u); \delta\big)$ is not empty. Thus according to Definition 1.6, $\rho\big(x, \Theta^I_{x,\delta,\eta}(u)\big) < \delta$, and so $B^I_\eta(x,\delta) \subset B_\rho(x,\delta)$, as asserted. □

3.3. (x, δ, η)-dominance.

We want to show that if $x \in K$ and if $0 < \delta \le 1$, then for an appropriate choice of n-tuple $I \in \mathbb{I}_n$ and all sufficiently small η, the mapping $\Theta^I_{x,\delta,\eta} : \mathbb{B}(0;1) \to B^I_\eta(x,\delta)$ is a diffeomorphism, and the exponential ball $B^I_\eta(x,\delta)$ is comparable to the Carnot-Carathéodory ball $B_\rho(x,\delta)$. When this is the case, the inverse mapping $(\Theta^I_{x,\delta,\eta})^{-1}$ will give a natural set of coordinates on the exponential ball $B^I_\eta(x,\delta)$, and hence also on the comparable Carnot-Carathéodory ball. The choice of the n-tuple I which makes this work depends on the following key concept.

Definition 3.6. *Let $x \in \Omega$, let $\delta > 0$, and let $0 < \kappa \le 1$. Then $I \in \mathbb{I}_n$ is* (x,δ,κ)-dominant *if*

$$|\lambda_I(x)|\, \delta^{d(I)} \ge \kappa \max_{J\in\mathbb{I}_n} |\lambda_J(x)|\, \delta^{d(J)}.$$

For each $x \in \Omega$ and $\delta > 0$ there always exists at least one n-tuple $I \in \mathbb{I}_n$ so that

$$|\lambda_I(x)|\, \delta^{d(I)} = \max_{J\in\mathbb{I}_n} |\lambda_J(x)|\, \delta^{d(J)},$$

and thus I is $(x,\delta,1)$-dominant. However, if I is $(x,\delta,1)$-dominant, there need not exist any open neighborhood U of x so that I is $(y,\delta,1)$-dominant for all $y \in U$. On the other hand, if I is $(x,\delta,1)$-dominant and if $\kappa < 1$, it follows by continuity that there will be an open neighborhood U so that I is (y,δ,κ)-dominant for all $y \in U$. This is one of the reasons for the introduction of the parameter κ.

3.4. The local structure of Carnot-Carathéodory balls.

We now state the main result of this chapter, which describes the local geometry of the Carnot-Carathéodory balls $\big\{B_\rho(x,\delta)\big\}$ by relating them to the exponential balls $\big\{B^I_\eta(x,\delta)\big\}$ defined in Section 3.1. As usual, $\Omega \subset \mathbb{R}^n$ is a connected open set, and $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ is a control system on Ω. The constants d, D, N_0, and N_1 are defined in equation (3.3).

THEOREM 3.7. *Let $K \subset \widetilde{K} \subset \Omega$ be compact subsets with K contained in the interior of $\widetilde{K}$. There exist positive constants $0 < \tau < 1$ and $0 < \eta < 1$ depending only on $\nu(K)$ and $\big|\big|\mathcal{Y}\big|\big|_{\widetilde{K},N_1}$ so that if $x \in K$ and if $0 < \delta \leq 1$, there exists $I \in \mathbb{I}_n$ so that the following statements hold.*

(1) *If $J\Theta^I_{x,\delta,\eta}$ is the Jacobian determinant of the mapping $\Theta^I_{x,\delta,\eta}$, then*
$$J\Theta^I_{x,\delta,\eta}(0) = \eta^n \big|\lambda_I(x)\big|\, \delta^{d(I)},$$
and for every $u \in \mathbb{B}(0;1)$ we have
$$\frac{\eta^n}{4}|\lambda_I(x)|\,\delta^{d(I)} \leq J\Theta^I_{x,\delta,\eta}(u) \leq 4\,\eta^n\,|\lambda_I(x)|\,\delta^{d(I)}. \tag{3.9}$$

(2) *The balls $B_\rho(x,\delta)$ and $B^I_\eta(x,\delta)$ are comparable. Precisely, we have*
$$B_\rho(x,\tau\,\delta) \subset B^I_\eta(x,\delta) \subset B_\rho(x,\delta). \tag{3.10}$$

(3) *The mapping $\Theta^I_{x,\delta,\eta} : \mathbb{B}(0;1) \to B^I_\eta(x,\delta)$ is one-to-one and onto.*

(4) *Let $\{Z_1,\ldots,Z_q\}$ be the unique smooth real vector fields on the unit ball $\mathbb{B}(0;1)$ such that*
$$d\Theta^I_{x,\delta,\eta}[Z_j] = \eta\,\delta^{d_j}\,Y_j. \tag{3.11}$$
Then $\mathcal{Z} = \{Z_1,\ldots,Z_q; d_1,\ldots,z_q\}$ is a control structure on $\mathbb{B}(0;1)$ uniform in x and δ. Precisely
$$\begin{aligned} ||\mathcal{Z}||_{\mathbb{B}(0;1),M} &\leq ||\mathcal{Y}||_{K,M}, \\ \tau_2(\mathcal{Z}) &\geq \tau_2(\mathcal{Y}). \end{aligned} \tag{3.12}$$
Moreover, for the special vector fields $\{Z_{i_1},\ldots,Z_{i_n}\}$, we can write
$$Z_{i_j} = \frac{\partial}{\partial u_j} + \sum_{k=1}^n \alpha_{j,k}\,\frac{\partial}{\partial u_k} \tag{3.13}$$
and
$$\sup_{u\in\mathbb{B}(0;1)} \sum_{j,k=1}^n |\alpha_{j,k}(u)|^2 \leq \frac{1}{4}. \tag{3.14}$$

An immediate consequence of parts (3) and (1) of Theorem 3.7 is that
$$\frac{\eta^n}{4}\,c_n\,\big|\lambda_I(x)\big|\,\delta^{d(I)} \leq \big|B^I_\eta(x,\delta)\big| \leq 4\,\eta^n\,c_n\,\big|\lambda_I(x)\big|\,\delta^{d(I)}$$
where $\big|E\big|$ is the Lebesgue measure of a set $E \subset \mathbb{R}^n$, and c_n is the volume of the unit ball in $\mathbb{R}^n$. Thus equation (3.7) gives the following estimate for the volume of the Carnot-Carathéodory ball $B_\rho(x,\delta)$.

COROLLARY 3.8. *Let $K \subset \Omega$ be a compact subset. There exists a constants $\eta > 0$ and $C > 0$ so that if $x \in K$ and if $0 < \delta \leq 1$, then*
$$C^{-1}\,\eta^n\,\Lambda(x,\delta) \leq \big|B_\rho(x,\delta)\big| \leq C\,\eta^n\,\Lambda(x,\delta).$$

4. Proof of the main theorem

In this section, we present a proof of Theorem 3.7. Fix a connected open subset $\Omega \subset \mathbb{R}^n$, and let $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ be a control system on Ω. Fix compact subsets $K \subset \widetilde{K} \subset \Omega$ such that K is contained in the interior of $\widetilde{K}$. Let η_0 be the constant given in Proposition 3.3.

The proof is rather long and technical. Subsection 4.1 contains some preliminary results of an algebraic nature that do not involve any estimates. Subsection 4.2 then uses these algebraic results to obtain estimates at a point x assuming that an n-tuple $I \in \mathbb{I}_n$ is (x, δ, κ)-dominant. Subsection 4.3 uses the theory of Taylor expansions developed in Chapter 2, along with the results of subsection 4.2, to obtain estimates on the exponential balls $\{B^I_\eta(x,\delta)\}$. Subsection 4.4 presents some topological results needed later in the proof. Finally, subsections 4.5, 4.6, 4.7, and 4.8 present the proofs of parts (1), (2), (3), and (4) of Theorem 3.7.

4.1. Algebraic preliminaries.

Recall from (H-2) of Definition 1.3 that there are functions $c^m_{j,k} \in \mathcal{E}(\Omega)$ so that for $1 \leq l, m \leq q$

$$[Y_j,\, Y_k] = \sum_{m=1}^{q} c^m_{j,k}\, Y_m \tag{4.1}$$

and

$$c^m_{j,k}(x) \equiv 0 \qquad \text{unless } d_m \leq d_j + d_k. \tag{4.2}$$

The coefficients $\{c^m_{j,k}\}$ are called the *structure functions* of the control system $\mathcal{Y}$. More generally, we have the following.

PROPOSITION 4.1. *Let if $L = (l_1, \ldots, l_r) \in \mathbb{I}_r$. Then we can write the iterated commutator*

$$Y_{[L]} = \big[Y_{l_r},\, \big[Y_{i_{r-1}},\, \cdots [Y_{l_2},\, Y_{l_1}]\big]\big] = \sum_{m=1}^{q} c^m_L\, Y_j \tag{4.3}$$

where $c^m_L \in \mathcal{E}(\Omega)$ and

$$c^m_L(x) \equiv 0 \text{ unless } d_m \leq d_{l_1} + \cdots + d_{l_r} = d(L). \tag{4.4}$$

The coefficients $\{c^j_L\}$ are linear combinations of the structure functions and their Y-derivatives up to order $(r-1)$.

PROOF. We argue by induction on $r \geq 2$. The case $r = 2$ is the hypothesis (H-2) of Definition 1.3. Let $L' = (l_1, \ldots, l_r, l_{r+1}) \in \mathbb{I}_{r+1}$ and $L = (l_1, \ldots, l_r)$. Then

$$\begin{aligned}
Y_{[L']} &= \big[Y_{l_{r+1}},\, [Y_{i_r},\, \cdots [Y_{l_2},\, Y_{l_1}]]\big] \\
&= \big[Y_{l_{r+1}},\, Y_{[L]}\big] \\
&= \sum_{m=1}^{q} \big[Y_{l_{r+1}},\, c^m_L Y_m\big] \\
&= \sum_{m=1}^{q} Y_{l_{r+1}}[c^m_L]\, Y_m + \sum_{k=1}^{q} c^k_L\, \big[Y_{l_{r+1}},\, Y_k\big] \\
&= \sum_{m=1}^{q} \Big[Y_{l_{r+1}}[c^m_L] + \Big(\sum_{k=1}^{q} c^k_L\, c^m_{l_{r+1},k}\Big)\Big] Y_m.
\end{aligned} \tag{4.5}$$

Now if $Y_{l_{r+1}}[c_L^m] \neq 0$, the by induction we have $d_m \leq d(L) < d(L')$. Also if $c_L^k\, c^m_{l_{r+1},k} \neq 0$ we must have $d_k \leq d(L)$ and $d_m \leq d_{l_{r+1}} + d_k$. Thus $d_m \leq d_{l_{r+1}} + d(L) = d(L')$. The statement about the structure of the coefficients also follows by induction and an examination of the formula in equation (4.5). This completes the proof. □

For the rest of this subsection, fix $I \in \mathbb{I}_n$. We shall suppose that $I = (1, \dots, n)$ in order to simplify the notational burden[6]. This can always be achieved by re-ordering the elements of $\{Y_1, \dots, Y_q\}$. For each point $x \in \Omega_I$, the tangent vectors $\{(Y_1)_x, \dots, (Y_n)_x\}$ span the tangent space T_x. Thus there are real numbers[7] $\{a_j^k(x)\}$ so that for $1 \leq j \leq q$ we can write

$$(Y_j)_x = \sum_{k=1}^{n} a_j^k(x)\,(Y_k)_x. \tag{4.6}$$

We can now use Cramer's rule to solve equation (4.6) for the n unknown values $\{a_j^1(x), \dots, a_j^n(x)\}$.

Proposition 4.2. *For $1 \leq j \leq q$, for $1 \leq k \leq n$, and for each $x \in \Omega_I$*

$$a_j^k(x) = \frac{\lambda_J(x)}{\lambda_I(x)} \tag{4.7}$$

where $J = (1, \dots, k-1, j, k+1, \dots, n) \in \mathbb{I}_n$ is obtained from $I = (1, \dots, n)$ by replacing k with j.

Note that $\mathcal{E}(\Omega_I)$ is a module over the ring $\mathcal{E}(\Omega)$. We need to deal with linear combinations of products of the functions $\{a_j^k\}$, with coefficients coming from $\mathcal{E}(\Omega)$. It will be convenient to keep track of these expressions in the following way.

Definition 4.3. *For every integer m and every positive integer s, $\mathcal{A}_s^m$ is the $\mathcal{E}(\Omega)$ submodule of $\mathcal{E}(\Omega_I)$ generated by all products of the form*

$$a_{j_1}^{k_1} \cdot a_{j_2}^{k_2} \cdots a_{j_r}^{k_r} \tag{4.8}$$

where $r \leq s$ and

$$m \leq (d_{k_1} + d_{k_2} + \cdots + d_{k_r}) - (d_{j_1} + d_{j_2} + \cdots + d_{j_r}) = d(K) - d(J). \tag{4.9}$$

We let $\mathcal{A}_1^0$ be the submodule generated by the function 1, so that $\mathcal{A}_1^0 = \mathcal{E}(\Omega)$. If $f \in \mathcal{A}_s^m$, we can write

$$f = \sum_{r=1}^{s} \sum_{\substack{J,K \in \mathbb{I}_r \\ m \leq d(K) - d(J)}} f_{K,J}\, a_{j_1}^{k_1} \cdot a_{j_2}^{k_2} \cdots a_{j_r}^{k_r}, \tag{4.10}$$

although the coefficients $f_{K,J} \in \mathcal{E}(\Omega)$ are not uniquely determined. We have the inclusions

$$\begin{aligned} \mathcal{A}_s^{m_1} &\supset \mathcal{A}_s^{m_2} && \text{if } m_1 \leq m_2, \\ \mathcal{A}_{s_1}^{m} &\subset \mathcal{A}_{s_2}^{m} && \text{if } s_1 \leq s_2. \end{aligned}$$

---

[6]This simplification is convenient, and we shall do this several times, even though the choice of n-tuple I may have changed.

[7]The numbers $\{a_j^1(x), \dots, a_j^n(x)\}$ depend on the choice of n-tuple I, and should perhaps be written $\{a_j^{I,1}(x), \dots, a_j^{I,n}(x)\}$. As long as I is understood, we suppress the additional parameter.

We show that the modules $\{\mathcal{A}_s^m\}$ behave well under differentiation by the vector fields $\{Y_1, \ldots, Y_q\}$. The key is the following calculation.

PROPOSITION 4.4. *For $1 \leq l \leq q$, $1 \leq k \leq n$, and $1 \leq j \leq q$ we have*

$$Y_l[a_j^k] = \sum_{m=1}^{q} c_{l,j}^m \, a_m^k + \sum_{i=1}^{n} \sum_{m=1}^{q} c_{l,i}^m \, a_j^i \, a_m^k. \tag{4.11}$$

In particular, $Y_l[a_j^k] \in \mathcal{A}_2^{d_k - d_j - d_l}$.

PROOF. Using equations (4.1) and (4.6), it follows that on Ω_I we can write

$$[Y_l, Y_j] = \sum_{k=1}^{n} \Big[\sum_{m=1}^{q} c_{l,j}^m \, a_m^k \Big] Y_k. \tag{4.12}$$

On the other hand, we can also write

$$\begin{aligned}
[Y_l, Y_j] &= \sum_{i=1}^{n} \big[Y_l, \, a_j^i \, Y_i\big] \\
&= \sum_{i=1}^{n} \big(Y_l[a_j^i] \, Y_i + a_j^i \, [Y_l, \, Y_i] \big) \\
&= \sum_{i=1}^{n} Y_l[a_j^i] \, Y_i + \sum_{i=1}^{n} \sum_{m=1}^{q} a_j^i \, c_{l,i}^m \, Y_m \\
&= \sum_{k=1}^{n} Y_l[a_j^k] \, Y_k + \sum_{i=1}^{n} \sum_{m=1}^{q} \sum_{k=1}^{n} a_j^i \, c_{l,i}^m \, a_m^k \, Y_k \\
&= \sum_{k=1}^{n} \Big[Y_l[a_j^k] + \sum_{i=1}^{n} \sum_{m=1}^{q} c_{l,i}^m \, a_j^i \, a_m^k \Big] Y_k.
\end{aligned} \tag{4.13}$$

Since $\{(Y_1)_x, \ldots, (Y_n)_x\}$ is a basis for T_x on Ω_I we can equate the coefficients of Y_k in equations (4.12) and (4.13), and we obtain equation (4.11).

In the first sum on the right of equation (4.11) we have $c_{l,j}^m \equiv 0$ unless $d_m \leq d_l + d_j$, and so $d_k - d_m \geq d_k - d_j - d_l$. Thus the first sum belongs to $\mathcal{A}_1^{d_k - d_j - d_l} \subset \mathcal{A}_2^{d_k - d_j - d_l}$. In the second sum, we have $c_{l,i}^m \equiv 0$ unless $d_m \leq d_l + d_i$. Thus $(d_i + d_k) - (d_j + d_m) \geq d_i + d_k - d_j - d_l - d_i = d_k - d_j - d_l$. Thus the second sum belongs to $\mathcal{A}_2^{d_k - d_j - d_l}$, and this completes the proof. □

We can now describe Y-derivatives of elements of $\mathcal{A}_s^m$. The following follows easily from the product rule and Proposition 4.4.

PROPOSITION 4.5. *Let $f \in \mathcal{A}_s^m$, and suppose f is written as in equation (4.10) with coefficients $\{f_{K,J}\}$. Then*

(1) For $1 \leq l \leq q$,

$$Y_l[f] \in \mathcal{A}_{s+1}^{m - d_l},$$

with coefficients which can be written as linear combinations of the functions $\big\{Y_l[f_{K,J}]\big\}$ and products of the functions $\big\{f_{K,J}\big\}$ and the structure functions $\big\{c_{l,i}^m\big\}$.

(2) *More generally, if* $L = (l_1, \ldots, l_r) \in \mathbb{I}_r$ *then*

$$Y_L[f] = Y_{i_r} Y_{i_{r-1}} \cdots Y_{i_2} Y_{i_1}[f] \in \mathcal{A}_{s+r}^{m-d(L)}$$

and the coefficients of $Y_L[f]$ *can be written as linear combinations of products of functions which are up to* r*-fold derivatives of the functions* $f_{K,J}$ *and* $r-1$*-fold derivatives of the coefficient fucntions* $\{c_{l,i}^m\}$ *with respect to the vector fields* $\{Y_{i_1}, \ldots, Y_{i_r}\}$.

We next derive formulas for derivatives of the determinants $\{\lambda_J\}$, $J \in \mathbb{I}_n$. We begin with the special determinant λ_I where $I = (1, \ldots, n)$.

PROPOSITION 4.6. *For* $x \in \Omega_I$ *and* $1 \leq l \leq q$ *we have* $Y_l[\lambda_I] = f_l\, \lambda_I$ *where*

$$f_l(x) = \nabla \cdot Y_l(x) + \sum_{k=1}^{n} \sum_{j=1}^{q} c_{l,k}^j(x)\, a_j^k(x).$$

In particular, the coefficient $f_l \in \mathcal{A}_1^{-d_l}$, *and is a linear combination of derivatives of the coefficients of the vector field* Y_l *and the structure functions.*

PROOF. Starting with the formula from Lemma 1.9 in Chapter 2, and then using equations (4.1) and (4.6), we have

$$\begin{aligned}
Y_l[\lambda_I] &= (\nabla \cdot Y_l)\, \lambda_I + \sum_{k=1}^{n} \det(Y_1, \ldots, Y_{k-1}, [Y_l, Y_k], Y_{k+1}, \ldots, Y_n) \\
&= (\nabla \cdot Y_l)\, \lambda_I + \sum_{k=1}^{n} \sum_{j=1}^{q} c_{l,k}^j \det(Y_1, \ldots, Y_{k-1}, Y_j, Y_{k+1}, \ldots, Y_n) \\
&= (\nabla \cdot Y_l)\, \lambda_I + \sum_{k=1}^{n} \sum_{j=1}^{q} \sum_{r=1}^{n} c_{l,k}^j\, a_j^r \det(Y_1, \ldots, Y_{k-1}, Y_r, Y_{k+1}, \ldots, Y_n) \\
&= (\nabla \cdot Y_l)\, \lambda_I + \sum_{k=1}^{n} \sum_{j=1}^{q} c_{l,k}^j\, a_j^k \det(Y_1, \ldots, Y_{k-1}, Y_k, Y_{k+1}, \ldots, Y_n) \\
&= \Big(\nabla \cdot Y_l + \sum_{k=1}^{n} \sum_{j=1}^{q} c_{l,k}^j\, a_j^k\Big)\, \lambda_I.
\end{aligned}$$

Now $\nabla \cdot Y_l \in \mathcal{A}_1^0 \subset \mathcal{A}_1^{-d_l}$. Moreover, since $c_{l,k}^j(x) \equiv 0$ unless $d_j \leq d_l + d_k$, $d_k - d_j \geq d_k - d_l - d_k = -d_l$. This completes the proof. □

COROLLARY 4.7. *Let* $L = (l_1, \ldots, l_r) \in \mathbb{I}_n$. *Then*

$$Y_{l_1} Y_{l_2} \cdots Y_{l_r}[\lambda_I] = f_L\, \lambda_I,$$

where $f_L \in \mathcal{A}_r^{-d(L)}$. *The coefficients of* f_L *can be written as linear combinations of up to* r*-fold derivatives of the coefficients of* Y_j *and up to* $(r-1)$*-fold derivatives of the structure functions.*

PROOF. This follows easily by induction, using Propositions 4.5 and 4.6. □

PROPOSITION 4.8. *Let* $L = (l_1, \ldots, l_r) \in \mathbb{I}_r$, *and let* $J \in \mathbb{I}_n$. *Then*

$$Y_{l_1} Y_{l_2} \cdots Y_{l_r}[\lambda_J] = f_{L,J}\, \lambda_I$$

where $f_{L,J} \in \mathcal{A}_{n+r}^{d(I)-d(J)-d(K)}$. The coefficients of $f_{L,J}$ can be written as linear combinations of up to r-fold derivatives of the coefficients of the vector fields $\{Y_1, \ldots, Y_q\}$ and up to $(r-1)$-fold derivatives of the structure functions.

PROOF. Let $J = (j_1, \ldots, j_n)$. Then

$$\begin{aligned}
\lambda_J &= \det(Y_{j_1}, \ldots, Y_{j_n}) \\
&= \det\left(\sum_{k_1=1}^{n} a_{j_1}^{k_1}\, Y_{k_1}, \ldots, \sum_{k_n=1}^{n} a_{j_n}^{k_n}\, Y_{k_n}\right) \\
&= \sum_{k_1=1}^{n} \cdots \sum_{k_n=1}^{n} a_{j_1}^{k_1} \cdots a_{j_n}^{k_n}\, \det(Y_{k_1}, \ldots, Y_{k_n}) \\
&= \Big[\sum_{\sigma \in \mathfrak{S}_n} (-1)^{\sigma} a_{j_1}^{\sigma(1)} \cdots a_{j_n}^{\sigma(n)}\Big]\, \lambda_I \\
&= f_J\, \lambda_I,
\end{aligned}$$

where $\mathfrak{S}_n$ is the group of permutations of $\{1, \ldots, n\}$. Now if $\sigma \in \mathfrak{S}_n$,

$$\big(d_{\sigma(1)} + \cdots + d_{\sigma(n)}\big) - \big(d_{j_1} + \cdots + d_{j_n}\big) = d(I) - d(J).$$

Thus $f_J \in \mathcal{A}_n^{d(I)-d(J)}$. The proposition then follows by using Proposition 4.5 and Corollary 4.7. □

4.2. Estimates at x using (x, δ, κ)-dominance.

If we assume that an n-tuple $I \in \mathbb{I}_n$ is (x, δ, κ)-dominant, we can obtain estimates at x for functions $f \in \mathcal{A}_s^m$ and for the determinant functions $\{\lambda_J\}$.

PROPOSITION 4.9. *Let $0 < \kappa \leq 1$, let $0 < \delta \leq 1$, and let $x \in \Omega$. Suppose $I \in \mathbb{I}_n$ is (x, δ, κ)-dominant. We assume, for simplicity of notation, that $I = (1, \ldots, n)$, so we can use the notation introduced in subsection 4.1.*

(1) Every generator $a_{j_1}^{k_1} \cdots a_{j_r}^{k_r}$ of $\mathcal{A}_s^m$ satisfies

$$\big|a_{j_1}^{k_1}(x) \cdots a_{j_r}^{k_r}(x)\big| \leq \kappa^{-s}\, \delta^m. \tag{4.14}$$

(2) Let $L = (l_1, \ldots, l_r) \in \mathbb{I}_r$. Suppose that $f \in \mathcal{A}_s^m$. Then

$$\big|Y_{l_1} \cdots Y_{l_r}[f](x)\big| \leq C\, \kappa^{-s-r}\, \delta^{m-d(L)} \tag{4.15}$$

where C depends only on the supremum at x of up to r derivatives of the coefficients of f and up to $(r-1)$ derivatives of the coefficients of the vector fields $\{Y_{l_1}, \ldots, Y_{l_r}\}$ and the structure functions $\{c_{l,j}^m\}$.

(3) Let $L = (l_1, \ldots, l_r) \in \mathbb{I}_r$, and suppose $J \in \mathbb{I}_n$. Then

$$\big|Y_{l_1} \cdots Y_{l_r}[\lambda_J](x)\big| \leq C\, \kappa^{-n-r}\, \delta^{d(I)-d(J)-d(L)}\, \big|\lambda_I(x)\big|, \tag{4.16}$$

where C depends only on the supremum at x of up to r derivatives of the coefficients of the vector fields $\{Y_1, \ldots, Y_q\}$, and up to $(r-1)$ derivatives of the structure functions.

PROOF. Since I is (x,δ,κ)-dominant and $\lambda_J(x) \neq 0$ for some $J \in \mathbb{I}_n$, we certainly have $x \in \Omega_I$. Thus according to Proposition 4.2, $a_j^k(x) = \lambda_J(x)/\lambda_I(x)$ where J is obtained from I by replacing k by j. However, since I is (x,δ,κ)-dominant, this implies that

$$\left|a_j^k(x)\right| \leq \kappa^{-1}\,\delta^{d(I)-d(J)} = \kappa^{-1}\,\delta^{d_k-d_j}.$$

Hence

$$\left|a_{j_1}^{k_1}(x)\cdots a_{j_r}^{k_r}(x)\right| \leq \kappa^{-r}\,\delta^{(d_{k_1}+\cdots+d_{k_r})-(d_{j_1}+\cdots+d_{j_r})} \leq \kappa^{-s}\,\delta^m,$$

and this gives statement (1).

It follows from Proposition 4.5 that if $L = (l_1,\ldots,l_r) \in \mathbb{I}_r$ and if $f \in \mathcal{A}_s^m$, then $Y_{l_1}\cdots Y_{l_r}[f] \in \mathcal{A}_{s+r}^{m-d(L)}$ where the coefficients can be written as linear combinations of up to r-fold derivatives of the coefficients of f and up to $(r-1)$-fold derivatives of the structure constants. On the other hand, if $f \in \mathcal{A}_s^m$, it follows from part (1) of the Proposition that $|f(x)| \leq C\,\kappa^{-s}\,\delta^{-m}$, where C depends on the number of generators of $\mathcal{A}_s^m$ and on the supremum at x of the values of the coefficients of f. This proves statement (2).

Finally, Proposition 4.8 shows that $Y_{l_1}\,Y_{l_2}\cdots Y_{l_r}[\lambda_J] = f_{L,J}\,\lambda_I$ where $f_{L,J} \in \mathcal{A}_{n+r}^{d(I)-d(J)-d(K)}$. Thus part (2) implies part (3), and completes the proof. □

The next proposition gives a lower bound for $|\lambda_I(x)|$ on compact subsets when I is (x,δ,η)-dominant. Recall from Definition 1.4 that if $K \subset \Omega$ is compact, we put $\nu(K) = \inf_{y\in K}\,\max_{J\in\mathbb{I}_n}\,\left|\lambda_J(y)\right|$. The quantity $\nu(K)$ is a measure of the linear independence of the vector fields in $\mathcal{Y}$, and $\nu(K) > 0$ since the vectors $\{(Y_1)_y,\ldots,(Y_q)_y\}$ span T_y for every $y \in K$. In equation (3.3) we also put $N_0 = \max_{I,J\in\mathbb{I}_n}|d(I)-d(J)|$.

PROPOSITION 4.10. *Let $K \subset \Omega$ be compact. Let $x \in K$, let $0 < \delta \leq 1$, let $0 < \kappa \leq 1$, and suppose $I \in \mathbb{I}_n$ is (x,δ,κ)-dominant. Then*

$$\delta^{N_0} \leq \frac{1}{\kappa\,\nu(K)}\left|\lambda_I(x)\right|. \tag{4.17}$$

PROOF. Since I is (x,δ,κ)-dominant and $\delta \leq 1$, it follows that for all $L \in \mathbb{I}_n$ we have $\left|\lambda_I(x)\right| \geq \kappa\left|\lambda_L(x)\right|\delta^{d(L)-d(I)}$. In particular, if we choose L so that $|\lambda_L(x)| = \max_{J\in\mathbb{I}_n}|\lambda_J(x)|$ we have

$$\left|\lambda_I(x)\right| \geq \kappa\,\nu(K)\,\delta^{d(L)-d(I)} \geq \kappa\,\nu(K)\,\delta^{N_0},$$

which gives the desired result. □

4.3. Estimates on exponential balls.

In this section we use Taylor expansions in canonical coordinates and the estimates obtained in Section 4.2 to obtain estimates for determinants $\{\lambda_J\}$ and functions $f \in \mathcal{A}_s^m$ on an exponential ball $B_\eta^I(x,\delta)$. As usual, we fix a connected open set $\Omega \subset \mathbb{R}^n$ and a control system $\mathcal{Y} = \{Y_1,\ldots,Y_q;d_1,\ldots,d_q\}$ on Ω. We also fix compact sets $K \subset \widetilde{K} \subset \Omega$ and the corresponding constant η_0 as in Proposition 3.3. The following lemma shows that if I is (x,δ,κ)-dominant, then for η sufficiently small, the determinant λ_I is essentially constant on $B_\eta^I(x,\delta)$. Recall that $d = \min_{1\leq j\leq q} d_j$, $D = \max_{1\leq j\leq q} d_j$, $N_0 = \sup_{J,L\in\mathbb{I}_n}\left|d(J)-d(L)\right|$, and N_1 is the smallest positive integer such that $N_1 d \geq N_0 + D$.

LEMMA 4.11. *Let $0 < \kappa \leq 1$. There is a constant $0 < \eta_1 \leq \eta_0$ depending only on κ and $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$ so that if $x \in K$, if $0 < \delta \leq 1$, if $I \in \mathbb{I}_n$ is (x,δ,κ)-dominant, and if $0 < \eta \leq \eta_1$, then for $y \in B_\eta^I(x,\delta)$ we have*

$$\big|\lambda_I(y) - \lambda_I(x)\big| \leq \frac{1}{2}\,\big|\lambda_I(x)\big|,$$

and hence

$$\frac{1}{2}\big|\lambda(x)\big| < \big|\lambda(y)\big| < \frac{3}{2}\big|\lambda(x)\big|.$$

PROOF. For simplicity of notation, we again assume that $I = (1,2,\ldots,n)$. Write

$$F(u_1,\ldots,u_n) = \lambda_I\big(\Theta_{x,\delta,\eta}^I(u_1,\ldots,u_n)\big) = \lambda_I\Big(\exp\big(\eta\sum_{k=1}^n u_k\,\delta^{d_k}\,Y_k\big)(x)\Big).$$

The statement that $y \in B_\eta^I(x,\delta)$ means that $y = \Theta_{x,\delta,\eta}^I(u)$ for some $|u| < 1$. Then $F(0) = \lambda_I(x)$ and $\lambda_I(y) = F(u)$.

If $u \in \mathbb{B}(0;1) \subset \mathbb{R}^n$, Lemma 3.9 in Chapter 2 applied to the function F gives

$$\Big|\big(\lambda_I(y) - \lambda_I(x)\big) - \sum_{k=1}^{N_1-1}\frac{\eta^k}{k!}\big(u_1\,\delta^{d_1}\,Y_1 + \cdots + u_n\,\delta^{d_n}\,Y_n\big)^k[\lambda_I](x)\Big| \leq C_{N_1}\,\delta^{N_1 d}\,\eta^{N_1}\,|u|^{N_1}$$

where C_{N_1} depends on the supremum on $\widetilde{K}$ of N_1 applications of the vector fields $\{Y_1,\ldots,Y_n\}$ to the function λ_I. Thus C_{N_1} depends only on the quantity $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$. It follows that

$$\big|\lambda_I(y) - \lambda_I(x)\big| \leq \sum_{k=1}^{N_1-1}\frac{\eta^k}{k!}\Big|\big(u_1\,\delta^{d_1}\,Y_1 + \cdots + u_n\,\delta^{d_n}\,Y_n\big)^k[\lambda_I](x)\Big| + C_{N_1}\,\delta^{N_1 d}\,\eta^{N_1}\,|u|^{N_1}.$$

We first estimate the error term $C_{N_1}\,\delta^{N_1 d}\,\eta^{N_1}\,|u|^{N_1}$. Since $|u| \leq 1$ and $N_1 d \geq N_0$, it follows from Proposition 4.10 that we have

$$C_{N_1}\,\delta^{N_1 d}\,\eta^{N_1}|u|^{N_1} < C_{N_1}\,\delta^{N_0}\,\eta^{N_1} \leq \frac{C_{N_1}}{\kappa\,\nu(K)}\big|\lambda_I(x)\big|\eta^{N_1} < \frac{1}{4}\,\big|\lambda_I(x)\big|$$

provided that $\eta^{N_1} \leq \frac{1}{4}\,C_{N_1}^{-1}\,\kappa\,\nu(K)$.

Thus in order to complete the proof of the lemma, it suffices to show that for sufficiently small η and $|u| \leq 1$ we have

$$\sum_{k=1}^{N_1-1}\frac{\eta^k}{k!}\Big|\big(u_1\,\delta^{d_1}\,Y_1 + \cdots + u_n\,\delta^{d_n}\,Y_n\big)^k[\lambda_I](x)\Big| \leq \frac{1}{4}\,\big|\lambda_I(x)\big|. \tag{4.18}$$

Using (3) from Proposition 4.9, we can estimate a typical term on the left hand side of the last inequality when $|u| \leq 1$ and $\eta < 1$ by

$$\begin{aligned}
\frac{\eta^k}{k!}|u_{l_1}|\cdots|u_{l_k}|\,\delta^{d_{l_1}+\cdots+d_{l_k}}\,&\big|Y_{l_1}\cdots Y_{l_k}[\lambda_I](x)\big| \\
&\leq \eta\,\delta^{d_{l_1}+\cdots+d_{l_k}}\,\Big|Y_{l_1}\cdots Y_{l_k}[\lambda_I](x)\Big| \\
&\leq C_k\,\eta\,\kappa^{-n-k}\,\delta^{d_{l_1}+\cdots+d_{l_k}}\,\delta^{-d_{l_1}-\cdots-d_{l_k}}\,\big|\lambda_I(x)\big| \\
&\leq C_k\,\eta\,\kappa^{-n-N_1}\,\big|\lambda_I(x)\big|
\end{aligned}$$

where C_k depends only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{K,N_1}$. Since there are only a fixed finite number of such terms, we can choose η_1 sufficiently small, depending only on κ and $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$, so that equation (4.18) is satisfied for all $0 < \eta \leq \eta_1$. This completes the proof. □

It follows from Lemma 4.11 that $\lambda_I(y) \neq 0$ for $y \in B^I_\eta(x,\delta)$. This gives us the following.

COROLLARY 4.12. *Suppose that $0 < \kappa \leq 1$, that $x \in K$, that $0 < \delta \leq 1$, and that $I \in \mathbb{I}_n$ is (x,δ,κ)-dominant. If η_1 is chosen as in Lemma 4.11 and if $0 < \eta \leq \eta_1$, then $B^I_\eta(x,\delta) \subset \Omega_I$.*

For any $J \in \mathbb{I}_n$, part (3) of Proposition 4.9 gives an estimate for the value of the Y-derivatives of λ_J at the point x. Using the same kind of argument as in Lemma 4.11, we can obtain essentially the same estimates at any point $y = \Theta^I_{x,\delta,\eta}(u)$ if $u \in \mathbb{B}(0;1)$. We restrict our attention to λ_J itself. Recall that N_1 is the smallest positive integer such that $N_1 d \geq N_0 + nD$.

PROPOSITION 4.13. *Let $0 < \kappa \leq 1$, $x \in K$, $0 < \delta \leq 1$, and $0 < \eta \leq \eta_1$, where η_1 is given in Lemma 4.11. Suppose $I \in \mathbb{I}_n$ is (x,δ,κ)-dominant (with $I = (1,\ldots,n)$ for notational simplicity). Then for $J \in \mathbb{I}_n$ and all $y \in B^I_\eta(x,\delta)$ we have*

$$\big|\lambda_J(y)\big| \leq C\,\kappa^{-n-N_1}\,\delta^{d(I)-d(J)}\,\big|\lambda_I(x)\big|,$$

where C depends only on κ, $\nu(K)$, and $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$.

PROOF. Put

$$G(u_1,\ldots,u_n) = \lambda_J\Big(\exp\big(\eta\sum_{k=1}^{n} u_k\,\delta^{d_k}\,Y_k\big)(x)\Big).$$

Then Lemma 3.9 in Chapter 2 gives

$$\Big|G(u) - \sum_{k=0}^{N_1-1}\frac{\eta^k}{k!}\big(u_1\,\delta^{d_1}\,Y_1 + \cdots + u_n\,\delta^{d_n}\,Y_n\big)^k\big[\lambda_J\big](x)\Big| \leq C_{N_1}\,\delta^{N_1 d}\,\eta^{N_1}\,|u|^{N_1}$$

where C_{N_1} depends only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$. Note that $N_1 d \geq N_0 + d(I) - d(J)$. Thus for the the error term we have

$$C_{N_1}\delta^{N_1 d}\,\eta^{N_1}|u|^{N_1} \leq C_{N_1}\delta^{N_0}\,\delta^{d(I)-d(J)} \leq \Big[\frac{C_{N_1}}{\kappa\,\nu(K)}\Big]\,\delta^{d(I)-d(J)}\,\big|\lambda_I(x)\big|,$$

where we have used Proposition 4.10. It remains to estimate a finite number of terms of the form

$$\frac{\eta^k}{k!}|u_{m_1}|\cdots|u_{m_k}|\,\delta^{d_{m_1}+\cdots+d_{m_k}}\big|Y_{m_1}\cdots Y_{m_k}[\lambda_J](x)\big|.$$

But using part (3) of Proposition 4.9, if all $|u_j| \leq 1$ this expression is bounded by

$$C_k\,\kappa^{-n-k}\,\delta^{d(I)-d(J)}\,\big|\lambda_I(x)\big|$$

where again C_k depends only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$. This completes the proof. □

We can also obtain estimates for functions $f \in \mathcal{A}^m_s$. We will need the following simple version.

PROPOSITION 4.14. *Let $0 < \kappa \leq 1$, $x \in K$, $0 < \delta \leq 1$, and $0 < \eta \leq \eta_1$, where η_1 is given in Lemma 4.11. Suppose $I \in \mathbb{I}_n$ is (x, δ, κ)-dominant (with $I = (1, \dots, n)$ for notational simplicity). If $y \in B^I_\eta(x, \delta)$, then*

$$\left|a^m_k(y)\right| \leq C\, \kappa^{-n-N_1}\, \delta^{d_m - d_k},$$

where C depends only on $\nu(\widetilde{K})$ and $\left|\left|\, \mathcal{Y}\, \right|\right|_{\widetilde{K},N_1}$.

PROOF. According to Proposition 4.2, we have $a^m_k(y) = \lambda_M(y)\, \lambda_I(y)^{-1}$ where $M \in \mathbb{I}_n$ is obtained from $I = (1, \dots, n)$ by replacing m by k. The result now follows from Proposition 4.13 and Lemma 4.11. □

4.4. Some topological remarks.

Many of our arguments would be considerably simplified if we knew *a priori* that the mapping $\Theta^I_{x,\delta,\eta}$ were globally one-to-one on $\mathbb{B}(0; 1)$. We will eventually show that this is true for η sufficiently small, but at first we will only be able to show that $J\Theta^I_{x,\delta,\eta}(u) \neq 0$ for $|u| < 1$ and η sufficiently small. It will follow from the open mapping theorem that $\Theta^I_{x,\delta,\eta}$ is then locally one-to-one. In this section, we provide the version of the open mapping theorem and some additional topological results which we shall need.

LEMMA 4.15. *Let $U \subset \mathbb{R}^n$ be open, and let $\Theta : U \to \mathbb{R}^n$ be a $\mathcal{C}^2$ mapping. Suppose that for every $u \in U$, $J\Theta(u) \neq 0$. Let $E \subset U$ be a compact, connected, simply connected subset, and let $W = \Theta(E) \subset \mathbb{R}^n$. For $u \in U$, let $\mathbb{B}(u, \epsilon)$ denote the open Euclidean ball centered at u with radius ϵ.*

(1) There exists $\epsilon_1, \epsilon_2 > 0$ so that:

   (a) for every $u \in E$, $\mathbb{B}(u, \epsilon_1) \subset U$;

   (b) for every $u \in E$, the mapping Θ is globally one-to-one on $\mathbb{B}(u, \epsilon_1)$;

   (c) for every $u \in E$, $\mathbb{B}(\Theta(u), \epsilon_2) \subset \Theta\big(\mathbb{B}(u, \epsilon_1)\big)$.

(2) If $[a, b] \subset \mathbb{R}$ is an interval, if $\varphi : [a, b] \to W$ is continuous and one-to-one, and if $\Theta(u_0) = a$ then there exists a unique continuous mapping $\theta : [a, b] \to E$ such that $\theta(a) = u_0$ and $\varphi(t) = \Theta\big(\theta(t)\big)$ for all $t \in [a, b]$.

In order to finally prove that $\Theta^I_{x,\delta,\eta}$ is globally one-to-one, we will use the following fact.

LEMMA 4.16. *Let $U \subset \mathbb{R}^n$ be open, and let $\Theta : U \to \mathbb{R}^n$ be a $\mathcal{C}^2$ mapping. Suppose that for every $u \in U$, $J\Theta(u) \neq 0$. Suppose $E_1 \subset E_2 \subset U$ are compact, connected, simply connected subsets, and supposet $W_j = \Theta(E_j)$. Suppose $W_1 \subset S \subset W_2$ where S is simply connected. Then Θ is globally one-to-one on E_1.*

4.5. Proof of Theorem 3.7, part (1).

In this subsection we give the proof of part (1) of Theorem 3.7. We fix a connected open set $\Omega \subset \mathbb{R}^n$ and a control system $\mathcal{Y} = \{Y_1, \ldots, Y_q; d_1, \ldots, d_q\}$ on Ω. We also fix compact sets $K \subset \widetilde{K} \subset \Omega$ with K contained in the interior of $\widetilde{K}$, and the corresponding constant η_1 as in Lemma 4.11. Let $x \in K$, let $0 < \delta \leq 1$, and let $0 < \kappa \leq 1$. Choose $I \in \mathbb{I}_n$ which is (x, δ, η)-dominant. As usual, we assume $I = (1, \ldots, n)$. Let $0 < \eta \leq \eta_1$ where η_1 is given in Lemma 4.11.

We start by computing the Jacobian of the mapping

$$\Theta^I_{x,\delta,\eta}(u) = \exp\big(\eta \sum_{k=1}^n u_k\, \delta^{d_k}\, Y_k\big)(x).$$

To do this, we need to compute $d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_u\big]$. It is easy to check that at the origin, $d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_0\big] = \eta\, \delta^{d_j} (Y_j)_x$, and so $J\Theta^I_{x,\delta,\eta}(0) = \eta^n \delta^{d(I)} \big|\lambda_I(0)\big|$. However, it is not true in general that $d\Theta_{x,\delta}\big[\big(\partial_{u_j}\big)_u\big] = \eta\, \delta^{d_j}\, (Y_j)_{\Theta_{x,\delta}(u)}$.

According to Corollary 4.12, if $y \in B^I_\eta(x, \delta)$, then $\lambda_I(y) \neq 0$, and so the tangent vectors $\{(Y_1)_y, \ldots, (Y_n)_y\}$ span the tangent space T_y. Thus if $\eta \leq \eta_1$ we can write $d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_u\big]$ as a linear combination of these vectors. The following Lemma gives a more precise result for η sufficiently small. Recall that N_1 is the smallest positive integer such that $N_1 d \geq N_0 + nD$.

Lemma 4.17. *Let $0 < \kappa \leq 1$. There exists $\eta_2 \leq \eta_1$ depending only on κ, $\nu(K)$, and $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$ with the following property. Suppose that $x \in K$, that $0 < \delta \leq 1$, and that $I \in \mathbb{I}_n$ is (x, δ, κ)-dominant. If $0 < \eta \leq \eta_2$, $|u| \leq 1$ and $y = \Theta^I_{x,\delta,\eta}(u)$, we have*

$$d\Theta^I_{x,\delta,\eta}\big[(\partial_{u_j})_u\big] = \eta\, \delta^{d_j} \Big[(Y_j)_y + \sum_{k=1}^n b_{j,k}(y)\, (Y_k)_y\Big] \tag{4.19}$$

where

$$|b_{j,k}(y)| \leq \eta\, \delta^{d_k - d_j}. \tag{4.20}$$

Proof. Let us write $u \cdot_\delta Y = \sum_{k=1}^n u_k\, \delta^{d_k}\, (Y_k)_y$. Then let $W^{N_1}_y \in T_y$ be the vector

$$W^{N_1}_y = \eta\, \delta^{d_j}\, (Y_j)_y + \sum_{l=2}^{N_1-1} \alpha_k \Big[\overbrace{\big[\eta\, u \cdot_\delta Y,\, \big[\eta\, u \cdot_\delta Y,\, \cdots \big[\eta\, u \cdot_\delta Y,\, \eta\, \delta^{d_j}\, (Y_j)_y\big]\big]\big]}^{l\text{-fold iterate}}\Big] \tag{4.21}$$

where the rational numbers $\{\alpha_k\}$ are defined in Chapter 2, Lemma 5.2. This Lemma shows that when $|u| \leq 1$ we have

$$\Big|d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_u\big] - W^{N_1}_Y\Big| \leq C_{N_1}\, \eta^{N_1}\, \delta^{N_1 d} \tag{4.22}$$

where C_{N_1} depends only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$. Let us also write $E^{N_1}_Y = d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_u\big] - W^{N_1}_Y$ so that

$$d\Theta^I_{x,\delta,\eta}\big[\big(\partial_{u_j}\big)_u\big] = W^{N_1}_y + E^{N_1}_y. \tag{4.23}$$

We first deal with the 'error term' $E^{N_1}_y$. According to equation (4.22), this is a vector with length at most $C_{N_1}\, \eta^{N_1}\, \delta^{N_1 d}$. The tangent vectors $\{(Y_1)_y, \ldots, (Y_n)_y\}$ span T_y, so we can write $E^{N_1}_y$ as a linear combinations of these vectors. Cramer's rule shows that we can bound the resulting coefficients by a constant C times the length of $E_y^{N_1}$ divided by $|\lambda_I(y)|$, where C depends only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},1}$. Thus using Proposition 4.10 and Lemma 4.11, each coefficient is bounded by

$$C_{N_1}\eta^{N_1}\delta^{N_1 d}|\lambda_I(y)|^{-1} \leq 2\,C_N\eta^{N_1}\delta^{N_1 d}\,|\lambda_I(x)|^{-1} \leq \frac{2C_N}{\kappa\,\nu(K)}\,\eta^{N_1}\,\delta^{N_1 d-N_0} \leq \frac{1}{2}\eta^2\,\delta^D$$

provided that $4\,C_N\,\eta^{N-2} \leq \kappa\,\nu(K)$. Thus we can write $E_y^{N_1} = \eta\,\delta^{d_j}\,\sum_{k=1}^n b_k\,(Y_k)_y$ where $|b_k| \leq \frac{1}{2}\,\eta\,\delta^{d_k-d_j}$.

Next we deal with the terms in the sum defining $W_y^{N_1}$ in equation (4.21). A typical term in the sum has the form

$$\eta^l\,u_{i_1}\cdots u_{i_{l-1}}\delta^{d_{i_1}+\cdots+d_{i_{l-1}}+d_j}\,\,\big[Y_{i_1},\,\big[Y_{i_2},\,\cdots\big[Y_{i_{l-1}},\,Y_j\big]\big]\big] \tag{4.24}$$

where $l \geq 2$. According to Proposition 4.1 with $L = (i_1,\ldots,i_{l-1},j) \in \mathbb{I}_l$, we can write

$$\big[Y_{i_1},\,\big[Y_{i_2},\,\cdots\big[Y_{i_{l-1}},\,Y_j\big]\big]\big] = \sum_{k=1}^{q} c_L^k\,Y_k = \sum_{m=1}^{n}\Big[\sum_{k=1}^{q} c_L^k\,a_k^m\Big]\,Y_m$$

where $c_L^k(y) \equiv 0$ unless $d_k \leq d(L)$. Thus the term (4.24) can be written

$$\eta^l\,\sum_{m=1}^{n}\Big[\sum_{k=1}^{q} u_{i_1}\cdots u_{i_{l-1}}\delta^{d(L)}\,c_L^k\,a_k^m\Big]\,Y_m.$$

If $|u| \leq 1$, we can estimate the coefficient of Y_m by

$$\eta^l\sum_{k=1}^{q}\delta^{d(L)}\,|c_L^k(y)|\,|a_k^m(y)| \leq C_m\,\eta^2\,\sum_{k=1}^{q}\delta^{d(L)+d_m-d_k}\,|c_L^k(y)| \leq C_m'\,q\,\eta^2\delta^{d_m}$$

where C_m and C_m' depend only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$, and we have used Corollary 4.14 to estimate $|a_k^m(y)|$. If we choose η sufficiently small, the sum of the various coefficients of Y_m is bounded by $\frac{1}{2}\,\eta\,\delta^{d_m}$, and this completes the proof. □

We now establish part (1) of Theorem 3.7.

Lemma 4.18. *Let $0 < \kappa \leq 1$. There exists $0 < \eta_2 \leq 1$ depending only on κ, $\nu(K)$, and $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_1}$ with the following property. Let $x \in K$, let $0 < \delta \leq 1$, and let $I \in \mathbb{I}_n$ be (x,δ,κ)-dominant. If $0 < \eta \leq \eta_2$, then for $u \in \mathbb{B}(0;1)$ we have*

$$\frac{\eta^n}{4}\,\big|\lambda_I(x)\big|\,\delta^{d(I)} \leq J\Theta^I_{x,\delta,\eta}(u) \leq 4\,\eta^n\,\big|\lambda_I(x)\big|\,\delta^{d(I)}. \tag{3.9}$$

Proof. Using Lemma 4.17, we have

$$\begin{aligned}
J\Theta^I_{x,\delta,\eta}(u) &= \Big|\det\Big(d\Theta^I_{x,\delta,\eta}\big((\partial_{u_1})_u\big),\dots,d\Theta^I_{x,\delta,\eta}\big((\partial_{u_1})_u\big)\Big)\Big| \\
&= \Big|\det\Big(\eta\,\delta^{d_1}\Big[(Y_1)_y+\sum_{k=1}^n b_{1,k}(Y_k)_y\Big],\dots,\eta\,\delta^{d_n}\Big[(Y_n)_y+\sum_{k=1}^n b_{n,k}(Y_k)_y\Big]\Big)\Big| \\
&= \Big|\eta^n\,\delta^{d(I)}\Big[\det\big((Y_1)_y,\dots,(Y_n)_y\big)+ \\
&\qquad +\sum_{k_1=1}^n\cdots\sum_{k_n=1}^n(b_{1,k_1}\cdots b_{n,k_n})\det\big((Y_{k_1})_y,\dots,(Y_{k_n})_y\big)\Big]\Big| \\
&= \Big|\eta^n\,\delta^{d(I)}\Big[\det\big((Y_1)_y,\dots,(Y_n)_y\big)+ \\
&\qquad +\sum_{\sigma\in\mathfrak{S}_n}(b_{1,\sigma(1)}\cdots b_{n,\sigma(n)})\det\big((Y_{\sigma(1)})_y,\dots,(Y_{\sigma(n)})_y\big)\Big]\Big| \\
&= \Big|\eta^n\,\delta^{d(I)}\,\lambda_I(y)\Big[1+\sum_{\sigma\in\mathfrak{S}_n}(-1)^\sigma\,(b_{1,\sigma(1)}\cdots b_{n,\sigma(n)})\Big]\Big|
\end{aligned}$$

since any determinant with a repeated row is zero. However, using Lemma 4.17 again, we have

$$\big|b_{1,\sigma(1)}\cdots b_{n,\sigma(n)}\big|\le\eta^n\,\delta^{(d_{\sigma(1)}+\cdots+d_{\sigma(n)})-(d_1+\cdots+d_n)}=\eta^n$$

and so

$$\Big|\sum_{\sigma\in\mathfrak{S}_n}(-1)^\sigma\,(b_{1,\sigma(1)}\cdots b_{n,\sigma(n)})\Big|\le n!\,\eta^n\,. \tag{4.25}$$

Thus if $n!\,\eta^n<\frac{1}{2}$ it follows that

$$\frac{\eta^n}{2}\,\delta^{d(I)}\,\big|\lambda_I(y)\big|\le J\Theta^I_{x,\delta,\eta}(u)\le\frac{3\eta^n}{2}\,\delta^{d(I)}\,\big|\lambda_I(y)\big|.$$

Combining this with the estimate in Lemma 4.11 gives the estimate in equation (3.9) and completes the proof. □

Since the mapping $\Theta^I_{x,\delta,\eta_2}$ has non-vanishing Jacobian determinant on the unit ball, we can apply Lemma 4.15 to a dilate. If $\eta<\eta_2$ we have

$$\Theta^I_{x,\delta,\eta}(u)=\Theta^I_{x,\delta,\eta_2}\big(\eta\,\eta_2^{-1}\,u\big)$$

and thus $\Theta^I_{x,\delta,\eta}$ is just the mapping $\Theta^I_{x,\delta,\eta_2}$ restricted to the Euclidean ball of radius $\eta\,\eta_2^{-1}<1$. Thus applying Lemma 4.15, we have the following.

Corollary 4.19. *With the notation of Lemma 4.18, let $\eta<\eta_2$. There exist $\epsilon>0$ so that*

(1) *for every $u\in\mathbb{B}(0;1)$, the mapping $\Theta^I_{x,\delta,\eta}$ is globally one-to-one on the ball $\mathbb{B}(u;\epsilon)$;*

(2) *if $\varphi:[a,b]\to\overline{B^I_\eta(x,\delta)}$ is continuous and one-to-one with $\varphi(a)=x$, there exists a unique continuous mapping $\theta:[a,b]\to\overline{\mathbb{B}(0;1)}$ so that $\theta(a)=0$ and $\Theta^I_{x,\delta,\eta}\big(\theta(t)\big)=\varphi(t)$ for $a\le t\le b$.*

We continue our investigation of the images of the vector fields $\{\partial_{u_j}\}$ under the differential of the mapping $\Theta^I_{x,\delta,\eta}$ that are described in Lemma 4.17. Let $\eta < \eta_2$ and let $\epsilon > 0$ be as in Corollary 4.19. Let $u \in \mathbb{B}(0;1)$, let $y = \Theta^I_{a,\delta,\eta}(u)$, and let V_y be the diffeomorphic image under $\Theta^I_{x,\delta,\eta}$ of the ball $\mathbb{B}(u,\epsilon)$. Let $\{W_1,\ldots,W_n\}$ be the vector fields on V_y where $W_j = d\Theta^I_{x,\delta,\eta}\big[\partial_{u_j}\big]$. Lemma 4.17 shows that

$$(\eta\,\delta^{d_j})^{-1}\,W_j = Y_j + \sum_{k=1}^{n} b_{j,k}\,Y_k = \sum_{k=1}^{n}\big(\delta_{j,k}+b_{j,k}\big)\,Y_k \tag{4.26}$$

where $b_{j,k} \in \mathcal{E}(V_y)$ and $\sup_{z\in V_y}|b_{j,k}(z)| \le \eta\,\delta^{d_k-d_j}$. We want to solve this system for the vector fields $\{Y_k\}$ in terms of the vector fields $\{W_j\}$. After possibly shrinking η again, we can achieve the following.

Lemma 4.20. *Let $\varepsilon > 0$. There exists $\eta_3 \le \eta_2$ so that if $\eta < \eta_3$, we have*

$$\eta\,\delta^{d_k}Y_k = W_k + \sum_{j=1}^{n}\beta_{j,k}\,W_j$$

where $\beta_{j,k} \in \mathcal{E}(V_y)$ and

$$\sup_{z\in V_y}|\beta_{j,k}(z)| \le \varepsilon.$$

Proof. Let B be the $n\times n$ complex matrix

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n}\\ b_{2,1} & b_{2,2} & \cdots & b_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ b_{n,1} & b_{n,2} & \cdots & b_{n,n}\end{bmatrix},$$

and let

$$D = \begin{bmatrix} \delta^{d_1} & 0 & \cdots & 0\\ 0 & \delta^{d_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \delta^{d_n}\end{bmatrix} \qquad \text{so that} \qquad D^{-1} = \begin{bmatrix} \delta^{-d_1} & 0 & \cdots & 0\\ 0 & \delta^{-d_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \delta^{-d_n}\end{bmatrix}.$$

Then equation (4.26) can be written as

$$\eta^{-1}\,D^{-1}[W] = (I+B)[Y].$$

Hence if $(I+B)$ is invertible, we can solve for the $\{Y_k\}$ by writing

$$[Y] = (I+B)^{-1}(\eta\,D)^{-1}[W].$$

On the other hand, we have

$$D\,(I+B)\,D^{-1} = I + \begin{bmatrix} b_{1,1} & \delta^{d_1-d_2}\,b_{1,2} & \cdots & \delta^{d_1-d_2}\,b_{1,n}\\ \delta^{d_2-d_1}\,b_{2,1} & b_{2,2} & \cdots & \delta^{d_2-d_n}\,b_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ \delta^{d_n-d_1}\,b_{n,1} & \delta^{d_n-d_2}\,b_{n,2} & \cdots & b_{n,n}\end{bmatrix} = I + B_1,$$

and now the entries of B_1 are all bounded in absolute value by η. If η is sufficiently small, we can write

$$(I+B_1)^{-1} = I + \sum_{j=1}^{\infty}(-1)^j B_1^n = I + \widetilde{B}_1$$

where the entries of $\widetilde{B}_1$ are all uniformly bounded by 2η. Putting everything together, and taking $\eta < \frac{\varepsilon}{2}$, we have

$$\begin{aligned} \eta\, D[Y] &= (\eta\, D)(I+B)^{-1}\,(\eta\, D)^{-1}[W] \\ &= (\eta\, D)D^{-1}(I+B_1)^{-1}\,D\,(\eta\, D)^{-1}[W] \\ &= (I+\widetilde{B}_1)[W]. \end{aligned}$$

This is equivalent to the statement of the Lemma. □

We can interpret Lemma 4.20 in the following way. If $|u_0| \leq 1$, the mapping $\Theta^I_{x,\delta,\eta}$ is a diffeomorphism of the open ball $\mathbb{B}(u_0,\epsilon)$ to its image, which is a neighborhood V_{y_0} of $y_0 = \Theta^I_{x_0,\delta,\eta}(u)$. There is an inverse mapping

$$\Psi = (\psi_1, \ldots, \psi_n) : V_y \to \mathbb{B}(u_0,\epsilon),$$

so that if $u = (u_1, \ldots, u_n) \in \mathbb{B}(u_0,\epsilon)$ we have

$$u_k = \psi_k\big(\Theta^I_{x,\delta,\eta}(u)\big).$$

Thus we can regard the functions $\{\psi_1, \ldots, \psi_n\}$ as coordinates near y. Now if $u \in \mathbb{B}(u_0,\epsilon)$ and $y = \Theta^I_{x,\delta,\eta}(u)$ then

$$\begin{aligned} W_j[\psi_k](y) &= d\Theta^I_{x,\delta,\eta}[\partial_{v_j}][\psi_k](y) \\ &= \partial_{u_j}\big[\psi_k \circ \Theta^I_{x,\delta,\eta}\big](u) \\ &= \partial_{u_j}\big[u_k\big](u) \\ &= \delta_{j,k}. \end{aligned}$$

Thus $\eta\,\delta^{d_j}\,Y_j[\psi_k](y) = \delta_{j,k} + \beta_{k,j}(y)$. This gives the following.

Corollary 4.21. *With $\eta \leq \eta_3$, if $\Psi = (\psi_1, \ldots, \psi_n)$ is the inverse to the mapping $\Theta^I_{x,\delta,\eta}$ on the neighborhood V_{y_0}, then for $y \in V_{y_0}$*

$$\big|Y_j[\psi_k](y)\big| \leq \begin{cases} (1+\varepsilon)\,\eta^{-1}\,\delta^{-d_j} & \text{if } j = k \\ \varepsilon\,\eta^{-1}\,\delta^{-d_j} & \text{if } j \neq k. \end{cases} \tag{4.27}$$

4.6. Proof of Theorem 3.7, part (2).

We now can prove part (2) of Theroem 3.7.

Lemma 4.22. *Let $0 < \kappa \leq 1$, and let $0 < \eta \leq \eta_3$ where η_3 is given in Lemma 4.20. There is a constant C, depending only on $\big|\big|\,\mathcal{Y}\,\big|\big|_{\widetilde{K},N_3}$, and constant $\alpha > 0$ depending only on n, N_0, and d, so that if*

$$\tau < C\,\kappa^\alpha,$$

then if $x \in K$, $0 < \delta \leq 1$, $I \in \mathbb{I}_n$ is (x,δ,κ)-dominant, and $0 < \sigma \leq 1$, we have

$$B_\rho(x, \tau\,\eta^{1/d}\sigma^{\frac{D}{d}}\,\delta) \subset B^I_\eta(x, \sigma\delta).$$

Proof. Let $\tau > 0$ and let $y \in B_\rho(x, \tau\,\eta^{\frac{1}{d}}\,\sigma^{\frac{D}{d}}\,\delta)$. Then there is an absolutely continuous mapping $\varphi = \varphi : [0,1] \to \Omega$ with $\varphi(0) = x$ and $\varphi(1) = y$ such that for almost all $t \in [0,1]$ we can write

$$\varphi_y'(t) = \sum_{j=1}^{q} b_j(t)(Y_j)_{\varphi_y(t)} \tag{4.28}$$

with

$$\sup_{1\leq j\leq q} |b_j(t)| \leq (\tau\,\eta^{\frac{1}{d}}\,\sigma^{\frac{D}{d}}\,\delta)^{d_j} \leq \eta\,(\tau\,\sigma^{\frac{D}{d}}\,\delta)^{d_j}. \tag{4.29}$$

Without loss of generality, we can assume that the mapping φ is one-to-one.

Recall that

$$B_\eta^I(x,\sigma\delta) = \Big\{ y \in \Omega \,\Big|\, \big(\exists u \in \mathbb{B}(0;1)\big)\big(y = \Theta_{x,\delta,\eta}^I(D_\sigma[u])\big)\Big\}.$$

Let

$$E = \Big\{ t \in [0,1] \,\Big|\, \varphi(s) \in B_\eta^I(x,\sigma\delta) \text{ for all } 0 \leq s \leq t \Big\}.$$

Since $B_\eta^I(x,\sigma\delta)$ is open and φ is continuous, the set E is relatively open in $[0,1]$. Our object is to show that if τ is sufficiently small, then $1 \in E$, and hence $y = \varphi(1) \in B_\eta^I(x,\sigma\delta)$. This would imply $B_\rho(x,\tau\,\sigma^{\frac{D}{d}}\,\delta) \subset B_\eta^I(x,\sigma\delta)$.

Now if $1 \notin E$, it follows that $E = [0,a)$ with $a \leq 1$. Moreover, $\varphi(a)$ belongs to the boundary of $B_\eta^I(x,\sigma\delta)$, and so $\varphi(a) = \Theta_{x,\delta,\eta}^I(D_\sigma[u])$ with $|u| = 1$. According to part (2) of Corollary 4.19, we can lift the mapping φ to a mapping $\theta : [0,a] \to \overline{\mathbb{B}(0;1)}$ so that $\theta(0) = 0$, $\big|\theta(s)\big| < 1$ for $0 \leq s < a$, $\big|\theta(a)\big| = 1$, and for $0 \leq s \leq a$,

$$\varphi(s) = \exp\big(\eta \sum_{k=1}^{n} \sigma^{d_k}\theta_k(s)\,(\delta)^{d_k}\,Y_k\big)(x) = \Theta_{x,\delta,\eta}^I\big(D_\sigma[\theta(s)]\big). \tag{4.30}$$

Since the mapping $\Theta_{x,\delta,\eta}^I$ is locally one-to-one and the mapping θ and φ are one-to-one, it follows that $\Theta_{x,\delta,\eta}^I$ is actually globally one-to-one on some small open neighborhood of the image $\theta\big([0,a]\big)$. If we write the inverse mapping as $\Psi = (\psi_1,\ldots,\psi_n)$, then these components are well-defined functions in some neighborhood of $\varphi\big([0,a]\big)$. Thus we can write

$$D_\sigma[\theta(s)] = \big(\psi_1\big(\varphi(s)\big),\ldots,\psi_n\big(\varphi(s)\big)\,\big).$$

Since $\theta(0) = 0$, we can use the fundamental theorem of calculus, equation (4.28), and equation (4.6) to calculate:

$$\begin{aligned}
\sigma^{d_k}\theta_k(a) &= \sigma^{d_k}\theta_k(a) - \sigma^{d_k}\theta_k(0)\\
&= \psi_k\big(\varphi(a)\big) - \psi_k\big(\varphi(0)\big)\\
&= \int_0^a \frac{d}{ds}\big[\psi_k\big(\varphi(s)\big)\big]\,ds\\
&= \int_0^a \sum_{l=1}^{q} b_l(s)\,Y_l[\psi_k]\big(\varphi(s)\big)\,ds\\
&= \sum_{l=1}^{q}\sum_{j=1}^{n} \int_0^a b_l(s)\,a_l^j\big(\varphi(s)\big)\,Y_j[\psi_k]\big(\varphi(s)\big)\,ds
\end{aligned}$$

But for $0 \leq s \leq a$ we have

$$\begin{aligned}\big|b_l(s)\, a_l^j\big(\varphi(s)\big)\, Y_j[\psi_k]\big(\varphi(s)\big)\big| &\leq \eta\, (\tau\, \sigma^{\frac{D}{d}}\, \delta)^{d_l}\, C\kappa^{-n-N_0}\delta^{d_j-d_l}\, \eta^{-1}\, \delta^{-d_j} \\ &\leq C\, \kappa^{-n-N_0}\tau^{d_l}(\sigma^{\frac{D}{d}})^{d_l} \\ &\leq C\, \kappa^{-n-N_0}\, \tau^d\, \sigma^{d_k} \\ &< \sigma^{d_k}\end{aligned}$$

since $Dd_l \geq dd_k$, provided that $C\, \kappa^{-n-N_0}\, \tau^d < 1$. Here C depends only on $\big|\big|\, \mathcal{Y}\, \big|\big|_{\widetilde{K},N_3}$. This is a contradiction, and completes the proof. □

4.7. Proof of Theorem 3.7, part (3).

We can now show that for η sufficiently small, the mapping $\Theta^I_{x,\delta,\eta}$ is globally one-to-one. The key is the following comparison of exponential balls.

Lemma 4.23. *Suppose that $I, J \in \mathbb{I}_n$, that $x \in K$, that $0 < \delta \leq 1$, and that both I and J are (x, δ, κ)-dominant. Then there exist constants $0 < \tau_2 < \tau_1 < 1$ depending only on κ, η, and $\big|\big|\, \mathcal{Y}\, \big|\big|_{\widetilde{K},N_1}$ so that*

$$B^J_\eta(x, \tau_2\delta) \subset B^I_\eta(x, \tau_1\delta) \subset B^J_\eta(x, \delta).$$

Proof. Since I is (x, δ, κ)-dominant, it follows from Lemma 4.22 that if $\tau_1 = C\, \kappa^\alpha \eta^{\frac{1}{d}}$, we have $B_\rho(x, \tau_1\delta) \subset B^I_\eta(d, \delta)$. Here C depends only on $\big|\big|\, \mathcal{Y}\, \big|\big|_{\widetilde{K},N_3}$. On the other hand, Proposition 3.5 shows that $B^J(x, \tau_1\delta) \subset B_\rho(x, \tau_1\delta)$, and so

$$B^J(x, \tau_1\delta) \subset B^I_\eta(x, \delta). \tag{A}$$

Next, since J is (x, δ, κ)-dominant, we have

$$\begin{aligned}|\lambda_J(x)|\, (\tau_1\delta)^{d(J)} &\geq \kappa\, \tau_1^{d(J)} \sup_{L\in\mathbb{I}_n} |\lambda_L(x)|\, \delta^{d(L)} \\ &= \kappa\, \tau_1^{d(J)-d(L)} \sup_{L\in\mathbb{I}_n} |\lambda_L(x)|\, (\tau_1\delta)^{d(L)} \\ &\geq \kappa\, \tau_1^{N_0}\, |\lambda_L(x)|\, (\tau_1\delta)^{d(L)}.\end{aligned}$$

Thus J is also $\big(x, (\tau_1\delta), (\kappa\tau_1^{N_0})\big)$-dominant. Using Lemma 4.22 again, if $\tau_2 = C\, (\kappa\tau_1^{N_0})^\alpha$, then we have $B_\rho(x, \tau_2\delta) \subset B^J_\eta(x, \tau_1\delta)$, where C again depends only on $\big|\big|\, \mathcal{Y}\, \big|\big|_{\widetilde{K},N_3}$. But then by Proposition 3.5, we have $B^I_\eta(x, \tau_2\delta) \subset B_\rho(x, \tau_2\delta)$. This gives us

$$B^I_\eta(x, \tau_2\delta) \subset B^J_\eta(x, \tau_1\delta). \tag{B}$$

Combining (A) and (B) gives the desired result. □

Lemma 4.24. *There exists $\eta > 0$ and $\kappa > 0$ so that if $x \in K$ and $0 < \delta \leq 1$ there exists $I \in \mathbb{I}_n$ which is (x, δ, κ)-dominant and $\Theta^I_{x,\delta,\eta} : \mathbb{B}(0; 1) \to B^I_\eta(x, \delta)$ is globally one-to-one.*

Proof. Let $x_0 \in K$. Choose $I_0 \in \mathbb{I}_n$ such that

$$\begin{aligned}d(I_0) &= \min\, \big\{d(J)\, \big|\, J \in \mathbb{I}_n,\ \lambda_J(x_0) \neq 0\big\}, \\ \big|\lambda_{I_0}(x_0)\big| &= \max_{d(J)=d(I_0)} \big|\lambda_J(x_0)\big|.\end{aligned}$$

Then there exists $0 < \kappa_0 \leq \frac{1}{2}$ depending on the point x_0 so that

$$\big|\lambda_{I_0}(x_0)\big| \geq 2\kappa_0 \sup_{J\in\mathbb{I}_n} \big|\lambda_J(x_0)\big|,$$

and hence, since $d(I_0)$ is minimal among indices J with $\lambda_J(x_0) \neq 0$, it follows that

$$\big|\lambda_{I_0}(x_0)\big|\,\delta^{d(I_0)} \geq 2\kappa_0 \sup_{J\in\mathbb{I}_n} \big|\lambda_J(x_0)\big|\,\delta^{d(J)}$$

for all $0 \leq \delta \leq 1$. Thus I_0 is $(x_0, \delta, 2\kappa_0)$-dominant, and hence also (x_0, δ, κ_0)-dominant for all $0 < \delta \leq 1$.

Choose an open neighborhood $\widetilde{W}$ of x_0 contained in $\widetilde{K}$ such that for $x \in \widetilde{W}$ we have

$$\big|\lambda_{I_0}(x)\big| \geq \kappa_0 \sup_{J\in\mathbb{I}_n} \big|\lambda_J(x)\big|.$$

Thus I_0 is $(x, 1, \kappa_0)$-dominant for all $x \in \widetilde{W}$.

Let η_3 be the constant from Lemma 4.20. Define a mapping $\Theta : \widetilde{W} \times \mathbb{B}(0;1) \to \widetilde{W} \times \Omega$ by setting

$$\Theta(x, u) = \big(x, \Theta^{I_0}_{x,1,\eta_3}(u)\big).$$

Then the Jacobian determinant of the mapping $\Theta^{I}_{x,1,\eta_3}$ at the point $(x_0, 0)$ is $J\Theta(x_0, 0) = (\eta_3)^n \big|\lambda_{I_0}(x_0)\big| \neq 0$, and by the open mapping theorem, it follows that there is an open neighborhood $x_0 \in W \subset \widetilde{W}$ and a constant $\epsilon > 0$ so that the mapping Θ is globally one-to-one on the set $W \times \mathbb{B}(0, \epsilon)$. If we let $\eta_4 = \epsilon\,\eta_3$, it follows that the mapping $\widetilde{\Theta}(x, u) = \big(x, \Theta^{I_0}_{x,1,\eta_4}(u)\big)$ is globally one-to-one on $W \times \mathbb{B}(0;1)$.

Thus we are in the following situation. For $x_0 \in K$, we have found $I_0 \in \mathbb{I}_n$, constants $0 < \kappa_0 \leq \frac{1}{2}$ and $0 < \eta_4 < 1$, and a neighborhood W of x_0 in Ω so that

- I_0 is (x_0, δ, κ_0)-dominant for all $0 < \delta \leq 1$.
- I_0 is $(x, 1, \kappa_0)$-dominant for $x \in W$.
- The mapping $\Theta^{I_0}_{x,\delta,\eta} : \mathbb{B}(0;1) \to B^I_\eta(x, \delta)$ is globally one-to-one for all $0 < \eta \leq \eta_4$, all $0 < \delta \leq 1$, and all $x \in W$..

However, it is important to note that it is *not* necessarily true that I_0 is (x, δ, κ_0)-dominant if $x \neq x_0$ and $\delta < 1$. As soon as we move away from x_0, there may be n-tuples $J \in \mathbb{I}_n$ with $d(J) < d(I_0)$ such that $\lambda_J(x_0) = 0$ but $\lambda_J(x) \neq 0$. It then may happen that for δ small we have $\kappa_0 \big|\lambda_J(x)\big|\delta^{d(J)} >> \big|\lambda_{I_0}(x)\big|\,\delta^{d(I_0)}$.

We can, however, proceed as follows. For each $x \in W$ we can choose a sequence of n-tuples $I_0, I_1, \ldots, I_M \in \mathbb{I}_n$ and a sequence of positive numbers $1 = \delta_0 > \delta_1 > \cdots > \delta_M > 0$ so that

$$\big|\lambda_{I_j}(x)\big|\,\delta^{d(I_j)} \geq \frac{\kappa_0}{2} \sup_{J\in\mathbb{I}_n} \big|\lambda_J(x)\big|\,\delta^{d(J)} \quad \text{for } 0 \leq j \leq M-1 \text{ and } \delta_{j+1} \leq \delta \leq \delta_j$$

$$\big|\lambda_{I_M}(x)\big|\,\delta^{d(I_M)} \geq \frac{\kappa_0}{2} \sup_{J\in\mathbb{I}_n} \big|\lambda_J(x)\big|\,\delta^{d(J)} \quad \text{for } 0 \leq \delta \leq \delta_M$$

The number M can depend on x. However, we can assume that $d(I_{j+1}) < d(I_j)$ for $0 \leq j \leq M-1$. In particular, each chosen n-tuple occurs only once, and thus M is at most the total number of elements of $\mathbb{I}_n$, and hence is bounded independently of x.

In particular, we can apply Lemma 4.23 since I_0 and I_1 are both $(x, \delta_1, \frac{1}{2}\kappa_2)$-dominant. It follows that there are constants $0 < \tau_1' < \tau_1 < 1$ so that

$$B_\eta^{I_1}(x, \tau_1'\delta_1) \subset B_\eta^{I_0}(x, \tau_1\delta_1) \subset B_\eta^{I_1}(x, \delta_1).$$

Since the mapping $\Theta^{I_0}_{x,1,\eta}$ is globally one-to-one, the set $B_\eta^{I_0}(x, \tau_1\delta_1)$ is simply connected. It follows from Lemma 4.16 that the mapping $\Theta^{I_1}_{x,\delta_1,\eta}$ is globally one-to-one on the set $\big\{u \in \mathbb{B}(0;1) \,\Big|\, |u| < (\tau_1')^D\big\}$. In particular, $B_\eta^{I_1}(x, \tau_1'\delta_1)$ is simply connected.

We repeat the argument. Note that I_1 and I_2 are both $(x, \delta_2, \frac{1}{2}\kappa_0)$-dominant. Since $\tau_1' < 1$, we have for $j = 1, 2$

$$\begin{aligned} |\lambda_{I_j}(x)|(\tau_1'\delta_2)^{d(I_j} &\geq {\tau_1'}^{d(I_j)-d(L)}\frac{\kappa_0}{2} \sup_{L\in\mathbb{I}_n} |\lambda(x)|\,(\tau_1'\delta_2)^{d(L)} \\ &\geq \frac{1}{2}\kappa_0{\tau_1'}^{N_0} \sup_{L\in\mathbb{I}_n} |\lambda(x)|\,(\tau_1'\delta_2)^{d(L)} \end{aligned}$$

Thus both I_1 and I_2 are $\big(x, (\tau_1'\delta_2), \frac{1}{2}\kappa_0(\tau_1')^{N_0}\big)$-dominant.

It now follows from Lemma 4.23 that there are constants $0 < \tau_2' < \tau_2$ so that

$$B_\eta^{I_2}(x, \tau_2'\tau_1'\delta_2) \subset B_\eta^{I_1}(x, \tau_2\tau_1'\delta_2) \subset B_\eta^{I_2}(x, \tau_1'\delta_2).$$

Since $\tau_2 \leq 1$ it follows that $B_\eta^{I_1}(x, \tau_2\tau_1'\delta_2) \subset B_\eta^{I_1}(x, \tau_1'\delta_1)$, which we know is the image of a globally one-to-one mapping. Then $B_\eta^{I_1}(x, \tau_2\tau_1'\delta_2)$ is simply connected, and another application of Lemma 4.16 shows that $\Theta^{I_2}_{x,\delta_2,\eta}$ is globally one-to-one on the set $\big\{u \in \mathbb{B}(0;1) \,\Big|\, |u| < (\tau_2'\tau_1')^D\big\}$.

We can now repeat this process M times. We find a sequence $1 \geq \epsilon_1 \geq \epsilon_2 \cdots \geq \epsilon_M > 0$ so that $\Theta^{I_j}_{x,\delta_j,\eta}$ is globally one to one on the set $\big\{u \in \mathbb{B}(0;1) \,\Big|\, |u| < \epsilon_j\big\}$. Then if we put $\eta_5 = \epsilon_M$ we have shown that $\theta^{I_j}_{x,\delta_j,\eta_5}$ is globally one-to-one on the unit ball $\mathbb{B}(0;1)$.

Let $\eta = \eta_3$. The mapping $\Theta^{I_0}_{x,1,\eta}$ is globally one-to-one, and so $B_\eta^{I_0}(x, \epsilon_1\delta_1)$ is simply connected. If $\Theta^{I_1}_{x,\epsilon_2\delta_1,\eta}$ were not globally one-to-one, there would be a line segment in $\mathbb{B}(0;1)$ which $\Theta^{I_1}_{x,\epsilon_2\delta_1,\eta}$ maps to a closed curve in $B_\eta^{I_1}(x, \epsilon_2\delta_1)$. But this curve can be deformed to a point in $B_\eta^{I_0}(x, \epsilon\delta_1)$, and hence it can be deformed to a point in $B_\eta^{I_1}(x, \delta_1)$, which is impossible. Thus $\Theta^{I_1}_{x,\epsilon_2\,\delta_1,\eta}$ is globally one-to-one.

We now repeat this argument N times, and conclude that all the mappings are globally one-to-one. This completes the proof, since E is compact. □

4.8. Proof of Theorem 3.7, part (4).

Let $\{Z_j\}$ be the vector fields on $\mathbb{B}(0;1)$ such that $d\Theta^I_{x,\delta,\eta}[Z_j] = \eta\,\delta^{d_j}\,Y_j$ on $B^I_\eta(x,\delta)$. We want to show that $\mathcal{Z} = \big\{Z_1, \ldots, Z_q; d_1, \ldots, d_q\big\}$ is a control system on $\mathbb{B}(0;1)$ and that $\big|\big|\,\mathcal{Z}\,\big|\big|_{\mathbb{B}(0,;1),N}$ and $\nu(\mathcal{Z})$ is bounded independently of x, δ, and η. According to Lemma 4.17, for $1 \leq j \leq n$ we have

$$d\Theta^I_{x,\delta,\eta}[\partial_{u_j}] = d\Theta^I_{x,\delta,\eta}\Big[Z_j + \sum_{k=1}^{n} b_{j,k}\,Z_k\Big],$$

and since $d\Theta^I_{x,\delta,\eta}$ is one-to-one, it follows that

$$\partial_{u_j} = Z_j + \sum_{k=1}^{n} b_{j,k}\, Z_k,$$

where $\{|b_{j,k}|\}$ are uniformly bounded Thus arguing as in Lemma 4.20, we can solve for the vector fields $\{Z_k\}$ and write

$$\eta\, Z_k = \partial_{u_k} + \sum_{j=1}^{n} \beta_{j,k}\, \partial_{u_j}$$

for $1 \leq k \leq n$, where $\{|\beta_{j,k}|\}$ are uniformly bounded. For other indices, recall that for $n+1 \leq l \leq q$ we have

$$Y_l = \sum_{m=1}^{n} a_l^m\, Y_m,$$

and hence

$$\eta\, \delta^{d_l}\, Y_l = \sum_{m=1}^{n} \delta^{d_l - d_m}\, a_l^m\, (\eta\, \delta^{d_m}\, Y_m).$$

But this means that

$$Z_l = \sum_{m=1}^{n} \delta^{d_l - d_m}\, a_l^m\, Z_m,$$

and $\{|\delta^{d_l - d_m}\, a_l^m|\}$ are uniformly bounded.

5. Smooth Metrics

We now use the information about the local structure of Carnot-Carathéodory balls to construct a smooth version of the metric ρ.

5.1. Differentiable metrics.

The key is the construction of local bump functions which behave correctly under differentiation.

Proposition 5.1. *Let $u = (u_1, \dots, u_n) \in \mathbb{B}(0;1)$ and suppose $|u| < \eta < 1$. Then there exists $v \in \mathbb{B}(0;1)$ with $|v| < 1$ and $u_j = \eta^{\frac{d_j}{D}}\, v_j$ for $1 \leq j \leq n$.*

Proof. Let $w = \eta^{-1} u$, so that $|w| < 1$ and $u = \eta\, w$. Write

$$\eta = \big(\eta^{\frac{1}{D}}\big)^D = \big(\eta^{\frac{1}{D}}\big)^{d_j}\, \big(\eta^{\frac{D-d_j}{D}}\big)$$

Hence

$$\begin{aligned} u &= (\eta\, w_1, \dots, \eta\, w_n) \\ &= \big(\eta^{\frac{d_1}{D}}\, \eta^{\frac{D-d_j}{D}}\, w_1, \dots, (\eta^{\frac{d_n}{D}}\, \eta^{\frac{D-d_n}{D}}\, w_n\big) \\ &= \big(\eta^{\frac{d_1}{D}}\, v_1, \dots, \eta^{\frac{d_n}{D}}\, v_n\big) \end{aligned}$$

where $v_j = \eta^{\frac{D-d_j}{D}}\, w_j$.

□

COROLLARY 5.2. *If $u \in \mathbb{B}(0;1)$, with $|u| < \lambda \leq 1$, then*

$$\Theta^I_{x,\delta,\eta}(u) \in B^I_\eta(x, \lambda^{\frac{1}{D}}\,\delta) \subset B_\rho\big(x, \lambda^{\frac{1}{D}}\,\delta\big).$$

PROOF. Let $|u| < \lambda < 1$. Then we have

$$\begin{aligned}
\Theta^I_{x,\delta,\eta}(u) &= \exp\Big(\eta \sum_{k=1}^{n} u_j\,\delta^{d_j} Y_j\Big)(x) \\
&= \exp\Big(\eta \sum_{k=1}^{n} v_j\,\big(\lambda^{\frac{1}{D}}\delta\big)^{d_j} Y_j\Big)(x) \\
&= \Theta^I_{x,\lambda^{\frac{1}{D}}\delta,\eta}(v) \in B^I_\eta(x, \lambda^{\frac{1}{D}}\delta) \subset B_\rho\big(x, \lambda^{\frac{1}{D}}\delta\big)
\end{aligned}$$

□

LEMMA 5.3. *Let $K \subset \widetilde{K} \subset \Omega$ be a compact sets with K contained in the interior of $\widetilde{K}$. Let η, τ be the corresponding constants from Theorem 3.7. Then for $x \in K$ and $0 < \delta \leq 1$ there exists $\varphi = \varphi_{x,\delta} \in \mathcal{E}(\Omega)$ with the following properties:*

(1) $0 \leq \varphi_{x,\delta}(y) \leq 1$ *for all* $y \in \Omega$.

(2) We have

$$\varphi_{x,\delta}(y) = \begin{cases} 0 & \text{if } \rho(x,y) \geq \delta, \\ 1 & \text{if } \rho(x,y) \leq \tau\eta^{\frac{1}{d}}\tau^{\frac{D}{d}}\,\delta. \end{cases}$$

(3) For every $J = (j_1, \ldots, j_r) \in \mathbb{I}_r$ there is a constant C_J, independent of x and δ, so that

$$\sup_{y\in\Omega} \big|Y_{j_r} Y_{j_{r-1}} \cdots Y_{j_2} Y_{j_1}[\varphi_{x,\delta}](y)\big| \leq C_J\,\delta^{-d(J)}.$$

PROOF. Given $x \in K$ and $0 < \delta \leq 1$, let $I \in \mathbb{I}_n$, η, and τ be as in the conclusions of Theorem 3.7. The mapping $\Theta^I_{x,\delta,\eta} : \mathbb{B}(0;1) \to B^I_\eta(x,\delta)$ is a diffeomorphism. Let us write $\Phi : B^I_\eta(x,\delta) \to \mathbb{B}(0;1)$ be the inverse mapping. Choose a function $\phi \in \mathcal{C}^\infty_0\big(\mathbb{B}(0;1)\big)$ so that $0 \leq \phi(u) \leq 1$ for all u, $\phi(u) \equiv 1$ for $|u| \leq \tau^D$, and $\phi(u) \equiv 0$ for $|u| \geq \frac{1}{2}$. Put $\varphi_{x,\delta}(u) = \phi\big(\Phi(y)\big)$. Then clearly $\varphi_{x,\delta} \in \mathcal{C}^\infty_0(B^I_\eta(x,\delta))$ and $0 \leq \varphi(y) \leq 1$ for all $y \in \Omega$.

Since $B^I_\eta(x,\delta) \subset B_\rho(x,\delta)$ and the support of $\varphi_{x,\delta}$ is contained in $B^I_\eta(x,\delta)$, it follows that $\varphi_{x,\delta}(y) = 0$ if $\rho(x,y) \geq \delta$.

If $y \in B^I_\eta(x,\tau\delta)$, then there exists $u \in \mathbb{B}(0;1)$ so that $y = \Theta^I_{x,\tau\delta,\eta}(u)$. This means that

$$y = \exp\Big(\eta \sum_{k=1}^{n} u_k \tau^{d_k}\delta^{d_k} Y_k\Big)(x) = \exp\Big(\eta \sum_{k=1}^{n} v_k \delta^{d_k} Y_k\Big)(x) = \Theta^I_{x,\delta,\eta}(v)$$

where $v_k = \tau^{d_k}\,u_k$. Since $\tau \leq 1$ it follows that $|v| \leq \tau^d |u| < \tau^d$. Thus $B^I_\eta(x,\tau\delta)$ is contained in the image under $\Theta^I_{x,\delta,\eta}$ of $\mathbb{B}(0;\tau^d)$, which is where we know $\phi \equiv 1$. Thus $\varphi_{x\,delta} \equiv 1$ on $B^I_\eta(x,\tau\delta)$. But it follows from part 2 of Theorem 3.7 that $B_\rho(x, \tau\eta^{\frac{1}{d}}\tau^{\frac{D}{d}}\,\delta) \subset B^I_\eta(x,\tau\delta)$. Thus if $\rho(x,y) \leq \tau\eta^{\frac{1}{d}}\tau^{\frac{D}{d}}\,\delta$, it follows that $\varphi_{x,\delta}(y) = 1$. This completes part (2), and condition (3) follows from the last part of Theorem 3.7. This completes the proof. □

Proposition 5.4. *Let $K \subset \Omega$ be compact, and let $C > 1$. There is a positive integer M so that if $0 < \delta \leq 1$ and if $\{B_\rho(x_j, \delta)\}$ is a disjoint collection of balls with centers $x_j \in K$, then every point of K is contained in at most M of the dilated balls $\{B_\rho(x_j, C\delta)\}$.*

Proof. Let $y \in K$ and suppose that $y \in B_\rho(x_j, C\delta)$ for $1 \leq j \leq M$. Then for these indices, $B_\rho(x_j, \delta) \subset B_\rho(x_j, C\delta) \subset B_\rho(y, 3C\delta)$, and $B_\rho(y, 3C\delta) \subset B_\rho(x_j, 4C\delta)$. Since the balls $\{B_\rho(x_j, \delta)\}$ are disjoint, it follows from the doubling property that

$$\begin{aligned}\big|B(y, 3C\delta)\big| &\geq \sum_{j=1}^{M} \big|B_\rho(x_j, \delta)\big| \\ &\geq A \sum_{j=1}^{M} \big|B_\rho(x_j, 4C\delta)\big| \\ &\geq AM \big|B(y, 3C\delta)\big|\end{aligned}$$

Thus $M \leq A^{-1}$. □

Lemma 5.5. *Let $K \subset \Omega$ be compact and let $0 < \delta \leq 1$. Then there exists $\omega = \omega_\delta \in \mathcal{C}^\infty(\Omega \times \Omega)$ with the following properties:*

(1) For all $x, y \in \Omega$ we have $0 \leq \omega_\delta(x, y) \leq M$.

(2) For all $x, y \in \Omega$ we have

$$\begin{aligned}\omega_\delta(x, y) &\geq 1 \qquad \text{if } \rho(x, y) \leq \delta \\ \omega_\delta(x, y) &= 0 \qquad \text{if } \rho(x, y) \geq C\delta.\end{aligned}$$

(3) For every $J = (j_1, \ldots, j_r) \in \mathbb{I}_r$ and $L = (l_1, \ldots, l_s) \in \mathbb{I}_s$ there is a constant $C_{J,L}$ independent of δ so that

$$\sup_{(x,y)\in\Omega\times\Omega} \big|[Y_{j_r} \cdots Y_{j_1}][Y_{l_s} \cdots Y_{l_1}][\omega_\delta](x, y)\big| \leq C_{J,L}\, \delta^{-d(J)-d(L)}.$$

Here $[Y_{j_r} \cdots Y_{j_1}]$ acts on the variable x and $[Y_{l_s} \cdots Y_{l_1}]$ acts on the variable y.

Proof. Consider the collection of balls $\{B_\rho(x, \delta)\}$ for $x \in K$, and choose a maximal sub-collection $\{B_\rho(x_j, \delta)$ so that $B_\rho(x_j, \delta) \cap B_\rho(x_k, \delta) = \emptyset$ if $j \neq k$. For any $y \in K$, the ball $B_\rho(y, \delta)$ must intersect one of the balls $B_\rho(x_j, \delta)$, and hence $y \in B_\rho(x_j, 2\delta)$ for some j. Thus the collection $\{B_\rho(x_j, 2\delta)\}$ covers K. By Proposition 5.4, each point in K belongs to at most M of these larger balls. But then if $x, y \in K$ with $\rho(x, y) \leq \delta$, there exists an index j so that both x and y belong to $B_\rho(x_j, 3\delta)$.

Now using Lemma 5.3, for each j there is a function $\varphi_j \in \mathcal{C}_0^\infty(\Omega)$ such that $\varphi_j(x) = 1$ if $\rho(x_j, x) \leq 3\delta$, $\varphi_j(x) = 0$ if $\rho(x_j, x) \geq C\,\delta$, and

$$\sup_{y\in\Omega} \big|Y_{j_r} \cdots Y_{j_1}[\varphi_j](y)\big| \leq C_J\, \delta^{-d(J)}$$

for every $J \in \mathbb{I}_r$. Put

$$\omega_\delta(x, y) = \sum_j \varphi_j(x)\varphi_j(y).$$

Note that each term in this sum is non-negative. if $\varphi_j(x)\varphi_j(y) \neq 0$, then $x, y \in B_\rho(x_j, C\,\delta)$. There are at most M balls $B_\rho(x_j, C\delta)$ which contain x. Thus for $x, y \in \Omega$, there are at most M terms in this sum which are non-vanishing. Since each term is bounded by 1, it follows that for all $x, y \in \Omega$,

$$0 \leq \omega_j(x, y) \leq M.$$

Next, if $\rho(x, y) \geq 2C\delta$, it follows that there are no balls $B_\rho(x_j, C\delta)$ containing both x and y, and hence each term $\varphi_j(x)\varphi_j(y) = 0$. This shows that

$$\omega_j(x, y) = 0 \qquad \text{if } \rho(x, y) \geq 2C\delta.$$

If $\rho(x, y) < \delta$, we have seen that there is an index j so that $x, y \in B_\rho(x_j, 3\delta)$. Hence $\varphi_j(x) = \varphi_j(y) = 1$, and it follows that

$$\omega_j(x, y) \geq 1 \qquad \text{if } \rho(x, y) < \delta.$$

Finally, we have

$$\big|[Y_{j_r} \cdots Y_{j_1}][Y_{l_s} \cdots Y_{l_1}][\omega_\delta](x, y)\big| \leq \sum_j \big|Y_{j_r} \cdots Y_{j_1}[\varphi_j](x)\big|\,\big|[Y_{l_s} \cdots Y_{l_1}[\varphi_j](y)\big|$$

For each$x, y \in \Omega$, here are at mots M non-zero terms in this sum, and by Lemma 5.3, each term term is bounded by $C_{J,L}\,\delta^{-d(J)-d(L)}$. Thus

$$\sup_{x,y\in\Omega} \big|[Y_{j_r} \cdots Y_{j_1}][Y_{l_s} \cdots Y_{l_1}][\omega_\delta](x, y)\big| \leq C_{J,L}\,M\,\delta^{-d(J)-d(L)}.$$

This completes the proof. □

Remark: *Variants of this a result appear in* [**Sta93**]*, and in* [**Koe00**] *and* [**Koe02**].

Theorem 5.6. *Let $K \subset \Omega$ be a compact set. There is a function $\widetilde{\rho} : \Omega \times \Omega \to [0, \infty 0$ which is infinitely differentiable away from the diagonal*

$$\triangle = \Big\{(x, y) \in \Omega \times \Omega \,\Big|\, x = y\Big\}$$

with the following properties.

(1) There is a constant $C > 0$ so that for all $x, y \in K$ with $x \neq y$ we have

$$C^{-1} \leq \frac{\rho(x, y)}{\widetilde{\rho}(x, y)} \leq C.$$

(2) For every $J \in \mathbb{I}_r$ and $L \in \mathbb{I}_s$ there is a constant $C_{K,L}$ so that for all $x, y \in K$ with $x \neq y$ we have

$$[Y_{j_r} \cdots Y_{j_1}][Y_{l_s} \cdots Y_{l_1}][\widetilde{\rho}(x, y)] \leq C_{J,L}\,\widetilde{\rho}(x, y)^{1-d(J)-d(L)}.$$

Proof. There are functions $\{\omega_j\}$ defined on $\Omega \times \Omega$ such that

$$\begin{aligned} 0 \leq \omega_j(x, y) &\leq M && \text{for all } x, y \in \Omega, \\ \omega_j(x, y) &\geq 1 && \text{when } 0 \leq \rho(x, y) \leq 2^{-j}, \\ \omega_j(x, y) &\equiv 0 && \text{when } \rho(x, y) \geq C\,2^{-j}. \end{aligned}$$

Now put

$$\widetilde{\rho}(x, y) = \begin{cases} 0 & \text{if } x = y \\ \Big[\sum_{j=1}^\infty 2^j\,\omega_j(x, y)\Big]^{-1} & \text{if } x \neq y. \end{cases}$$

Fix $x \neq y$. Let j_0 be the smallest non-negative integer such that $C\,2^{-j} \leq \rho(x,y)$. Then for $j \geq j_0$ we have $\omega_j(x,y) = 0$ and so

$$\begin{aligned}\sum_{j=0}^{\infty} 2^j\,\omega_j(x,y) &= \sum_{j=0}^{j_0-1} 2^j\,\omega_j(x,y)\\ &\leq M \sum_{j=0}^{j_0-1} 2^j\\ &< M\,2^{j_0} < 2CM\,\rho(x,y)^{-1}\end{aligned}$$

since by hypothesis $C\,2^{-(j_0-1)} > \rho(x,y)$. Next let j_1 be the largest integer such that $\rho(x,y) \leq 2^{-j_1}$. Then for $0 \leq j \leq j_1$ we have $\omega_j(x,y) \geq 1$, and so

$$\begin{aligned}\sum_{j=0}^{\infty} 2^j\,\omega_j(x,y) &\geq \sum_{j=0}^{j_1} 2^j\,\omega_j(x,y)\\ &\geq \sum_{j=0}^{j_1} 2^j\\ &> 2^{j_1} > \frac{1}{2}\,\rho(x,y)^{-1}\end{aligned}$$

since by hypothesis $2^{-j_1-1} < \rho(x,y)$. Thus it follows that for $x \neq y$ we have

$$\frac{1}{2CM}\,\rho(x,y) < \Big[\sum_{j=0}^{\infty} 2^j\,\omega_j(x,y)\Big]^{-1} < 2\rho(x,y),$$

and so $\widetilde{\rho}$ is comparable to ρ, proving (1).

The differential inequality in (2) follows in the same way, using the fact that

$$\big|[Y_{j_r}\cdots Y_{j_1}][Y_{l_s}\cdots Y_{l_1}][\omega_j](x,y)\big| \leq C_{J,L}\,(2^{-j})^{-d(J)-d(L)}.$$

□

5.2. Scaled bump functions.

CHAPTER 4

Subelliptic estimates and hypoellipticity

1. Introduction

In this chapter we study of the L^2-Sobolev regularity properties of second order partial differential operators of the form $\mathcal{L} = X_0 - \sum_{j=1}^{p} X_j^2 + iY_0 + c$, where $\{X_0, X_1, \ldots, X_p\}$ and Y_0 are smooth real vector fields and c is a smooth complex-valued function. Under an appropriate finite type hypothesis on the vector fields $\{X_0, \ldots, X_p\}$ and an appropriate assumption about the size of the vector field Y_0, we shall derive what are called *subelliptic estimates* for $\mathcal{L}$ in the scale of L^2-Sobolev spaces $H^s(\mathbb{R}^n)$. The definitions and properties of these Sobolev spaces are discussed in Chapter 10.

2. Subelliptic estimates

2.1. Statement of the main theorem. Let $\Omega \subset \mathbb{R}^n$ be a connected open set. Let $X_0,\, X_1, \ldots, X_p,\, Y_0 \in T(\Omega)$ be smooth real vector fields, and let $c \in \mathcal{C}^\infty(\Omega)$ be a (possibly) complex-valued function. We study the second order partial differential operator

$$\mathcal{L} = X_0 - \sum_{j=1}^{p} X_j^2 + i\, Y_0 + c. \tag{2.1}$$

A subelliptic estimate for $\mathcal{L}$ asserts the existence of a constant $\epsilon > 0$ with the following property. Suppose $\zeta \prec \zeta' \in \mathcal{C}_0^\infty(\Omega)$. (Recall that this means that $\zeta'(x) = 1$ for all x in the support of ζ). Then for all $s \in \mathbb{R}$, there is a constant C_s so that if $u \in \mathcal{D}'(\Omega)$ is a distribution,

$$\big|\big|\, \zeta\, u \,\big|\big|_{s+\epsilon} \leq C_s \Big[\big|\big|\, \zeta' \mathcal{L}[u] \,\big|\big|_s + \big|\big|\, \zeta' u \,\big|\big|_s \Big].$$

Note that the three terms $\zeta_2 u$, $\zeta_1 \mathcal{L}[u]$, and $\zeta_1 u$ are distributions with compact support in Ω. Part of the content of subellipticity is that if the two terms on the right hand side are finite, then the left hand side is also finite.

In the operator $\mathcal{L}$, the term $i\,Y_0 + c$ is regarded as a perturbation of the main term $X_0 - \sum_{j=1}^{p} X_j^2$. In order to obtain subelliptic[1] estimates, we shall assume the vector fields $\{X_0, \ldots, X_p\}$ are of finite type and the vector field Y_0 is suitably dominated by the vector fields $\{X_1, \ldots, X_p\}$. Precisely, we assume:

---

[1]The expression 'subelliptic estimate' derives from a comparison of the above estimate with the classical estimates for elliptic operators. Thus if $\triangle = \sum_{j=1}^{n} \partial_{x_j}^2$ is the Laplace operator on $\mathbb{R}^n$ and if $\zeta \prec \zeta' \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, then for every $s \in \mathbb{R}^n$ there is a constant C_s so that if $u \in \mathcal{D}'(\mathbb{R}^n)$,

$$\big|\big|\, \zeta u \,\big|\big|_{s+2} \leq C_s \big[\big|\big|\, \zeta' \triangle[u] \,\big|\big|_s + \big|\big|\, \zeta' u \,\big|\big|_s \big].$$

(H-1) The iterated commutators of length less than or equal to m of the vector fields $\{X_0, \ldots, X_p\}$ uniformly[2] span the tangent space $T_x(\mathbb{R}^n)$ for every $x \in \Omega$.

(H-2) There exist constants $\eta > 0$ and C_0, $C_1 < \infty$ so that for all $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$

$$\big|\big(Y_0[\varphi], \varphi\big)_0\big| \leq (1-\eta)\sum_{j=1}^{p} \big|\big|\, X_j[\varphi]\, \big|\big|_0^2 + C_0 \big|\big|\, \varphi\, \big|\big|_0^2, \quad \text{and} \tag{2.2}$$

$$\big|\big(Y_0[\varphi], \psi\big)_0\big| \leq C_1 \Big[\sum_{j=1}^{p} \big(\big|\big|\, X_j[\varphi]\, \big|\big|_0^2 + \big|\big|\, X_j[\psi]\, \big|\big|_0^2\big) + \big|\big|\, \varphi\, \big|\big|_0^2 + \big|\big|\, \psi\, \big|\big|_0^2\Big]. \tag{2.3}$$

Write the vector fields $X_j = \sum_{k=1}^n a_{j,k}\, \partial_{x_k}$ and $Y_0 = \sum_{k=1}^n b_k\, \partial_{x_k}$ where $\{a_{j,k}\}$, $b_k \in \mathcal{C}_{\mathbb{R}}^\infty(\Omega)$. By shrinking the set Ω, we can always assume in addition that:

(H-3) All derivatives of the functions $a_{j,k}$, b_k, c are bounded on the set Ω.

Before stating the basic subelliptic estimate for $\mathcal{L}$ in the Sobolev space H^s, we need to discuss the nature of the constants that appear in the theorem. For later application it will be important to see that the bounds we obtain are uniform in the sense that they only depend on the parameter s, on the size of the coefficients of the operator $\mathcal{L}$, and on the choice of various cut-off functions $\{\zeta_j\} \subset \mathcal{C}_0^\infty(\Omega)$ that are used in the proof.

DEFINITION 2.1. *Fix $\zeta_0, \ldots, \zeta_M \in \mathcal{C}_0^\infty(\Omega)$. For any non-negative integer N, set*

$$\begin{aligned} B(N) = \sup_{x\in V}\ \sup_{|\alpha|\leq N} \Big[\sum_{j=1}^{p}\sum_{k=1}^{n} \big|\partial^\alpha a_{j,k}(x)\big| + \sum_{k=1}^{n} \big|\partial^\alpha b_k(x)\big| \\ + \big|\partial^\alpha c(x)\big| + \sum_{j=0}^{M}\sum_{k=1}^{n} \big|\partial^\alpha \zeta_j(x)\big|\Big]. \end{aligned} \tag{2.4}$$

If $s \in \mathbb{R}$, a constant C_s is allowable if there is an integer N_s so that C_s depends only on s and on $B(N_s)$.

With this definition in hand, we now state the main result of this chapter.

THEOREM 2.2. *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $\mathcal{L}$ be the second order partial differential operator given in equation (2.1). Suppose the vector fields $\{X_0, X_1, \ldots, X_p, Y_0\}$ and the function c satisfy hypotheses (H-1), (H-2), and (H-3). Set $\epsilon = 2\cdot 4^{-m}$, and fix $\zeta \prec \zeta' \in \mathcal{C}_0^\infty(\Omega)$. For every $s \in \mathbb{R}$ there is an allowable constant C_s so that if $u \in \mathcal{D}'(\Omega)$ and if $\zeta' u \in H^s(\mathbb{R}^n)$ and $\zeta'\mathcal{L}[u] \in H^s(\mathbb{R}^n)$, then $\zeta u \in H^{s+\epsilon}(\mathbb{R}^n)$, $\zeta X_j u \in H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$ for $1 \leq j \leq p$, and*

$$\big|\big|\, \zeta u\, \big|\big|_{s+\epsilon} + \sum_{j=1}^{p} \big|\big|\, \zeta X_j u\, \big|\big|_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\, \zeta'\mathcal{L}[u]\, \big|\big|_s + \big|\big|\, \zeta' u\, \big|\big|_s\Big]. \tag{2.5}$$

[2]This means that there is a constant $\eta > 0$ so that for all $x \in \Omega$, there is an n-tuple of iterated commutators of length at most m whose determinant at x is bounded below in absolute value by η. This can always be achieved by shrinking Ω.

2.2. Hypoellipticity of $\mathcal{L}$. The classical Weyl Lemma asserts that if u is a distribution on $\mathbb{R}^n$ and if $\triangle[u]$ is smooth on an open set $U \subset \mathbb{R}^n$, then u itself is smooth on U. This qualitative property of the Laplace operator $\triangle$ and other elliptic operators is called hypoellipticity.

DEFINITION 2.3. *If L is a linear differential operator with $\mathcal{C}^\infty$-coefficients on an open set $\Omega \subset \mathbb{R}^n$, then L is hypoelliptic provided that for every open set $U \subset \Omega$, if $u \in \mathcal{D}'(\Omega)$ is a distribution and if $L[u] \in \mathcal{C}^\infty(U)$, then $u \in \mathcal{C}^\infty(U)$.*

An important consequence of Theorem 2.2 is that the operator $\mathcal{L}$ is *hypoelliptic*.

COROLLARY 2.4. *Suppose that $\mathcal{L} = X_0 + \sum_{j=1}^p X_j^2 + iY_0 + c$, and that the vector fields $\{X_0, X_1, \ldots, X_p, Y_0\}$ and the function c satisfy hypotheses (H-1), (H-2), and (H-3). Then $\mathcal{L}$ is hypoelliptic.*

PROOF. Let $\varphi \in \mathcal{C}_0^\infty(U)$. Choose $\zeta_0, \zeta_1, \ldots \in \mathcal{C}_0^\infty(U)$ with $\varphi \prec \zeta_j$ and $\zeta_{j+1} \prec \zeta_j$ for all j. Since $\zeta_0 u$ has compact support, there exists an $s \in \mathbb{R}$ such that $\zeta_0 u \in H^s(\mathbb{R}^n)$. Since by hypothesis $\zeta_0 \mathcal{L}[u] \in H^s(\mathbb{R}^n)$, the inequality then shows that $\zeta_1 u \in H^{s+\epsilon}(\mathbb{R}^n)$. We can then repeat this argument N times to show that $\zeta_N u \in H^{s+N\epsilon}(\mathbb{R}^n)$. It follows that $\varphi u \in \bigcap_t H^t(\mathbb{R}^n) \subset \mathcal{C}^\infty(\mathbb{R}^n)$. Since this is true for any $\varphi \in \mathcal{C}_0^\infty(U)$, it follows that $u \in \mathcal{C}^\infty(U)$. □

2.3. Commentary on the argument. The proof of Theorem 2.2 uses the Schwarz inequaltiy, the calculus of pseudodifferential operators, and a long and intricate series of calculations involving integration by parts. The argument proceeds in two main steps. The first deals with estimates when the distribution $u \in \mathcal{D}'(\Omega)$ is actually a function $\varphi \in \mathcal{C}_0^\infty(\Omega)$. Since one is assuming the data is differentiable, these are called *a priori* estimates. The second step then deals with the passage from smooth functions to general distributions $u \in \mathcal{D}'(\Omega)$.

Let us isolate some key elements in the proof of the *a priori* estimates. The first is the proof that if $\varphi \in \mathcal{C}_0^\infty(\Omega)$, then $\sum_{j=1}^p \big|\big| X_j\varphi \big|\big|_s^2$ is controlled by $\Re e\big(\mathcal{L}[\varphi], \varphi\big)_s$ and $\big|\big| \varphi \big|\big|_s^2$. This is the subject of Section 3. In particular, the conclusion of Lemma 3.12 is that

$$\frac{\eta}{8}\sum_{j=1}^p \big|\big| X_j\varphi \big|\big|_s^2 \leq \Re e\big(\mathcal{L}[\varphi], \varphi\big)_s + C_s\big|\big| \varphi \big|\big|_s^2$$

where η is the constant that appears in hypothesis (H-2), equation (2.2).

For $s = 0$, this estimates follows easily from an argument involving integration by parts. The proof is given in Lemma 4.2. In order to pass to a general parameter s, we write $\big|\big| X_j\varphi \big|\big|_s = \big|\big| \Lambda^s X_j\varphi \big|\big|_0$ and $\Re e\big(\mathcal{L}[\varphi], \varphi\big)_s = \Re e\big(\Lambda^s\mathcal{L}[\varphi], \Lambda^s\varphi\big)_0$. Roughly speaking, we would like to replace $\big|\big| \Lambda^s X_j\varphi \big|\big|_0$ by $\big|\big| X_j\Lambda^s\varphi \big|\big|_0$, and $\Re e\big(\Lambda^s\mathcal{L}[\varphi], \Lambda^s\varphi\big)_0$ by $\Re e\big(\mathcal{L}[\Lambda^s\varphi], \Lambda^s\varphi\big)_0$, and then apply the case $s = 0$. Unfortunately, $\Lambda^s\varphi$ is not compactly supported, so we introduce additional cut-off functions to deal with this difficulty. In addition, we need to control the commutators of X_j and $\mathcal{L}$ with pseudodifferential operators A^s of order s. These manipulations generate a large number of terms, all of which must be bounded by the right hand side of equation (2.5).

In dealing with the commutators, note that an error term $\big|\big| [X_j,\, A^s]\,\varphi \big|\big|_0$ is dominated by $\big|\big| \varphi \big|\big|_s$ since $[X_j,\, A^s]$ is an operator of order s. However an error term $\big|\big| [\mathcal{L},\, A^s]\,\varphi \big|\big|_0$ cannot be handled so easily since $[\mathcal{L},\, A^s]$ is an operator of order $s+1$. Thus a second key ingredient in the proof is the observation that such commutators are not just an operator of order $s+1$, but can be controlled by operators of order s composed with the vector fields $\{X_1, \ldots, X_p\}$ or their adjoints. In Proposition 3.6 below we see that one can write

$$[\mathcal{L}, A^s] = \sum_{j=1}^{p} B_j^s X_j + B_0^s$$

$$[\mathcal{L}, A^s] = \sum_{j=1}^{p} X_j^* B_j^s + \widetilde{B}_0^s$$

where $\{B_j^s\}$ and $\{\widetilde{B}_j^s\}$ are pseudodifferential operators of order s. This is used in Lemmas 3.7 and 3.10 to estimate the commutator $[\mathcal{L}, A^s]$ in terms of the vectors $\{X_j\}$.

If the vectors $\{X_1, \ldots, X_p\}$ spanned the tangent space at each point, we could immediately estimate $|| \varphi ||_{s+1}^2$ in terms of $\sum_{j=1}^{p} || X_j \varphi ||_s^2$. In order to use the much weaker finite type hypothesis (H-1) and complete the proof of the *a priori* estimates, we need to employ a third key observation: if we have a favorable estimate

$$|| Z\varphi ||_{s+a-1} \lesssim || \mathcal{L}\varphi ||_s + || \varphi ||_s$$

for a vector field Z and some $a > 0$, then we get a possibly less good but still favorable estimate

$$|| [X_j, Z]\varphi ||_{s+b-1} \lesssim || \mathcal{L}\varphi ||_s + || \varphi ||_s$$

for all $0 \leq j \leq p$ and some $0 < b \leq a$. This is the content of Lemma 4.5 below. More generally, in Section 4 we study spaces of subelliptic multipliers, which are pseudodifferential operators A such that

$$|| A\varphi ||_{s+a-1} \lesssim || \mathcal{L}\varphi ||_s + || \varphi ||_s$$

for some $a > 0$.

In order to pass from *a priori* estimates of smooth functions to estimates for general distributions $u \in \mathcal{D}'(\Omega)$, we multiply u by a cut-off function $\zeta \in \mathcal{C}_0^\infty(\Omega)$, and then consider the 'mollifier' or convolution $T_t[\zeta\, u] = \zeta\, u * \chi_t$ where $\chi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and $\chi_t(x) = t^{-n}\chi(t^{-1}x)$. For $t > 0$ sufficiently small, $T_t[\zeta\, u] \in \mathcal{C}_0^\infty(\Omega)$, and so we obtain

$$|| T_t[\zeta\, u] ||_{s+\frac{1}{2}\epsilon} \lesssim || \mathcal{L}\big[T_t[\zeta\, u]\big] ||_s + || T_t[\zeta\, u] ||_s$$

from the *a priori* estimate. To complete the proof of Theorem 2.2, we then need to control the commutator $[\mathcal{L}, T_t\zeta]$. This is done in Section 5. Finally in Section 6, we give examples of vector fields Y_0 which satisfy hypothesis (H-2).

3. Estimates for $\sum_{j=1}^{p} || X_j \varphi ||_s^2$

In this section we prove that for every $s \in \mathbb{R}$ there is an allowable constant C_s so that

$$\sum_{j=1}^{p} || X_j \varphi ||_s^2 \leq C_s \Big[|| \mathcal{L}[\varphi] ||_s^2 + || \varphi ||_s^2 \Big]$$

for every $\varphi \in \mathcal{C}_0^\infty(\Omega)$.

3.1. Integration by parts. Recall that the inner product of two functions $f, g \in L^2(\Omega)$ is given by

$$\big(f, g\big)_0 = \int_\Omega f(x)\, \overline{g(x)}\, dx,$$

and the norm of $f \in L^2(\Omega)$ is denoted by $||\, f\, ||_0 = \sqrt{\big(f, f\big)_0}$. If $X = \sum_{k=1}^n a_k\, \partial_{x_k}$ is a smooth real vector field, the divergence of X is the smooth function

$$\nabla \cdot X(x) = \sum_{k=1}^n \frac{\partial a_k}{\partial x_k}(x).$$

The *formal adjoint* of the vector field X is the operator X^* whose action on a function φ is given by

$$X^*[\varphi](x) = -X[\varphi](x) - \nabla \cdot X(x)\, \varphi(x).$$

The following proposition summarizes the elementary properties of formal adjoints that arise in integration by parts.

PROPOSITION 3.1. *Let X, Y be smooth real vector fields on an open set $\Omega \subset \mathbb{R}^n$, let $f \in \mathcal{C}^\infty_{\mathbb{R}}(\Omega)$ be a smooth real-valued function, and let $\varphi, \psi \in \mathcal{C}^\infty_0(\Omega)$ be (possibly) complex-valued functions with compact support. Then*

$$\big(X[\varphi], \psi\big)_0 = \big(\varphi, X^*[\psi]\big)_0 = -\big(\varphi, X[\psi]\big)_0 - \big(\varphi, \nabla \cdot X\, \psi\big)_0; \tag{3.1}$$

$$\big(X[\varphi], f\, \psi\big)_0 = -(\varphi, f\, X[\psi])_0 + \big(\varphi, X^*[f]\, \psi\big)_0; \tag{3.2}$$

$$\big(XY[\varphi], \psi\big)_0 = -\big(Y[\varphi], X[\psi]\big)_0 - \big(Y[\varphi], \nabla \cdot X \psi\big)_0; \tag{3.3}$$

$$\Re e\big(X[\varphi], \varphi\big)_0 = -\frac{1}{2}(\varphi, \nabla \cdot X\, \varphi)_0; \tag{3.4}$$

$$\Re e\big(X[\varphi], f\, \varphi\big)_0 = \frac{1}{2}\big(\varphi, X^*[f]\, \varphi\big)_0; \tag{3.5}$$

$$\Re e\big(X^2[\varphi], \varphi\big)_0 = -||\, X[\varphi]\, ||_0^2 - \big(\varphi, X^*\big[\nabla \cdot X\big]\, \varphi\big)_0. \tag{3.6}$$

PROOF. Let $X = \sum_{j=1}^n a_j\, \partial_{x_j}$ be a smooth real vector field on Ω, and suppose that $\varphi, \psi \in \mathcal{C}^\infty_0(\Omega)$ are complex-valued functions. The function $a_j \varphi\, \overline{\psi}$ has compact support in Ω, and so $\int_\Omega \partial_{x_j}(a_j \varphi\, \overline{\psi})\, dx = 0$ by the fundamental theorem of calculus. Thus integration by parts gives

$$\begin{aligned}
\big(X[\varphi],\, \psi\big)_0 &= \sum_{j=1}^n \int_\Omega a_j(x) \partial_{x_j}(x)\, \overline{\psi(x)}\, dx \\
&= \sum_{j=1}^n \int_\Omega \partial_{x_j}(a_j\, \varphi\, \overline{\psi})(x)\, dx - \sum_{j=1}^n \int_\Omega \varphi(x)\, \partial_{x_j}(a_j\, \overline{\psi})(x)\, dx \\
&= -\int_\Omega \varphi(x)\, \overline{\Big[\sum_{j=1}^n a_j(x)\, \partial_{x_j} \psi(x)\Big]}\, dx - \int_\Omega \varphi(x)\, \overline{\Big[\sum_{j=1}^n \partial_{x_j} a_j(x)\Big] \psi(x)}\, dx \\
&= -\big(\varphi, X[\psi]\big)_0 - \big(\varphi, \nabla \cdot X\, \psi\big)_0.
\end{aligned}$$

This gives equation (4.1). Equations (3.2) through (3.6) follow from repeated applications of this identity. □

We can now calculate the formal adjoint of the operator $\mathcal{L}$.

PROPOSITION 3.2. *Let* $\mathcal{L} = -\sum_{j=1}^p X_j^2 + X_0 + iY_0 + c$ *be a second order partial differential operator on an open set* $\Omega \subset \mathbb{R}^n$ *as given in equation (2.1). If* $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$*, then* $\big(\mathcal{L}[\varphi], \psi\big)_0 = \big(\varphi, \mathcal{L}^*[\psi]\big)_0$ *where*

$$\mathcal{L}^* = -\sum_{j=1}^p X_j^2 - X_0 + \sum_{j=1}^p 2\,\nabla\!\cdot\! X_j\, X_j + iY_0 + \widetilde{c}, \tag{3.7}$$

and $\widetilde{c}(x) = c(x) - div[X_0](x) + i\,div[Y_0](x) - \sum_{j=1}^p \Big[X_j\big[\nabla\!\cdot\! X_j\big] + (\nabla\!\cdot\! X_j)^2\Big](x).$

PROOF. Using equation (3.3) we have

$$\begin{aligned}
\big(X_j^2[\varphi], \psi\big)_0 &= -\big(X_j[\varphi], X_j[\psi]\big)_0 - \big(X_j[\varphi], \nabla\!\cdot\! X_j\psi\big)_0 \\
&= \big(\varphi, X_j^2[\psi]\big)_0 + \big(\varphi, \nabla\!\cdot\! X_j X_j[\psi]\big)_0 \\
&\quad + \big(\varphi, X_j\big[\nabla\!\cdot\! X_j\psi\big]\big)_0 + \big(\varphi, (\nabla\!\cdot\! X_j)^2\psi\big)_0 \\
&= \big(\varphi, X_j^2[\psi]\big)_0 + \big(\varphi, 2\nabla\!\cdot\! X_j X_j[\psi]\big)_0 \\
&\quad + \big(\varphi, \Big[X_j\big[\nabla\!\cdot\! X_j\big] + (\nabla\!\cdot\! X_j)^2\Big]\psi\big)_0
\end{aligned}$$

Summing and using equation (4.1) gives the desired formula. □

REMARK 3.3. *If we set* $\widetilde{X}_0 = -X_0 + 2\sum_{j=1}^p \nabla\!\cdot\! X_j\, X_j$*, then*

$$\mathcal{L}^* = -\sum_{j=1}^p X_j^2 + \widetilde{X}_0 + iY_0 + \widetilde{c},$$

and so has $\mathcal{L}^*$ *has essentially the same form as* $\mathcal{L}$*. Note that the linear span of the vector fields* $\{X_0, X_1, \ldots, X_p\}$ *is the same as the linear span of* $\{\widetilde{X}_0, X_1, \ldots, X_p\}$*.*

3.2. The basic L^2-identity and L^2-inequality.

The key first step in deriving L^2-estimates for the operator $\mathcal{L}$ is to observe that the quadratic form $\big(\mathcal{L}[\varphi], \varphi\big)_0$ can be rewritten in terms of the vector fields X_0, $X_1, \ldots, X_p$, Y_0. We have the following *basic identity*.

LEMMA 3.4. *For* $\varphi \in \mathcal{C}_0^\infty(\Omega)$*,*

$$\begin{aligned}
\Re e\big(\mathcal{L}[\varphi], \varphi\big)_0 &= \sum_{j=1}^p ||\, X_j[\varphi]\,||_0^2 - \Im m\big(Y_0[\varphi], \varphi\big)_0 \\
&\quad + \frac{1}{2}\sum_{j=1}^p \big(\varphi, X_j^*\big[\nabla\!\cdot\! X_j\big]\varphi\big)_0 - \frac{1}{2}\big(\varphi, div[X_0]\varphi\big)_0 + \Re e\big(c\varphi, \varphi\big)_0
\end{aligned} \tag{3.8}$$

PROOF. We have

$$\big(\mathcal{L}[\varphi], \varphi\big)_0 = -\sum_{j=1}^p \big(X_j^2[\varphi], \varphi\big)_0 + \big(X_0[\varphi], \varphi\big)_0 + i\big(Y_0[\varphi], \varphi\big)_0 + \big(c\varphi, \varphi\big)_0$$

It follows from equation (3.6) that

$$-\big(X_j^2[\varphi], \varphi\big)_0 = ||\, X_j[\varphi]\,||_0^2 + \big(X_j[\varphi], \mathrm{div}(X_j)\,\varphi\big)_0,$$

so that

$$\big(\mathcal{L}[\varphi],\varphi\big)_0 = \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_0^2 + \sum_{j=1}^{p} \big(X_j[\varphi], \nabla\cdot X_j\varphi\big)_0 + \big(X_0[\varphi],\varphi\big)_0 + i\big(Y_0[\varphi],\varphi\big)_0 + \big(c\varphi,\varphi\big)_0.$$

Taking real parts and using equations (3.4) and (3.5) gives the equality stated in (3.8). □

Note that the last three terms in equation (3.8) do not involve any differentiation of the function φ. Thus these terms can be estimated in terms of $\big|\big|\,\varphi\,\big|\big|_0^2$ and are usually viewed as error terms while the first two terms in (3.8) are the main terms. We now use the first hypothesis on the vector field Y_0; recall we assume that for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$,

$$\big|\Im m\,\big(Y_0[\varphi],\varphi\big)_0\big| \leq (1-\eta)\sum_{j=1}^{p}\big|\big| X_j[\varphi] \big|\big|_0^2 + C_0\big|\big|\,\varphi\,\big|\big|_0^2. \tag{2.2}$$

This leads to the following *basic estimate*.

Lemma 3.5. *Suppose that the vector field Y_0 satisfies (2.2). Then for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$*

$$\eta\sum_{j=1}^{p}\big|\big| X_j[\varphi] \big|\big|_0^2 \leq \Re e\big(\mathcal{L}[\varphi],\varphi\big)_0 + C_1\big|\big|\,\varphi\,\big|\big|_0^2 \tag{3.9}$$

where $C_1 = C_0 + \sup_{x\in\Omega}\Big[\frac{1}{2}|div[X_0](x)| + \frac{1}{2}\sum_{j=1}^{p}|X_j^*\big[\nabla\cdot X_j\big](x)| + |c(x)|\Big]$.

Proof. Using equations (3.8) and (2.2), if $\varphi \in \mathcal{C}_0^\infty(U)$ we have

$$\begin{aligned}\sum_{j=1}^{p}\big|\big| X_j[\varphi] \big|\big|_0^2 &\leq \Re e\big(\mathcal{L}[\varphi],\varphi\big)_0 + (1-\eta)\sum_{j=1}^{p}\big|\big| X_j[\varphi] \big|\big|_0^2 + C_0\big|\big|\,\varphi\,\big|\big|_0^2 \\ &\qquad - \frac{1}{2}\sum_{j=1}^{p}\big(\varphi, X_j^*\big[\nabla\cdot X_j\big]\varphi\big)_0 + \frac{1}{2}\big(\varphi, \mathrm{div}[X_0]\varphi\big)_0 + \Re e\big(c\varphi,\varphi\big)_0 \\ &\leq \Re e\big(\mathcal{L}[\varphi],\varphi\big)_0 + (1-\eta)\sum_{j=1}^{p}\big|\big| X_j[\varphi] \big|\big|_0^2 + C_1\big|\big|\,\varphi\,\big|\big|_0^2,\end{aligned}$$

and this is equivalent to equation (4.3). □

3.3. Commutators. In this section we collect the statements and proofs of several estimates for commutators of pseudodifferential operators. In practice one needs to deal with a large number of commutators, and this can make it difficult to keep track of the allowable constants that arise in the estimates. We attempt to balance the need for precise descriptions of the allowable constants with the need for conceptually concise statements by using the following procedure. The first time we make an estimate, we write an explicit formula for the constant that shows it is allowable, but after this first appearance, we replace the explicit formula with a generic symbol C or C_s.

Recall that if A^s and B^t are operators of order s and t then the commutator $[A^s, B^t]$ is a pseudodifferential operator of order $s+t-1$. In particular, when we commute a pseudodifferential operator A^s of order s with the second order operator $\mathcal{L}$, the commutator $[\mathcal{L}, A^s]$ is of order $s+1$. However, because $\mathcal{L}$ has the special form given in equation (2.1), we can say more about the form of $[\mathcal{L}, A^s]$. This is based on the following elementary observation.

PROPOSITION 3.6. *Let $\mathcal{L} = -\sum_{j=1}^{p} X_j^2 + X_0 + iY_0 + c$, and let A^s be a pseudodifferential operator of order s. Then we can write*

$$[\mathcal{L}, A^s] = \sum_{j=1}^{p} B_j^s\, X_j + B_0^s \tag{3.10}$$

$$[\mathcal{L}, A^s] = \sum_{j=1}^{p} X_j^*\, B_j^s + \widetilde{B}_0^s \tag{3.11}$$

where

$$\begin{aligned}
B_j^s &= -2\,[X_j,\, A^s]\,,\\
B_0^s &= -\sum_{j=1}^{p} [X_j,\, [X_j,\, A^s]] + [X_0,\, A^s] + i\,[Y_0,\, A^s] + [g,\, A^s]\,,\\
\widetilde{B}_j^s &= \ \ 2\,[X_j,\, A^s]\\
\widetilde{B}_0^s &= \ \ \sum_{j=1}^{p}\Big([X_j,\, [X_j,\, A^s]] + 2\nabla\cdot X_j\,[X_j,\, A^s]\Big) + [X_0,\, A^s] + i\,[Y_0,\, A^s] + [g,\, A^s]\,.
\end{aligned}$$

The pseudodifferential operators $\{B_0^s, \widetilde{B}_0^s, B_1^s, \dots, B_p^s\}$ have order s. If A^s is properly supported, so are these operators.

PROOF. Since $\mathcal{L} = -\sum_{j=1}^{p} X_j^2 + X_0 + iY_0 + c$, we have

$$[\mathcal{L},\, A^s] = -\sum_{j=1}^{p}\big[X_j^2,\, A^s\big] + [X_0,\, A^s] + i\,[Y_0,\, A^s] + [c,\, A^s]\,.$$

But

$$\big[X_j^2,\, A^s\big] = X_j\,[X_j,\, A^s] + [X_j,\, A^s]\,X_j = \begin{cases} 2\,[X_j,\, A^s]\,X_j + [X_j,\, [X_j,\, A^s]]\,,\\[2ex] 2X_j\,[X_j,\, A^s] - [X_j,\, [X_j,\, A^s]]\,,\end{cases}$$

and this gives the desired formulas. □

We now establish two important consequences.

LEMMA 3.7. *Let A^s be a pseudodifferential operator of order s, let $U \Subset \Omega$ be a relatively compact open subset, and assume that $\varphi \in \mathcal{C}_0^\infty(U)$ implies $A^s[\varphi] \in \mathcal{C}_0^\infty(\Omega)$. There is an allowable constant C_s so that for all $\varphi \in \mathcal{C}_0^\infty(U)$*

$$\big|\big|\ [\mathcal{L},\, A^s]\,[\varphi]\,\big|\big|_0 \leq C_s\,\Big[\sum_{j=1}^{p} \big|\big|\, X_j[\varphi]\,\big|\big|_s + \big|\big|\,\varphi\,\big|\big|_s\Big], \tag{3.12}$$

and consequently

$$\big|\big| \mathcal{L}A^s[\varphi] \big|\big|_0 \leq C_s \Big[\big|\big| \mathcal{L}[\varphi] \big|\big|_s + \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big]. \tag{3.13}$$

Proof. Use Proposition 3.6 to write $[\mathcal{L}, A^s] = \sum_{j=1}^{p} B_j^s X_j + B_0^s$. Then

$$\begin{aligned}\big|\big| \,[\mathcal{L}, A^s]\,[\varphi] \big|\big|_0 &\leq \sum_{j=1}^{p} \big|\big| B_j^s X_j[\varphi] \big|\big|_0 + \big|\big| B_0^s[\varphi] \big|\big|_0 \\ &\leq \sum_{j=1}^{p} \big|\big| B_j^s \Lambda^{-s} \big|\big| \,\big|\big| X_j[\varphi] \big|\big|_s + \big|\big| B_0^s \Lambda^{-s} \big|\big| \,\big|\big| \varphi \big|\big|_s,\end{aligned}$$

and this give inequality (3.12). But then

$$\begin{aligned}\big|\big| \mathcal{L}A^s[\varphi] \big|\big|_0 &\leq \big|\big| A^s\mathcal{L}[\varphi] \big|\big|_0 + \big|\big| \,[\mathcal{L}, A^s]\,[\varphi] \big|\big|_0 \\ &\leq \big|\big| A^s\Lambda^{-s} \big|\big| \,\big|\big| \mathcal{L}[\varphi] \big|\big|_s + \big|\big| \,[\mathcal{L}, A^s]\,[\varphi] \big|\big|_0.\end{aligned}$$

Combining this with inequality (3.12) gives inequality (3.13). □

Remark 3.8. *We will see below in Corollary 3.13 that* $\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s \leq C_s\Big[\big|\big| \mathcal{L}\varphi \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big]$. *Hence it follows from equations (3.12) and (3.13) in Lemma 3.7 that in fact we have*

$$\big|\big| \,[\mathcal{L}, A^s]\, \big|\big|_0 \leq C_s\Big[\big|\big| \mathcal{L}\varphi \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big] \tag{3.12'}$$

$$\big|\big| \mathcal{L}A^s[\varphi] \big|\big|_0 \leq C_s\Big[\big|\big| \mathcal{L}\varphi \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big] \tag{3.13'}$$

Remark 3.9. *We will later need to commute $\mathcal{L}$ with two special kinds of operators of order zero. We take this opportunity to indicate the nature of the operators B_j^0 in these cases.*

(1) If $M[\varphi](x) = \zeta(x)\,\varphi(x)$ where $\zeta \in \mathcal{C}_0^\infty(\Omega)$, then M is a pseudodifferential operator of order zero. If X is a vector field and c is multiplication by a function, then

$$\begin{aligned}[c, M]\,[\varphi] &= 0 \\ [X, M]\,[\varphi] &= X[\zeta]\,\varphi \\ [X, [X, M]] &= X^2[\zeta]\,\varphi.\end{aligned}$$

(2) If $\chi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ and if $T[f](x) = \chi * f(x) = \int_{\mathbb{R}^n} f(y)\chi(x-y)\,dy$, then T is a pseudodifferential operator of order zero. If X is a vector field and c is multiplication by a function, then

$$\begin{aligned}[c, T]\,[\varphi](x) &= \int \big(c(x) - c(y)\big)\varphi(y)\chi(x-y)\,dy \\ [X, T]\,[\varphi] &= T[\nabla\cdot X\varphi] \\ [X, [X, T]] &= T[\nabla\cdot X^2\varphi]\end{aligned}$$

If we have an inner product rather than a norm in Lemma 3.7, we can obtain an improved estimate involving an arbitrarily small multiple of $\sum_{j=1}^{p} \big|\big|\, X_j[\varphi] \,\big|\big|_s$. This will be important, for example, in the proof of Lemma 3.12 below.

LEMMA 3.10. *Let A^s be a pseudodifferential operator of order s, let $U \Subset \Omega$ be a relatively compact open subset, and assume that $\varphi \in \mathcal{C}_0^\infty(U)$ implies $A^s[\varphi] \in \mathcal{C}_0^\infty(\Omega)$. There is an allowable constant C_s so that for every $\delta > 0$ and every $\varphi \in \mathcal{C}_0^\infty(U)$*

$$\Big|\Big([\mathcal{L},\, A^s]\,[\varphi], A^s[\varphi]\Big)\Big| \leq \delta \sum_{j=1}^{p} \big|\big|\, X_j[\varphi] \,\big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big|\, \varphi \,\big|\big|_s^2, \tag{3.14}$$

and consequently

$$\Re e\Big(\mathcal{L}A^s[\varphi], A^s[\varphi]\Big) \leq \Re e\Big(A^s\mathcal{L}[\varphi], A^s[\varphi]\Big) + \delta \sum_{j=1}^{p} \big|\big|\, X_j[\varphi] \,\big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big|\, \varphi \,\big|\big|_s^2. \tag{3.15}$$

PROOF. Again use Proposition 3.6 to write $[\mathcal{L},\, A^s]\,[\varphi] = \sum_{j=1}^{p} B_j^s X_j[\varphi] + B_0^s[\varphi]$ where $\{B_0^s, B_1^s, \ldots, B_p^s\}$ are pseudodifferential operators of order s and carry $\mathcal{C}_0^\infty(U)$ to $\mathcal{C}_0^\infty(\Omega)$. Using the inequality

$$\Big|\sum_{j=1}^{p} a_j b_j\Big| \leq \delta \sum_{j=1}^{p} |a_j|^2 + \frac{1}{4\delta}\sum_{j=1}^{p} |b_j|^2,$$

it follows that

$$\begin{aligned}
\Big|\Big([\mathcal{L},\, A^s]\,[\varphi], A^s[\varphi]\Big)\Big| &\leq \Big|\Big(B_0^s[\varphi], A^s[\varphi]\Big)\Big| + \sum_{j=1}^{p} \Big|\Big(B_j^s X_j[\varphi], A^s[\varphi]\Big)\Big| \\
&\leq \big|\big|\, \Lambda^{-s}(B_0^s)^* A^s \Lambda^{-s} \,\big|\big|\, \big|\big|\, \varphi \,\big|\big|_s^2 \\
&\qquad + \sum_{j=1}^{p} \big|\big|\, X_j[\varphi] \,\big|\big|_s \big|\big|\, \Lambda^{-s}(B_j^s)^* A^s \Lambda^{-s} \,\big|\big|\, \big|\big|\, \varphi \,\big|\big|_s \\
&\leq \delta \sum_{j=1}^{p} \big|\big|\, X_j[\varphi] \,\big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big|\, \varphi \,\big|\big|_s^2
\end{aligned}$$

This completes the proof. □

We shall need one additional technical lemma whose proof involves estimates of commutators.

LEMMA 3.11. *Suppose that T is a properly supported pseudodifferential operator of order $2s-1$ and that Y is a vector field on Ω such that*

$$\big|\big|\, Y[\varphi] \,\big|\big|_{4s-1} \leq C\Big[\big|\big|\, \mathcal{L}[\varphi] \,\big|\big|_0 + \big|\big|\, \varphi \,\big|\big|_0\Big]. \tag{3.16}$$

Then there is an allowable constant C_s so that

$$\sum_{j=1}^{p} \Big|\Big(X_j^2 Y[\varphi],\, T[\varphi]\Big)\Big| \leq C_s\Big[\big|\big|\, \mathcal{L}[\varphi] \,\big|\big|_0 + \big|\big|\, \varphi \,\big|\big|_0\Big]. \tag{3.17}$$

PROOF. We have

$$\begin{aligned}\Big(X_j^2 Y[\varphi], T[\varphi]\Big) &= \Big(X_j Y[\varphi], X_j^* T[\varphi]\Big)\\ &= \Big(X_j Y[\varphi], T X_j^*[\varphi]\Big) + \Big(X_j Y[\varphi], \big[X_j^*,\, T\big]\,[\varphi]\Big)\\ &= \Big(T^* X_j Y[\varphi], X_j^*[\varphi]\Big) + \Big(Y[\varphi], X_j^*\,\big[X_j^*,\, T\big]\,[\varphi]\Big)\\ &= \Big(X_j T^* Y[\varphi], X_j^*[\varphi]\Big) + \Big([T^*,\, X_j]\,Y[\varphi], X_j^*[\varphi]\Big)\\ &\quad + \Big(Y[\varphi], \big[X_j^*,\, T\big]\,X_j^*[\varphi]\Big) + \Big(Y[\varphi], \big[X_j^*,\, \big[X_j^*,\, T\big]\big]\,[\varphi]\Big).\end{aligned}$$

Note that $[T^*,\, X_j]$, $\big[X_j^*,\, T\big]$, and $\big[X_j^*,\, \big[X_j^*,\, T\big]\big]$ are also properly supported pseudodifferential operators of order $2s-1$. Thus for the last three terms we have

$$\begin{aligned}\Big|\Big([T^*,\, X_j]\,Y[\varphi], X_j^*[\varphi]\Big)\Big| &\le \big|\big|\,[T^*,\, X_j]\,\Lambda^{-2s+1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1}\big|\big|\,X_j^*[\varphi]\,\big|\big|_0\\ \Big|\Big(Y[\varphi], \big[X_j^*,\, T\big]\,X_j^*[\varphi]\Big)\Big| &\le \big|\big|\,\Lambda^{-2s+1}\,[T^*,\, X_j]\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1}\big|\big|\,X_j^*[\varphi]\,\big|\big|_0\\ \Big|\Big(Y[\varphi], \big[X_j^*,\, \big[X_j^*,\, T\big]\big]\,[\varphi]\Big)\Big| &\le \big|\big|\,\Lambda^{-2s+1}\,\big[X_j^*,\, \big[X_j^*,\, T\big]\big]\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1}\,\big|\big|\,\varphi\,\big|\big|_0.\end{aligned}$$

For the first term, we have

$$\Big|\Big(X_j T^* Y[\varphi], X_j^*[\varphi]\Big)\Big| \le \big|\big|\,X_j T^* Y[\varphi]\,\big|\big|_0\,\big|\big|\,X_j^*[\varphi]\,\big|\big|_0.$$

But

$$\big|\big|\,X_j T^* Y[\varphi]\,\big|\big|_0^2 \le C\Big[\Re e\Big(\mathcal{L}[T^* Y[\varphi]], T^* Y[\varphi]\Big) + \big|\big|\,T^* Y[\varphi]\,\big|\big|_0^2\Big]$$

Now

$$\big|\big|\,T^* Y[\varphi]\,\big|\big|_0 \le \big|\big|\,T^* \Lambda^{-2s+1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1},$$

and

$$\begin{aligned}\Big(\mathcal{L}[T^* Y[\varphi]], T^* Y[\varphi]\Big) &= \Big(T^* Y \mathcal{L}[\varphi], T^* Y[\varphi]\Big) + \Big([\mathcal{L},\, T^* Y]\,[\varphi], T^* Y[\varphi]\Big)\\ &= \Big(\mathcal{L}[\varphi], Y^* T T^* Y[\varphi]\Big)\\ &\qquad + \sum_{j=1}^{p}\Big(B_j^{2s} X_j[\varphi], T^* Y[\varphi]\Big) + \Big(B_0^{2s}[\varphi], T^* Y[\varphi]\Big)\\ &= \Big(\mathcal{L}[\varphi], Y^* T T^* Y[\varphi]\Big)\\ &\qquad + \sum_{j=1}^{p}\Big(X_j[\varphi], (B_j^{2s})^* T^* Y[\varphi]\Big) + \Big(\varphi, (B_0^{2s})^* T^* Y[\varphi]\Big)\end{aligned}$$

where the operators B_j^{2s} are properly supported of order $2s$. It follows that all the operators $\big\{Y^* T T^*, (B_0^{2s})^* T^*, \dots, (B_p^{2s})^* T^*\big\}$ are of order $4s-1$. Hence

$$\Big|\Big(\mathcal{L}[T^* Y[\varphi]], T^* Y[\varphi]\Big)\Big| \le C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0 + \sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_0 + \big|\big|\,\varphi\,\big|\big|_0\Big]\big|\big|\,Y[\varphi]\,\big|\big|_{4s-1}.$$

Putting all the estimates together, we obtain the desired estimate (3.17). □

3.4. The passage from L^2 to H^s. An important ingredient in the proof of sub-elliptic estimates for $\mathcal{L}$ is the fact that an estimate in the space $L^2(\mathbb{R}^n)$ can often be 'bootstrapped' into estimates in all Sobolev spaces $H^s(\mathbb{R}^n)$. The key to such arguments is the ability to control commutators and to localize the operator Λ^s. Let us try to understand the structure of these arguments.

Suppose A and B are operators and there is an estimate which says that for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$ we have $\big|\big| A[\varphi] \big|\big|_0 \leq C \big|\big| B[\varphi] \big|\big|_0$. We want to use this to establish a Sobolev space version $\big|\big| A[\varphi] \big|\big|_s \leq C_s \big|\big| B[\varphi] \big|\big|_s$. If it were true that:

(i) the operator Λ^s carries the space $\mathcal{C}_0^\infty(\Omega)$ to itself, and
(ii) the operator Λ^s commutes with the operators A and B,

then we could simply write

$$\begin{aligned} \big|\big| A[\varphi] \big|\big|_s &= \big|\big| \Lambda^s A[\varphi] \big|\big|_0 = \big|\big| A\big[\Lambda^s[\varphi]\big] \big|\big|_0 \\ &\leq C \big|\big| B\big[\Lambda^s[\varphi]\big] \big|\big|_0 = C \big|\big| \Lambda^s B[\varphi] \big|\big|_0 = C \big|\big| B[\varphi] \big|\big|_s. \end{aligned}$$

Of course, (i) is false unless s is a non-negative even integer, and (ii) is rarely true.

To deal with the fact that (ii) may be false, we must use the various estimates for commutators that were established in Section 3.3. To correct statement (i), let $U \Subset \Omega$ be a relatively compact open set, and choose $\zeta_1 \prec \zeta_2 \in \mathcal{C}_0^\infty(\Omega)$ such that $\zeta_1(x) = \zeta_2(x) = 1$ for all $x \in U$. If $\varphi \in \mathcal{C}_0^\infty(U)$, then $\varphi = \zeta_1\varphi$ and hence

$$\Lambda^s[\varphi] = \zeta_2\Lambda^s\zeta_1\varphi + (1-\zeta_2)\Lambda^s\zeta_1\varphi$$

Note that $\zeta_2\Lambda^s\zeta_1\varphi \in \mathcal{C}_0^\infty(\Omega)$, and so we can apply the L^2-estimate to this term. On the other hand, ζ_1 and $(1-\zeta_2)$ have disjoint supports. Thus the operator $(1-\zeta_2)\Lambda^s\zeta_1$ is infinitely smoothing and hence extends to a bounded operator from $H^s(\mathbb{R}^n)$ to $H^t(\mathbb{R}^n)$ for any s and t.

With this background discussion out of the way, we now state and prove the two main bootstrap results. We first obtain an extension of the 'basic estimate' from Lemma 4.2. Recall that this asserts that if the vector field Y_0 satisfies hypothesis (H-2), then for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$,

$$\eta \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_0^2 \leq \Re e\big(\mathcal{L}[\varphi], \varphi\big)_0 + C_1 \big|\big| \varphi \big|\big|_0^2. \tag{4.3}$$

LEMMA 3.12. *Suppose that the vector field Y_0 satisfies condition (2.2). Let $U \Subset \Omega$ be a relatively compact open subset. For every $s \in \mathbb{R}$ there is an allowable constant C_s so that for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$,*

$$\frac{\eta}{8} \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 \leq \Re e\big(\mathcal{L}[\varphi], \varphi\big)_s + C_s \big|\big| \varphi \big|\big|_s^2. \tag{3.18}$$

PROOF. Choose $\zeta_1 \prec \zeta_2 \in \mathcal{C}_0^\infty(\Omega)$ such that $\zeta_1(x) = 1$ for all $x \in U$. We will sometimes simplify notation by writing $T^s = \zeta_2\Lambda^s\zeta_1$. If $\varphi \in \mathcal{C}_0^\infty(U)$, we have

$\zeta_1 X_j[\varphi] = X_j[\varphi]$ and so

$$
\begin{aligned}
\frac{\eta}{4}\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 &= \frac{\eta}{4}\sum_{j=1}^{p} \big|\big| \Lambda^s X_j[\varphi] \big|\big|_0^2 \\
&\leq \frac{\eta}{2}\sum_{j=1}^{p} \big|\big| \zeta_2\Lambda^s\zeta_1 X_j\varphi \big|\big|_0^2 + \frac{\eta}{2}\sum_{j=1}^{p} \big|\big| (1-\zeta_2)\Lambda^s\zeta_1 X_j\varphi \big|\big|_0^2 \\
&\leq \eta\sum_{j=1}^{p} \big|\big| X_j T^s[\varphi] \big|\big|_0^2 + \eta\sum_{j=1}^{p} \big|\big| \,[T^s,\, X_j]\,[\varphi] \big|\big|_0^2 \\
&\qquad + \frac{\eta}{2}\sum_{j=1}^{p} \big|\big| (1-\zeta_2)\Lambda^s\zeta_1 X_j\varphi \big|\big|_0^2 \\
&= I + II + III.
\end{aligned}
$$

To estimate the terms II and III, note that $[T^s,\, X_j]$ is a pseudodifferential operator of order s, and so $\big|\big|\, [T^s,\, X_j]\,[\varphi] \big|\big|_0 \leq \big|\big|\, [T^s,\, X_j]\,\Lambda^{-s} \big|\big| \big|\big|\, \varphi \big|\big|_s$. Also, since ζ_1 and $(1-\zeta_2)$ have disjoint supports, the operator $(1-\zeta_2)\Lambda^s\zeta_1 X_j$ is infinitely smoothing, and hence $\big|\big| (1-\zeta_2)\Lambda^s\zeta_1 X_j\varphi \big|\big|_0 \leq \big|\big| (1-\zeta_2)\Lambda^s\zeta_1 X_j\Lambda^{-s} \big|\big| \big|\big| \varphi \big|\big|_s$. Thus we have

$$
II + III \leq C_s \big|\big| \varphi \big|\big|_s^2.
$$

To deal with the main term $I = \eta\sum_{j=1}^{p} \big|\big| X_j T^s[\varphi] \big|\big|_0^2$, note that $T^s\varphi \in \mathcal{C}_0^\infty(\Omega)$, and so we can apply Lemma 4.2 to obtain

$$
\begin{aligned}
\eta\sum_{j=1}^{p} \big|\big| X_j T^s[\varphi] \big|\big|_0^2 &\leq \Re e\Big(\mathcal{L}\big[T^s[\varphi]\big], T^s[\varphi]\Big) + C_1 \big|\big| T^s[\varphi] \big|\big|_0^2 \\
&\leq \Re e\Big(\mathcal{L}\big[T^s[\varphi]\big], T^s[\varphi]\Big) + C_1 \big|\big| T^s\Lambda^{-s} \big|\big|^2 \big|\big| \varphi \big|\big|_s^2
\end{aligned}
$$

But now we use Lemma 3.10 to conclude that for any $\delta > 0$,

$$
\begin{aligned}
I &\leq \Re e\Big(T^s\mathcal{L}[\varphi], T^s[\varphi]\Big) + \delta\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big| \varphi \big|\big|_s^2 \\
&= \Re e\Big(\zeta_2\Lambda^s\mathcal{L}[\varphi], \zeta_2\Lambda^s[\varphi]\Big) + \delta\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big| \varphi \big|\big|_s^2 \\
&= \Re e\Big(\Lambda^s\mathcal{L}[\varphi], \zeta_2^2\Lambda^s[\varphi]\Big) + \delta\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big| \varphi \big|\big|_s^2 \\
&= \Re e\Big(\Lambda^s\mathcal{L}[\varphi], \Lambda^s[\varphi]\Big) + \Re e\Big(\Lambda^s\mathcal{L}[\varphi], (\zeta_2^2-1)\Lambda^s\zeta_1[\varphi]\Big) \\
&\qquad + \delta\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big| \varphi \big|\big|_s^2 \\
&= \Re e\Big(\mathcal{L}[\varphi], \varphi\Big)_s + \Re e\Big(\Lambda^s\mathcal{L}[\varphi], (\zeta_2^2-1)\Lambda^s\zeta_1[\varphi]\Big) \\
&\qquad + \delta\sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big| \varphi \big|\big|_s^2.
\end{aligned}
$$

Again using the fact that the functions ζ_1 and $1-\zeta_2$ have disjoint supports, it follows that the operator $(\zeta_2^2-1)\Lambda^s\zeta_1$ is infinitely smoothing, so

$$\left|\Re e\Big(\Lambda^s\mathcal{L}[\varphi],(\zeta_2^2-1)\Lambda^s\zeta_1[\varphi]\Big)\right| \le \big|\big|\,\Lambda^{-s}\zeta_1\Lambda^s(\zeta_2^2-1)\Lambda^s\mathcal{L}\Lambda^{-s}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_s^2$$

so that

$$\eta\sum_{j=1}^{p}\big|\big|\,X_jT^s[\varphi]\,\big|\big|_0^2 \le \Re e\Big(\mathcal{L}[\varphi],\varphi\Big)_s+\delta\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_s^2+C_s(1+\delta^{-1})\big|\big|\,\varphi\,\big|\big|_s^2.$$

Putting the estimates for I, II, and III together, we see that

$$\frac{\eta}{4}\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_s^2 \le \Re e\Big(\mathcal{L}[\varphi],\varphi\Big)_s+\delta\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_s^2+C_s(1+\delta^{-1})\big|\big|\,\varphi\,\big|\big|_s^2.$$

Thus if we let $\delta=\frac{\eta}{8}$ we obtain the estimate in (3.20), and this completes the proof. □

Corollary 3.13. *Suppose that Y_0 satisfies the condition (2.2). Let $U \Subset \Omega$ be a relatively compact subset. Then for every $s\in\mathbb{R}$ there is an allowable constant C_s so that for all $\varphi\in\mathcal{C}_0^\infty(U)$*

$$\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_s^2 \le C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_s^2+\big|\big|\,\varphi\,\big|\big|_s^2\Big]. \tag{3.19}$$

Our next result shows that we can obtain results for more general pseudodifferential operator A^1 of order 1 which are analogous to the estimate in (3.19). This will be used in the discussion of subelliptic multipliers in Section 4.2.

Lemma 3.14. *Suppose that A^1 is a properly supported pseudodifferential operator of order 1, and suppose there are constants $0<a\le 1$ and $C<\infty$ so that for all $\varphi\in\mathcal{C}_0^\infty(\Omega)$*

$$\big|\big|\,A^1[\varphi]\,\big|\big|_{a-1} \le C\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0+\big|\big|\,\varphi\,\big|\big|_0\Big]. \tag{3.20}$$

Let $U\Subset\Omega$ be a relatively compact subset. Then for every $s\in\mathbb{R}$ there is an allowable constant C_s so that for all $\varphi\in\mathcal{C}_0^\infty(U)$

$$\big|\big|\,A^1[\varphi]\,\big|\big|_{a-1+s} \le C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_s+\big|\big|\,\varphi\,\big|\big|_s\Big]. \tag{3.21}$$

Proof. Choose $\zeta_1\prec\zeta_2\prec\zeta_3\in\mathcal{C}_0^\infty(\Omega)$ with $\zeta_1\equiv 1$ on U and $|\zeta_j(x)|\le 1$ for all $x\in\Omega$. Then

$$\begin{aligned}
\big|\big|\,A^1[\varphi]\,\big|\big|_{s+a-1} &= \big|\big|\,\Lambda^{s+a-1}A^1\zeta_1[\varphi]\,\big|\big|_0\\
&\le \big|\big|\,\Lambda^{s+a-1}\zeta_2A^1\zeta_1[\varphi]\,\big|\big|_0+\big|\big|\,\Lambda^{s+a-1}(1-\zeta_2)A^1\zeta_1[\varphi]\,\big|\big|_0\\
&\le \big|\big|\,\Lambda^{a-1}\zeta_3\Lambda^s\zeta_2A^1\zeta_1[\varphi]\,\big|\big|_0+\big|\big|\,\Lambda^{a-1}(1-\zeta_3)\Lambda^s\zeta_2A^1\zeta_1[\varphi]\,\big|\big|_0\\
&\qquad+\big|\big|\,\Lambda^{s+a-1}(1-\zeta_2)A^1\zeta_1[\varphi]\,\big|\big|_0\\
&= I+II+III.
\end{aligned}$$

Because of the disjoint supports, the two operators $\Lambda^{a-1}(1-\zeta_3)\Lambda^s\zeta_2A^1\zeta_1$ and $\Lambda^{s+a-1}(1-\zeta_2)A^1\zeta_1$ are infinitely smoothing, and so

$$\begin{aligned}
II &\le \big|\big|\,(1-\zeta_3)\Lambda^{s+a-1}\zeta_2A^1\zeta_1\Lambda^{-s}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_s\\
III &\le \big|\big|\,\Lambda^{s+a-1}(1-\zeta_2)A^1\zeta_1\Lambda^{-s}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_s.
\end{aligned}$$

To deal with the term $I = \big|\big|\, \Lambda^{a-1}\zeta_3\Lambda^s\zeta_2 A^1\zeta_1[\varphi]\,\big|\big|_0$ write

$$\begin{aligned} I &\leq \big|\big|\, \Lambda^{a-1}A^1\zeta_3\Lambda^s\zeta_2\zeta_1[\varphi]\,\big|\big|_0 + \big|\big|\, \Lambda^{a-1}\left[\zeta_3\Lambda^s\zeta_2,\, A^1\right]\zeta_1[\varphi]\,\big|\big|_0 \\ &= I_a + I_b. \end{aligned}$$

Since A^1 is of order 1, the operator $\Lambda^{a-1}\left[\zeta_3\Lambda^s\zeta_2,\, A^1\right]\zeta_1$ has order $s + a - 1 \leq s$. Hence

$$I_b \leq \big|\big|\, \left[\zeta_3\Lambda^{s+a-1}\zeta_2,\, A^1\right]\zeta_1\Lambda^{-s}\,\big|\big|\, \big|\big|\, \varphi\,\big|\big|_s.$$

Finally, using the hypothesis (3.20) on A^1 we have

$$\begin{aligned} I_a &= \big|\big|\, \Lambda^{a-1}A^1\zeta_3\Lambda^s\zeta_2\zeta_1[\varphi]\,\big|\big|_0 = \big|\big|\, A^1\zeta_3\Lambda^s\zeta_2\zeta_1[\varphi]\,\big|\big|_{a-1} \\ &\leq C\Big[\big|\big|\, \mathcal{L}[\zeta_3\Lambda^s\zeta_2\varphi]\,\big|\big|_s + \big|\big|\, \zeta_3\Lambda^s\zeta_2[\varphi]\,\big|\big|_s\Big] \\ &\leq C\Big[\big|\big|\, \mathcal{L}[\zeta_3\Lambda^s\zeta_2\varphi]\,\big|\big|_s + \big|\big|\, \zeta_3\Lambda^s\zeta_2\Lambda^{-s}\,\big|\big|\, \big|\big|\, [\varphi]\,\big|\big|_s\Big] \end{aligned}$$

But using Lemma 3.7 with $A^s = \zeta_3\Lambda^s\zeta_2$, and then Corollary 3.13 we have

$$\begin{aligned} \big|\big|\, \mathcal{L}[\zeta_3\Lambda^s\zeta_2\varphi]\,\big|\big|_s &\leq C_s\Big[\big|\big|\, \mathcal{L}[\varphi]\,\big|\big|_s + \sum_{j=1}^p \big|\big|\, X_j[\varphi]\,\big|\big|_s + \big|\big|\, \varphi\,\big|\big|_s\Big] \\ &\leq C_s\Big[\big|\big|\, \mathcal{L}[\varphi]\,\big|\big|_s + \big|\big|\, \varphi\,\big|\big|_s\Big]. \end{aligned}$$

Putting all the estimates together yields the estimate in equation (3.21) and completes the proof. □

4. Estimates for smooth functions

Our objective in this section is to establish a preliminary version of Theorem 2.2 where the distribution u is taken to be a compactly supported smooth function φ, and cut-off functions ζ_0 and ζ_1 are not needed. In Theorem 2.2, one of the main results is that the compactly supported distribution $\zeta_1 u$ belongs to the space $H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$. In this preliminary version, φ is assumed to be smooth with compact support, and the corresponding result is often called an *a priori* sub-elliptic estimate.

THEOREM 4.1. *Suppose that hypotheses (H-1) and (H-2) are satisfied. Set* $\frac{1}{2}\epsilon = 2\cdot 4^{-m}$. *Let* $U \Subset \Omega$ *be a relatively compact open set. Then for all* $s \in \mathbb{R}$ *there is an allowable constant* C_s *such that for all* $\varphi \in \mathcal{C}_0^\infty(U)$ *we have*

$$\big|\big|\, \varphi\,\big|\big|_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\, \mathcal{L}[\varphi]\,\big|\big|_s + \big|\big|\, \varphi\,\big|\big|_s\Big]. \tag{4.1}$$

4.1. Introduction.

Despite its apparent simplicity, the proof of the estimate in (4.1) is quite intricate. Before starting on the details we begin with an overview of the argument. Note that if $\varphi \in \mathcal{C}_0^\infty(U)$ then

$$\big|\big|\, \varphi\,\big|\big|_{s+\frac{1}{2}\epsilon} \approx \sum_{k=1}^n \big|\big|\, \partial_{x_k}[\varphi]\,\big|\big|_{s+\frac{1}{2}\epsilon-1}.$$

Thus to establish the inequality (4.1) it suffices to show that for all $\varphi \in \mathcal{C}_0^\infty(U)$,

$$\big|\big|\, \partial_{x_k}[\varphi]\,\big|\big|_{s+\frac{1}{2}\epsilon-1} \leq C_s\Big[\big|\big|\, \mathcal{L}[\varphi]\,\big|\big|_s + \big|\big|\, \varphi\,\big|\big|_s\Big]$$

for $1 \leq k \leq n$. A fruitful approach to this problem is to consider the collection $\mathcal{M}(s, \frac{1}{2}\epsilon, U)$ of all real vector fields Y on Ω such that for all $\varphi \in \mathcal{C}_0^\infty(U)$,

$$\big|\big| Y[\varphi] \big|\big|_{s+\frac{1}{2}\epsilon-1} \leq C_s\Big[\big|\big| \mathcal{L}[\varphi] \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big].$$

for some allowable constant C_s. The collection of spaces of sub-elliptic multipliers $\{\mathcal{M}(s, \frac{1}{2}\epsilon, U)\}$ of order $\frac{1}{2}\epsilon$ turns out to have unexpected Lie-theoretic properties. In particular, the commutator of two vector fields, each of which is a sub-elliptic multiplier, is again a subelliptic multiplier but of smaller order. The objective, of course, is to show that for some $\frac{1}{2}\epsilon > 0$, each operator $\partial_{x_k} \in \mathcal{M}(s, \frac{1}{2}\epsilon, U)$, and these Lie-theoretic properties allow one to use the hypothesis (H-1). There are three critical steps in the argument.

(a) We show that the vector fields $\{X_1, \ldots, X_p\}$ belong to $\mathcal{M}(0, 1, U)$ and the vector field X_0 belongs to $\mathcal{M}(0, \frac{1}{2}, U)$.

(b) We show that if $Y \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$ then $[X_0, Y] \in \mathcal{M}(0, \frac{1}{4}\frac{1}{2}\epsilon, U)$ and $[X_j, Y] \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$ for $1 \leq j \leq p$.

(c) We show that if $Y \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$, then $Y \in \mathcal{M}(s, \frac{1}{2}\epsilon, U)$ for every $s \in \mathbb{R}$.

Suppose these facts are established. It then follows by induction from (a) and (b) that any iterated commutator of length k of the vector fields $\{X_0, X_1, \ldots, X_p\}$ belongs to the space $\mathcal{M}(0, 2 \cdot 4^{-k}, U)$. Then (c) shows that every such commutator belongs to $\mathcal{M}(s, 2 \cdot 4^{-k}, U)$. On the other hand, hypothesis (H-1) says that every partial derivative ∂_{x_k} can be written as a linear combination of commutators of length at most m, and hence $\partial_{x_k} \in \mathcal{M}(s, 2 \cdot 4^{-m}, U)$. This then gives the desired estimate with $\frac{1}{2}\epsilon = 2 \cdot 4^{-m}$.

4.2. The space of sub-elliptic multipliers. The Sobolev norms $\big|\big| \varphi \big|\big|_s$ are defined through the use of the pseudodifferential operator Λ^s. For this reason, in the definition of the space of sub-elliptic multipliers, it is convenient to allow appropriate pseudodifferential operators of order 1 rather than just vector fields.

Definition 4.2. *For every relatively compact open subset $U \Subset \Omega$ and every $0 < a \leq 1$, let $\mathcal{M}(s, a, U)$ denote the space of all properly supported pseudodifferential operators $A \in OP^1(\mathbb{R}^n)$ with the property that there exists an allowable constant $C_s < \infty$ so that for all $\varphi \in \mathcal{C}_0^\infty(U)$,*

$$\big|\big| A[\varphi] \big|\big|_{s+a-1} \leq C\Big[\big|\big| \mathcal{L}[\varphi] \big|\big|_s + \big|\big| \varphi \big|\big|_s\Big]. \tag{4.2}$$

Let $\big|\big| A \big|\big|_{s,a,U}$ denote the optimal constant C_s so that

$$\big|\big| A \big|\big|_{s,a,U} = \sup\Big\{\big|\big| A[\varphi] \big|\big|_{s+a-1} \Big| \big|\big| \mathcal{L}[\varphi] \big|\big|_s + \big|\big| \varphi \big|\big|_s \leq 1\Big\}. \tag{4.3}$$

We begin by establishing two basic properties of $\mathcal{M}(s, a, U)$. The first deals with compositions.

Proposition 4.3. *If $A \in OP^{1-a}(\mathbb{R}^n) \subset OP^1(\mathbb{R}^n)$ is properly supported, then $A \in \mathcal{M}(s, a, U)$ for every $s \in \mathbb{R}$, and*

$$\big|\big| A \big|\big|_{s,a,U} \leq \big|\big| \Lambda^{s+a-1} A \Lambda^{-s} \big|\big|. \tag{4.4}$$

If $B \in OP^{a-b}(\mathbb{R}^n)$ is properly supported and if $A \in \mathcal{M}(s,a,U)$, the composition $BA \in \mathcal{M}(s,b,U)$, and

$$\big|\big|\, BA \,\big|\big|_{s,b,U} \leq \big|\big|\, \Lambda^{s+b-1} B \Lambda^{-s-a+1} \,\big|\big| \,\big|\big|\, A \,\big|\big|_{s,a,U}. \tag{4.5}$$

In particular, the space $\mathcal{M}(s,a,U)$ is a module over the algebra of properly supported pseudodifferential operators of order zero.

Proof. If $A \in OP^{1-a}_p(\mathbb{R}^n)$, then $\Lambda^{s+a-1} A \Lambda^{-s}$ is a pseudodifferential operator of order zero. We have

$$\big|\big|\, A[\varphi] \,\big|\big|_{s+a-1} = \big|\big|\, \Lambda^{s+a-1} A \Lambda^{-s} \Lambda^{s}[\varphi] \,\big|\big|_0 \leq \big|\big|\, \Lambda^{s+a-1} A \Lambda^{-s} \,\big|\big| \,\big|\big|\, \varphi \,\big|\big|_s,$$

and this gives inequality (4.4). Next, since B is properly supported, the same is true of BA. Note that $\Lambda^{s+b-1} B \Lambda^{-s-a+1}$ is a pseudodifferential operator of order zero. Let $\varphi \in \mathcal{C}^\infty_0(U)$. Then

$$\begin{aligned} \big|\big|\, BA[\varphi] \,\big|\big|_{s+b-1} &= \big|\big|\, \Lambda^{s+b-1} BA[\varphi] \,\big|\big|_0 \\ &\leq \big|\big|\, \Lambda^{s+b-1} B \Lambda^{-s-a+1} \,\big|\big| \,\big|\big|\, A[\varphi] \,\big|\big|_{s+a-1} \\ &\leq \big|\big|\, \Lambda^{s+b-1} B \Lambda^{-s-a+1} \,\big|\big| \,\big|\big|\, A \,\big|\big|_{s,a,U} \big[\big|\big|\, \mathcal{L}[\varphi \,\big|\big|_s + \big|\big|\, \varphi \,\big|\big|_s\big], \end{aligned}$$

and this gives inequality (4.5). □

The second results shows that for $a \leq \frac{1}{2}$, the space $\mathcal{M}(s,a,U)$ is closed under taking adjoints.

Proposition 4.4. *If $A \in \mathcal{M}(s,a,U)$ and if $a \leq \frac{1}{2}$, then $A^* \in \mathcal{M}(s,a,U)$.*

Proof. The proof is another exercise in moving an operators around an inner product, and keeping track of the commutators. Write $T = \Lambda^{2s+2a-2}$. Then

$$\begin{aligned} \big|\big|\, A^*[\varphi] \,\big|\big|^2_{s+a-1} &= \big(\Lambda^{s+a-1} A^*[\varphi], \Lambda^{s+a-1} A^*[\varphi]\big) = \big(A^*[\varphi], T A^*[\varphi]\big) \\ &= \big(AA^*[\varphi], T[\varphi]\big) + \big(A^*[\varphi], [T,\, A^*]\,[\varphi]\big) \\ &= \big(A^*A[\varphi], T[\varphi]\big) + \big(\,[A,\, A^*]\,[\varphi], T[\varphi]\big) + \big(A^*[\varphi], [T,\, A^*]\,[\varphi]\big) \\ &= \big(A[\varphi], TA[\varphi]\big) + \big(A[\varphi], [A,\, T]\,[\varphi]\big) + \big(\,[A,\, A^*]\,[\varphi], T[\varphi]\big) \\ &\qquad + \big(A^*[\varphi], [T,\, A^*]\,[\varphi]\big) \\ &= \big|\big|\, A[\varphi] \,\big|\big|^2_{s+a-1} + \big(A[\varphi], [A,\, T]\,[\varphi]\big) + \big(\,[A,\, A^*]\,[\varphi], T[\varphi]\big) \\ &\qquad + \big(A^*[\varphi], [T,\, A^*]\,[\varphi]\big) \\ &= \big|\big|\, A[\varphi] \,\big|\big|^2_{s+a-1} + I + II + III. \end{aligned}$$

Now we make estimates of the last three terms. First

$$\begin{aligned} I &= \big|\big(\Lambda^{s+a-1} A[\varphi], \Lambda^{-s-a+1}\,[A,\, T]\,[\varphi]\big)\big| \\ &\leq \frac{1}{2}\big|\big|\, A[\varphi] \,\big|\big|^2_{s+a-1} + \frac{1}{2}\big|\big|\, \Lambda^{-2s-a+1}\,[A,\, T]\,[\varphi] \,\big|\big|^2_s \\ &\leq \frac{1}{2}\big|\big|\, A[\varphi] \,\big|\big|^2_{s+a-1} + \big|\big|\, \Lambda^{-s-a+1}\,[A,\, T]\,\Lambda^{-s} \,\big|\big|^2 \,\big|\big|\, \varphi \,\big|\big|^2_s \end{aligned}$$

since the order of $\Lambda^{-s-a+1}\,[A,\,T]\,\Lambda^{-s}$ is $a-2+o(A)\leq 0$. Next

$$\begin{aligned} II &= \big|\big(\Lambda^{s+2a-2}\,[A,\,A^*]\,[\varphi],\Lambda^{s}[\varphi]\big)\big| \\ &\leq \frac{1}{2}\big|\big|\,\Lambda^{2a-2}\,[A,\,A^*]\,[\varphi]\,\big|\big|_s^2+\frac{1}{2}\big|\big|\,\varphi\,\big|\big|_s^2 \\ &\leq \frac{1}{2}\Big[\big|\big|\,\Lambda^{s+2a-2}\,[A,\,A^*]\,\Lambda^{-s}\,\big|\big|^2+1\Big]\big|\big|\,\varphi\,\big|\big|_s^2 \end{aligned}$$

since the order of $\Lambda^{s+2a-2}\,[A,\,A^*]\,\Lambda^{-s}$ is $2a-3+2\,o(A)\leq 0$. Finally

$$\begin{aligned} \big|\big(A^*[\varphi],[T,\,A^*]\,[\varphi]\big)\big| &= \big(\Lambda^{s+a-1}A^*[\varphi],\Lambda^{-s-a+1}\,[T,\,A^*]\,[\varphi]\big)\big| \\ &\leq \frac{1}{2}\big|\big|\,A^*[\varphi]\,\big|\big|_{s+a-1}^2+\frac{1}{2}\big|\big|\,\Lambda^{-2s-a+1}\,[T,\,A^*]\,[\varphi]\,\big|\big|_s^2 \\ &\leq \frac{1}{2}\big|\big|\,A^*[\varphi]\,\big|\big|_{s+a-1}^2+\big|\big|\,\Lambda^{-s-a+1}\,[T,\,A^*]\,\Lambda^{-s}\,\big|\big|^2\,\big|\big|\,\varphi\,\big|\big|_s^2 \end{aligned}$$

since the order of $\Lambda^{-s-a+1}\,[T,\,A^*]\,\Lambda^{-s}$ is $a-2+o(A^*)\leq 0$. Putting these inequalities together, we get

$$\frac{1}{2}\big|\big|\,A^*[\varphi]\,\big|\big|_{s+a-1}^2\leq\frac{3}{2}\big|\big|\,A[\varphi]\,\big|\big|_{s+a-1}^2+C_s^2\,\big|\big|\,\varphi\,\big|\big|_s^2$$

where

$$\begin{aligned} C_s^2 = \big|\big|\,\Lambda^{-s-a+1}\,[A,\,T]\,\Lambda^{-s}\,\big|\big|^2+\big|\big|\,\Lambda^{-s-a+1}\,[T,\,A^*]\,\Lambda^{-s}\,\big|\big|^2 \\ +\frac{1}{2}\Big[\big|\big|\,\Lambda^{s+2a-2}\,[A,\,A^*]\,\Lambda^{-s}\,\big|\big|^2+1\Big] \end{aligned}$$

is an allowable constant. □

4.3. The main theorem on sub-elliptic multipliers. As outlined in Section 4.1, the proof of Theorem 4.1 follows from the following statements about the spaces of subelliptic multipliers $\mathcal{M}(s,\frac{1}{2}\epsilon,U)$.

LEMMA 4.5. *Let $U\Subset\Omega$ be a relatively compact open subset. Choose $\zeta_1\prec\zeta_2\in\mathcal{C}_0^\infty(\Omega)$ such that $\zeta_1(x)=1$ for all $x\in U$, and $|\zeta_j(x)|\leq 1$ for all $x\in\Omega$ and $j=1,\,2$. Then:*

(1) *For each $s\in\mathbb{R}$ there is an allowable constant $C_s<\infty$ so that for all $\varphi\in\mathcal{C}_0^\infty(U)$*

$$\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_s^2\leq C_s\,\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_s^2+\big|\big|\,\varphi\,\big|\big|_s^2\Big]. \tag{4.6}$$

Thus the vector fields $X_1,\ldots,X_p\in\mathcal{M}(s,1,U)$ for all $s\in\mathbb{R}$.

(2) *For each $s\in\mathbb{R}$ there is an allowable constant $C_s<\infty$ such that for all $\varphi\in\mathcal{C}_0^\infty(U)$*

$$\big|\big|\,X_0[\varphi]\,\big|\big|_s\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_{s+\frac{1}{2}}+\big|\big|\,\varphi\,\big|\big|_{s+\frac{1}{2}}\Big]. \tag{4.7}$$

Thus the vector field $X_0\in\mathcal{M}(s,\frac{1}{2},U)$ for all $s\in\mathbb{R}$.

(3) *If $A\in\mathcal{M}(0,a,U)$ then $A\in\mathcal{M}(s,a,U)$ for all $s\in\mathbb{R}$, and $\big|\big|\,A\,\big|\big|_{s,a,U}\leq C_{s,a,U}\big|\big|\,A\,\big|\big|_{0,a,U}$.*

(4) Suppose that Y_1 *and* Y_2 *are smooth real vector fields on* Ω *and that* $Y_j \in \mathcal{M}(0, a_j, U)$. *Then the commutator* $[Y_1, Y_2] \in \mathcal{M}(0, a, U)$ *provided that* $0 < a \leq \min\left\{\frac{1}{2}a_1,\ \frac{1}{2}a_2,\ \frac{1}{2}(a_1 + a_2 - 1)\right\}$.

(5) If a vector field $Y \in \mathcal{M}(0, a, U)$, *then for* $1 \leq j \leq p$, *each commutator* $[X_j, Y] \in \mathcal{M}(0, \frac{a}{2}, U)$.

(6) If a vector field $Y \in \mathcal{M}(0, a, U)$, *then* $[X_0, Y] \in \mathcal{M}(0, \frac{a}{4}, U)$.

Note that we have already established conclusion (1) in Corollary 3.13, and conclusion (3) in Lemma 3.14.

PROOF OF (2). If $\varphi \in \mathcal{C}_0^\infty(U)$ we have

$$\big|\big| X_0[\varphi] \big|\big|_s^2 = \Big(X_0[\varphi], \Lambda^{2s} X_0[\varphi]\Big) = \Big(X_0[\varphi], \zeta_2 \Lambda^{2s} X_0[\zeta_1 \varphi]\Big) = \Big(X_0[\varphi], A^{2s+1}[\varphi]\Big)$$

where $A^{2s+1} = \zeta_2 \Lambda^{2s} X_0 \zeta_1$. Since we have $X_0 = \mathcal{L} + \sum_{j=1}^p X_j^2 - iY_0 - c$, this shows that

$$\begin{aligned} \big|\big| X_0[\varphi] \big|\big|_s^2 &= \Big(\mathcal{L}[\varphi], A^{2s+1}[\varphi]\Big) + \sum_{j=1}^p \Big(X_j^2[\varphi], A^{2s+1}[\varphi]\Big) \\ &\qquad - i\Big(Y_0[\varphi], A^{2s+1}[\varphi]\Big) - \Big(c\varphi, A^{2s+1}[\varphi]\Big) \\ &= I + II + III + IV. \end{aligned}$$

We deal separately with each of these four terms. First

$$\begin{aligned} |I| = \Big|\Big(\mathcal{L}[\varphi], A^{2s+1}[\varphi]\Big)\Big| &= \Big|\Big(\Lambda^{s+\frac{1}{2}} \mathcal{L}[\varphi], \Lambda^{-s-\frac{1}{2}} A^{2s+1}[\varphi]\Big)\Big| \\ &\leq \big|\big| \Lambda^{-s-\frac{1}{2}} A^{2s+1} \Lambda^{-s-\frac{1}{2}} \big|\big|\ \big|\big| \mathcal{L}[\varphi] \big|\big|_{s+\frac{1}{2}} \big|\big| \varphi \big|\big|_{s+\frac{1}{2}} \\ &\leq C_s \Big[\big|\big| \mathcal{L}[\varphi] \big|\big|_{s+\frac{1}{2}}^2 + \big|\big| \varphi \big|\big|_{s+\frac{1}{2}}^2\Big]. \end{aligned}$$

Next we deal with the term II. We have

$$\begin{aligned} |II| &\leq \sum_{j=1}^p \Big|\Big(X_j^2[\varphi], A^{2s+1}[\varphi]\Big)\Big| = \sum_{j=1}^p \Big|\Big(X_j[\varphi], X_j^* A^{2s+1}[\varphi]\Big)\Big| \\ &\leq \sum_{j=1}^p \Big|\Big(X_j[\varphi], A^{2s+1} X_j^*[\varphi]\Big)\Big| + \sum_{j=1}^p \Big|\Big(X_j[\varphi], \big[X_j^*,\ A^{2s+1}\big]\, [\varphi]\Big)\Big| \\ &\leq \sum_{j=1}^p \big|\big| \Lambda^{-s-\frac{1}{2}} A^{2s+1} \Lambda^{-s-\frac{1}{2}} \big|\big|\ \big|\big| X_j[\varphi] \big|\big|_{s+\frac{1}{2}} \big|\big| X_j^*[\varphi] \big|\big|_{s+\frac{1}{2}} \\ &\qquad + \sum_{j=1}^p \big|\big| \Lambda^{-s-\frac{1}{2}} \big[X_j^*,\ A^{2s+1}\big] \Lambda^{-s-\frac{1}{2}} \big|\big|\ \big|\big| X_j[\varphi] \big|\big|_{s+\frac{1}{2}} \big|\big| \varphi \big|\big|_{s+\frac{1}{2}} \\ &\leq C_s \Big[\sum_{j=1}^p \big|\big| X_j[\varphi] \big|\big|_{s+\frac{1}{2}} + \sum_{j=1}^p \big|\big| X_j^*[\varphi] \big|\big|_{s+\frac{1}{2}} + \big|\big| \varphi \big|\big|_{s+\frac{1}{2}}\Big]. \end{aligned}$$

Each vector field $X_j \in \mathcal{M}(s + \frac{1}{2}, 1, U)$, and since $X_j^* = -X_j + \operatorname{div}(X_j)$ it follows that $X_j^* \in \mathcal{M}(s + \frac{1}{2}, 1, U)$. Thus $|II| \leq C_s\Big[\big|\big| \mathcal{L}[\varphi] \big|\big|_{s+\frac{1}{2}}^2 + \big|\big| \varphi \big|\big|_{s+\frac{1}{2}}^2\Big]$.

The term IV is easy to deal with since

$$|IV| = \Big|\Big(\Lambda^{s+\frac{1}{2}}\varphi, \Lambda^{-s-\frac{1}{2}}\bar{c}A^{2s+1}[\varphi]\Big)\Big| \leq \big|\big|\,\Lambda^{-s-\frac{1}{2}}\bar{c}A^{2s+1}\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|^2_{s+\frac{1}{2}}.$$

To deal with the term III, write

$$\begin{aligned}|III| &= \Big|\Big(Y_0[\varphi], A^{2s+1}[\varphi]\Big)\Big| = \Big|\Big(\Lambda^{s+\frac{1}{2}}\zeta_1 Y_0[\varphi], \Lambda^{s-\frac{1}{2}}X_0[\varphi]\Big)\Big| \\ &\leq \Big|\Big(\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1 Y_0[\varphi], \zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]\Big)\Big| + \Big|\Big((1-\zeta_2)\Lambda^{s+\frac{1}{2}}\zeta_1 Y_0[\varphi], \Lambda^{s-\frac{1}{2}}X_0[\varphi]\Big)\Big| \\ &= III_a + III_b.\end{aligned}$$

Since $(1-\zeta_2)$ and ζ_1 have disjoint supports, the operator $(1-\zeta_2)\Lambda^{s+\frac{1}{2}}\zeta_1 Y_0$ is infinitely smoothing,

$$III_b \leq \big|\big|\,(1-\zeta_2)\Lambda^{s+\frac{1}{2}}\zeta_1 Y_0\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\Lambda^{s-\frac{1}{2}}X_0\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|^2_{s+\frac{1}{2}}.$$

On the other hand, we have

$$\begin{aligned}III_a &\leq \Big|\Big(Y_0[\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\varphi], \zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]\Big)\Big| + \Big|\Big(\Big[\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1,\, Y_0\Big][\varphi], \zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]\Big)\Big| \\ &= III_{a,1} + III_{a,2}.\end{aligned}$$

Since the order of the operators $\Big[\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1,\, Y_0\Big]$ and $\Lambda^{s-\frac{1}{2}}X_0$ is $s+\frac{1}{2}$, we have

$$III_{a,2} \leq \big|\big|\,\Big[\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1,\, Y_0\Big]\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\Lambda^{s-\frac{1}{2}}X_0\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|^2_{s+\frac{1}{2}}.$$

Finally, using the hypothesis on the vector field Y_0 given in equation (2.3) with $\varphi = \zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1[\varphi]$ and $\psi = \zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]$,

$$\begin{aligned}III_{a,1} \leq\; &\big|\big|\,\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\varphi\,\big|\big|_0 + \big|\big|\,\zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]\,\big|\big|_0 \\ &+ \sum_{j=1}^{p}\big|\big|\,X_j[\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\varphi]\,\big|\big|_0 + \sum_{j=1}^{p}\big|\big|\,X_j[\zeta_3\Lambda^{s-\frac{1}{2}}X_0\varphi]\,\big|\big|_0\end{aligned}$$

Now

$$\begin{aligned}\big|\big|\,\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\varphi\,\big|\big|_0 &\leq \big|\big|\,\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_{s+\frac{1}{2}} \\ \big|\big|\,\zeta_3\Lambda^{s-\frac{1}{2}}X_0[\varphi]\,\big|\big|_0 &\leq \big|\big|\,\zeta_3\Lambda^{s-\frac{1}{2}}X_0\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_{s+\frac{1}{2}}\end{aligned}$$

and

$$\begin{aligned}\big|\big|\,X_j(\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1)[\varphi]\,\big|\big|_0 &\leq \big|\big|\,(\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1)X_j[\varphi]\,\big|\big|_0 + \big|\big|\,\Big[X_j,\,(\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1)\Big][\varphi]\,\big|\big|_0 \\ &\leq \big|\big|\,\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,X_j[\varphi]\,\big|\big|_{s+\frac{1}{2}} \\ &\qquad + \big|\big|\,\Big[X_j,\,(\zeta_2\Lambda^{s+\frac{1}{2}}\zeta_1)\Big]\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_{s+\frac{1}{2}}\end{aligned}$$

while

$$\begin{aligned}\big|\big|\,X_j(\zeta_3\Lambda^{s-\frac{1}{2}}X_0)[\varphi]\,\big|\big|_0 &\leq \big|\big|\,(\zeta_3\Lambda^{s-\frac{1}{2}}X_0)X_j[\varphi]\,\big|\big|_0 + \big|\big|\,\Big[X_j,\,(\zeta_3\Lambda^{s-\frac{1}{2}}X_0)\Big][\varphi]\,\big|\big|_0 \\ &\leq \big|\big|\,(\zeta_3\Lambda^{s-\frac{1}{2}}\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,X_j[\varphi]\,\big|\big|_{s+\frac{1}{2}} \\ &\qquad + \big|\big|\,\Big[X_j,\,(\zeta_3\Lambda^{s-\frac{1}{2}}X_0)\Big]\Lambda^{-s-\frac{1}{2}}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_{s+\frac{1}{2}}.\end{aligned}$$

Since we already know that each $X_j \in \mathcal{M}(s+\frac{1}{2}, 1, U)$, the completes the proof. □

PROOF OF (4). Suppose that $Y_j \in \mathcal{M}(0, a_j, U)$ for $j = 1, 2$, and suppose that $0 < \frac{1}{2}\epsilon \le \min\left\{\frac{1}{2}a_1,\ \frac{1}{2}a_2,\ \frac{1}{2}(a_1+a_2-1)\right\}$. For $\varphi \in \mathcal{C}_0^\infty(U)$,

$$\left|\left|\ [Y_1,\, Y_2]\, [\varphi]\, \right|\right|^2_{\frac{1}{2}\epsilon-1} \le \left|\Big(Y_1Y_2[\varphi], \zeta_2\Lambda^{2\frac{1}{2}\epsilon-2}\, [Y_1,\, Y_2]\, \zeta_1[\varphi]\Big)\right| + \left|\Big(Y_2Y_1[\varphi], \zeta_2\Lambda^{2\frac{1}{2}\epsilon-2}\, [Y_1,\, Y_2]\, \zeta_1[\varphi]\Big)\right|.$$

We deal with both terms the same way. Write $\zeta_2\Lambda^{2\frac{1}{2}\epsilon-2}\,[Y_1,\, Y_2]\,\zeta_1 = A^{2\frac{1}{2}\epsilon-1}$. Then

$$\begin{aligned}
\left|\Big(Y_1Y_2[\varphi], A^{2\frac{1}{2}\epsilon-1}[\varphi]\Big)\right| &= \left|\Big(Y_2[\varphi], Y_1^* A^{2\frac{1}{2}\epsilon-1}[\varphi]\Big)\right| \\
&\le \left|\Big(Y_2[\varphi], A^{2\frac{1}{2}\epsilon-1}Y_1^*[\varphi]\Big)\right| \\
&\qquad + \left|\Big(Y_2[\varphi], \Big[Y_1^*,\ A^{2\frac{1}{2}\epsilon-1}\Big]\,[\varphi]\Big)\right| \\
&\le \left|\left|\, \Lambda^{-a_2+1}A^{2\frac{1}{2}\epsilon-1}\Lambda^{-a_1+1}\, \right|\right|\, \left|\left|\, Y_1[\varphi]\, \right|\right|_{a_1-1} \left|\left|\, Y_2[\varphi]\, \right|\right|_{a_2-1} \\
&\qquad + \left|\left|\, \Lambda^{-a_2+1}, \Big[Y_1^*,\ A^{2\frac{1}{2}\epsilon-1}\Big]\, \right|\right|\, \left|\left|\, Y_2[\varphi]\, \right|\right|_{a_2-1} \left|\left|\, \varphi\, \right|\right|_0 \\
&\le C\Big[\left|\left|\, \mathcal{L}[\varphi]\, \right|\right|_0^2 + \left|\left|\, \varphi\, \right|\right|_0^2\Big]
\end{aligned}$$

because of our assumptions on $\frac{1}{2}\epsilon$. Thus $[Y_1,\, Y_2] \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$. □

PROOF OF (5). Suppose that a smooth real vector field $Y \in \mathcal{M}(0, a, U)$. Each vector field $X_j \in \mathcal{M}(0, 1, U)$. Using part (4), it follows that $[X_j,\, Y] \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$ if $\frac{1}{2}\epsilon \le \min\left\{\frac{1}{2}, \frac{a}{2}, \frac{1}{2}(1+a-1)\right\} = \frac{a}{2}$, and this is what we want to show. □

PROOF OF (6). Let Y be a smooth real vector field on Ω and suppose for all $\varphi \in \mathcal{C}_0^\infty(U)$ we have

$$\left|\left|\, Y[\varphi]\, \right|\right|_{\frac{1}{2}\epsilon-1} \le C\Big[\left|\left|\, \mathcal{L}[\varphi]\, \right|\right|_0 + \left|\left|\, \varphi\, \right|\right|_0\Big].$$

We must show that

$$\left|\left|\ [X_0,\, Y]\, [\varphi]\, \right|\right|_{s-1} \le C_s\Big[\left|\left|\, \mathcal{L}[\varphi]\, \right|\right|_0 + \left|\left|\, \varphi\, \right|\right|_0\Big]$$

provided that $s \le \min\left\{\frac{\frac{1}{2}\epsilon}{4}, \frac{1}{2}\right\}$. Set $T^{2s-1} = \zeta_2\Lambda^{2s-2}\,[X_0,\, Y]\,\zeta_1$. Then

$$\begin{aligned}
\left|\left|\ [X_0,\, Y]\, [\varphi]\, \right|\right|^2_{s-1} &= \Big(\, [X_0,\, Y]\, [\varphi], \Lambda^{2s-2}\, [X_0,\, Y]\, [\varphi]\Big) \\
&= \Big(\, [X_0,\, Y]\, [\varphi], \big(\zeta_2\Lambda^{2s-2}\, [X_0,\, Y]\, \zeta_1\big)[\varphi]\Big) \\
&= \Big(\, [X_0,\, Y]\, [\varphi], T^{2s-1}[\varphi]\Big) \\
&= \Big(X_0Y[\varphi], T^{2s-1}[\varphi]\Big) - \Big(YX_0[\varphi], T^{2s-1}[\varphi]\Big) \\
&= I - II.
\end{aligned}$$

These two terms are handled slightly differently. To deal with term II, write $X_0 = \mathcal{L} + \sum_{j=1}^{p} X_j^2 - iY_0 - c$. Then

$$\begin{aligned} II &= \Big(Y\mathcal{L}[\varphi], T^{2s-1}[\varphi]\Big) + \sum_{j=1}^{p} \Big(YX_j^2[\varphi], T^{2s-1}[\varphi]\Big) \\ &\qquad - i\Big(YY_0[\varphi], T^{2s-1}[\varphi]\Big) - \Big(Y[c\varphi], T^{2s-1}[\varphi]\Big) \\ &= II_a + II_b + II_c + II_d. \end{aligned}$$

Now

$$\begin{aligned} \big|II_a\big| &\leq \Big|\Big(\mathcal{L}[\varphi], T^{2s-1}Y[\varphi]\Big)\Big| + \Big|\Big(\mathcal{L}[\varphi].\, \big[Y,\, T^{2s-1}\big]\,[\varphi]\Big)\Big| \\ &\leq \big|\big|\, T^{2s-1}\Lambda^{-2s+1}\,\big|\big|\,\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1} \\ &\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2 + \big|\big|\,\varphi\,\big|\big|_0^2\Big] \end{aligned}$$

since $2s-1 < \frac{1}{2}\epsilon - 1$. Similarly

$$\begin{aligned} \big|II_c\big| &\leq \Big|\Big(Y_0[\varphi], T^{2s-1}Y[\varphi]\Big)\Big| + \Big|\Big(Y_0[\varphi],\, \big[Y,\, T^{2s-1}\big]\,[\varphi]\Big)\Big| \\ &\leq \big|\big|\, T^{2s-1}\Lambda^{-2s+1}\,\big|\big|\,\big|\big|\,Y_0[\varphi]\,\big|\big|_0\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1} \\ &\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2 + \big|\big|\,\varphi\,\big|\big|_0^2\Big] \end{aligned}$$

because of our assumptions on the vector field Y_0. We also have

$$\begin{aligned} \big|II_d\big| &\leq \Big|\Big(Y[c]\varphi, T^{2s-1}[\varphi]\Big)\Big| + \Big|\Big(c\,Y[\varphi], T^{2s-1}[\varphi]\Big)\Big| \\ &\leq \big|\big|\,\overline{Y[c]}T^{2s-1}\,\big|\big|\,\big|\big|\,\varphi\,\big|\big|_0^2 + \big|\big|\,\Lambda^{-2s+1}\overline{c}\,T^{2s-1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{2s-1}\,\big|\big|\,\varphi\,\big|\big|_0 \\ &\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2 + \big|\big|\,\varphi\,\big|\big|_0^2\Big]. \end{aligned}$$

Finally, to deal with the therm II_b, note that

$$YX_j^2 = X_j^2Y + [Y,\, X_j]\,X_j + X_j\,[Y,\, X_j]\,.$$

Thus

$$\begin{aligned} II_b &= \sum_{j=1}^{p} \Big(X_j^2Y[\varphi], T^{2s-1}[\varphi]\Big) + \sum_{j=1}^{p} \Big(\,[Y,\, X_j]\,X_j[\varphi], T^{2s-1}[\varphi]\Big) \\ &\qquad + \sum_{j=1}^{p} \Big(X_j\,[Y,\, X_j]\,[\varphi], T^{2s-1}[\varphi]\Big) \\ &= II_{b,1} + II_{b,2} + II_{b,3}. \end{aligned}$$

The bounds for $II_{b,1}$ follow from Lemma 3.11.

To deal with the terms in $II_{b,2}$ write

$$\begin{aligned} \Big(\,[Y,\, X_j]\,X_j[\varphi], T^{2s-1}[\varphi]\Big) &= \Big(X_j[\varphi], T^{2s-1}\,[Y,\, X_j]^*\,[\varphi]\Big) \\ &\qquad + \Big(X_j[\varphi], \big[[Y,\, X_j]\,,\, T^{2s-1}\big]\,[\varphi]\Big). \end{aligned}$$

We have

$$\Big|\Big(X_j[\varphi],\big[[Y,\,X_j]\,,T^{2s-1}\big]\,[\varphi]\Big)\Big| \leq \big|\big|\ \big[[Y,\,X_j]\,,\,T^{2s-1}\big]\ \big|\big|\,\big|\big|\,X_j[\varphi]\,\big|\big|_0\,\big|\big|\,\varphi\,\big|\big|_0$$
$$\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2+\big|\big|\,\varphi\,\big|\big|_0^2\Big]$$

since $s \leq \frac{1}{2}$. Also, since $Y \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$, it follows from part (5) of the Theorem that $[Y,\,X_j] \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$, and hence by Proposition 4.4, $[Y,\,X_j]^* \in \mathcal{M}(0, \frac{1}{2}\epsilon, U)$. Thus

$$\Big|\Big(X_j[\varphi], T^{2s-1}\,[Y,\,X_j]^*\,[\varphi]\Big)\Big| \leq \big|\big|\,T^{2s-1}\Lambda^{-\frac{1}{2}\epsilon+1}\,\big|\big|\,\big|\big|\ [Y,\,X_j]^*\,[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}$$
$$\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2+\big|\big|\,\varphi\,\big|\big|_0^2\Big]$$

since $s \leq \frac{1}{4}\frac{1}{2}\epsilon$.

We deal with the terms in $II_{b,3}$ in a similar way. Write

$$\Big(X_j\,[Y,\,X_j]\,[\varphi], T^{2s-1}[\varphi]\Big) = \Big(\,[Y,\,X_j]\,[\varphi], T^{2s-1}X_j^*[\varphi]\Big)$$
$$+\Big(\,[Y,\,X_j]\,[\varphi],\,\big[X_j^*,\,T^{2s-1}\big]\,[\varphi]\Big).$$

Then as before

$$\Big(\,[Y,\,X_j]\,[\varphi], T^{2s-1}X_j^*[\varphi]\Big) \leq \big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}T^{2s-1}\,\big|\big|\,\big|\big|\ [Y,\,X_j]\,[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\big|\big|\,X_j^*[\varphi]\,\big|\big|_0$$
$$\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2+\big|\big|\,\varphi\,\big|\big|_0^2\Big]$$

and

$$\Big|\Big(\,[Y,\,X_j]\,[\varphi],\,\big[X_j^*,\,T^{2s-1}\big]\,[\varphi]\Big)\Big||$$
$$\leq \big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}\,\big[X_j^*,\,T^{2s-1}\big]\ \big|\big|\,\big|\big|\ [Y,\,X_j]\,[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\,\big|\big|\,\varphi\,\big|\big|_0$$
$$\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0^2+\big|\big|\,\varphi\,\big|\big|_0^2\Big].$$

Now we turn to the term I. This time we write X_0 in terms of $\mathcal{L}^*$. Thus we have

$$X_0 = -\mathcal{L}^* + \sum_{j=1}^{p} X_j^2 + \sum_{j=1}^{p}\psi_j X_j - iY_0 - \widetilde{c}$$

where the functions $\{\psi_j\}$ and $\widetilde{c}$ belong to $\mathcal{C}^\infty(\Omega)$. Thus we can write

$$\begin{aligned}I &= -\Big(\mathcal{L}^*Y[\varphi], T^{2s-1}[\varphi]\Big) + \sum_{j=1}^{p}\Big(X_j^2Y[\varphi], T^{2s-1}[\varphi]\Big)\\ &\qquad+\sum_{j=1}^{p}\Big(\psi_jX_jY[\varphi], T^{2s-1}[\varphi]\Big) - i\Big(Y_0Y[\varphi], T^{2s-1}[\varphi]\Big) - \Big(\widetilde{c}\,Y[\varphi], T^{2s-1}[\varphi]\Big)\\ &= -I_a + I_b + I_c + I_d + I_e.\end{aligned}$$

We write

$$\begin{aligned}I_a &= \Big(Y[\varphi], T^{2s-1}\mathcal{L}[\varphi]\Big) + \Big(Y[\varphi],\,\big[\mathcal{L},\,T^{2s-1}\big]\,[\varphi]\Big)\\ &= \Big(Y[\varphi], T^{2s-1}\mathcal{L}[\varphi]\Big) + \sum_{j=1}^{p}\Big(Y[\varphi], B_j^{2s-1}X_j[\varphi]\Big) + \Big(Y[\varphi], B_0^{2s-1}[\varphi]\Big)\end{aligned}$$

Thus

$$
\begin{aligned}
\big|I_a\big| \leq \big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}T^{2s-1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0 \\
+\sum_{j=1}^{p}\big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}B_j^{2s-1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\big|\big|\,X_j[\varphi]\,\big|\big|_0 \\
+\big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}B_0^{2s-1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\,\big|\big|\,\varphi\,\big|\big|_0 \\
\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0+\big|\big|\,\varphi\,\big|\big|_0\Big].
\end{aligned}
$$

The corresponding estimate for the term I_b follows from Lemma 3.11. To deal with the term I_c write

$$
\begin{aligned}
\big|I_c\big| &\leq \sum_{j=1}^{p}\Big|\Big(Y[\varphi],\psi_jT^{2s-1}X_j^*[\varphi]\Big)\Big|+\sum_{j=1}^{p}\Big|\Big(Y[\varphi],\,\big[X_j^*,\,\psi_jT^{2s-1}\big]\,[\varphi]\Big)\Big| \\
&\leq \sum_{j=1}^{p}\big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}\psi_jT^{2s-1}\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1} \\
&\quad+\sum_{j=1}^{p}\big|\big|\,\Lambda^{-\frac{1}{2}\epsilon+1}\,\big[X_j^*,\,T^{2s-1}\big]\,\big|\big|\,\big|\big|\,Y[\varphi]\,\big|\big|_{\frac{1}{2}\epsilon-1}\,\big|\big|\,\varphi\,\big|\big|_0 \\
&\leq C_s\Big[\big|\big|\,\mathcal{L}[\varphi]\,\big|\big|_0+\big|\big|\,\varphi\,\big|\big|_0\Big].
\end{aligned}
$$

The terms I_d and I_e are estimated exactly as are the corresponding terms in term II. This completes the proof. □

5. The theorem for distributions

5.1. Mollifiers. Choose $\chi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ such that $\chi(-y)=\chi(y)$, $\int \chi(y)\,dy = 1$ and $\mathrm{suppt}(\chi) \subset \{y \in \mathbb{R}^n \,\big|\, |y| < 1\}$. For $t > 0$ set

$$\chi_t(x) = t^{-n}\,\chi(t^{-1}x).$$

Then $\mathrm{suppt}(\chi_t) \subset \{y \in \mathbb{R}^n \,\big|\, |y| < t\}$. For $f \in \mathcal{S}(\mathbb{R}^n)$ define

$$T_t[f](x) = f * \chi_t(x) = \int_{\mathbb{R}^n} f(y)\,\chi_t(x-y)\,dy,$$

or equivalently

$$T_t[f](x) = \int_{\mathbb{R}^n} e^{2\pi i<x,\xi>}\,\widehat{\psi_t}(\xi)\,\widehat{f}(\xi)\,d\xi = \int_{\mathbb{R}^n} e^{2\pi i<x,\xi>}\,\widehat{\psi}(t\xi)\,\widehat{f}(\xi)\,d\xi..$$

Then $\{T_t\} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is a family of pseudodifferential operators in $OP^0(\mathbb{R}^n)$ with uniformly bounded norms, and

$$\lim_{t\to 0} T_t[f](x) = f(x).$$

Note that since $\chi_t(x-y) = \chi_t(y-x)$,

$$\int_{\mathbb{R}^n} T_t[f](x)\,g(x)\,dx = \int_{\mathbb{R}^n} f(y)(\chi_t * g)(y)\,dy.$$

By duality, the operators $\{T_t\}$ extend to the space $\mathcal{S}'(\mathbb{R}^n)$ of tempered distributions on $\mathbb{R}^n$. If $u \in \mathcal{S}'(\mathbb{R}^n)$, then $T_t[u]$ is the distribution whose action on $f \in \mathcal{S}(\mathbb{R}^n)$ is given by

$$\big\langle T_t[u], f\big\rangle = \big\langle u, \chi_t * f\big\rangle.$$

If we write $\tau_y\chi_t(x) = \chi_t(x-y)$, then $\chi_t * f$ is a limit in $\mathcal{S}(\mathbb{R}^n)$ of sums of the form $\sum f(y_j)\tau_{y_j}(x)$. Since u is a continuous linear functional, it follows that

$$\big\langle T_t[u], f\big\rangle = \int_{\mathbb{R}^n} f(y)\big\langle u, \tau_y[\chi_t]\big\rangle\, dy.$$

Thus the distribution $T_t[u]$ is given by integration against the function $u_t(y) = \big\langle u, \tau_y[\chi_t]\big\rangle$. Moreover, it is easy to check that u_t is infinitely differentiable. In fact

$$\partial^\alpha u_t(y) = (-1)^{|\alpha|}\big\langle u, \tau_y[\partial^\alpha \chi_t]\big\rangle.$$

LEMMA 5.1. *Let $u \in \mathcal{E}'(\Omega)$ be a distribution with compact support $K \subset \Omega$. Let η be the distance from K to the complement of Ω, and let $u_t = T_t[u]$. Then $u_t \in \mathcal{C}_0^\infty(\Omega)$ for $0 < t < \eta$. Moreover,*

(1) If $u \in H^s(\mathbb{R}^n)$, then $\big|\big| u_t \big|\big|_s \leq \big|\big| u \big|\big|_s$, and $\lim_{t\to 0} \big|\big| u - u_t \big|\big|_s = 0$.
(2) If $\sup_{t>0} \big|\big| u_t \big|\big|_s = C < \infty$, then $u \in H^s(\mathbb{R}^n)$ and $\big|\big| u \big|\big|_s \leq C$.

5.2. The main result. Our object is to finally prove

THEOREM 2.2 *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $\mathcal{L}$ be the second order partial differential operator given in equation (2.1). Suppose the vector fields $\{X_0, X_1, \dots, X_p, Y_0\}$ and c satisfy hypotheses (H-1), (H-2), and (H-3). Set $\epsilon = 2\cdot 4^{-m}$, and fix $\zeta_1 \prec \zeta_0 \in \mathcal{C}_0^\infty(\Omega)$. For every $s \in \mathbb{R}$ there is an allowable constant C_s so that if $u \in \mathcal{D}'(\Omega)$ is a distribtuion on Ω and if $\zeta_0\mathcal{L}[u] \in H^s(\mathbb{R}^n)$ and $\zeta_0 u \in H^s(\mathbb{R}^n)$, then $\zeta_1 u \in H^{s+\epsilon}(\mathbb{R}^n)$, $\zeta_1 X_j u \in H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$, and*

$$\big|\big| \zeta_1 u \big|\big|_{s+\epsilon} + \sum_{j=1}^p \big|\big| \zeta_1 X_j u \big|\big|_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big| \zeta_0\mathcal{L}[u] \big|\big|_s + \big|\big| \zeta_0 u \big|\big|_s\Big]. \tag{2.5}$$

6. Examples of vector fields Y_0 satisfying hypothesis (H-2)

In many applications the vector field Y_0 has the form

$$Y_0 = \sum_{j=1}^p b_j X_j + \frac{1}{2}\sum_{k,l=1}^p c_{k,l}\,[X_k,\, X_l] \tag{6.1}$$

where $b_j, c_{k,l} \in \mathcal{C}^\infty(\Omega)$ are real-valued. The Lie bracket is anti-symmetric, and hence

$$\begin{aligned}\sum_{k,l=1}^p c_{k,l}\,[X_k,\, X_l] &= \frac{1}{2}\sum_{k,l=1}^p c_{k,l}\,[X_k,\, X_l] - \frac{1}{2}\sum_{k,l} c_{k,l}\,[X_l,\, X_k] \\ &= \sum_{k,l=1}^p \frac{(c_{k,l} - c_{l,k})}{2}\,[X_k,\, X_l]\,.\end{aligned}$$

Thus there is no loss in assuming $c_{k,l} = -c_{l,k}$. We will show that if the matrix $\{c_{k,l}\}$ is sufficiently small, the operator Y_0 given in (6.1) satisfies the hypotheses given in equations (2.2) and (2.3). Note that no size assumptions are needed on the coefficients $\{b_j\}$.

LEMMA 6.1. *Let* $Y_0 = \sum_{j=1}^p b_j X_j + \frac{1}{2}\sum_{k,l=1}^p c_{k,l}\,[X_k,\, X_l]$ *where* $b_j, c_{k,l} \in \mathcal{C}^\infty_{\mathbb{R}}(\Omega)$ *and* $c_{k,l} = -c_{l,k}$. *Suppose there exists* $\eta > 0$ *so that for all* $x \in \Omega$

$$\Big|\sum_{k,l=1}^p c_{k,l}(x)\xi_k\,\overline{\xi_l}\Big| \leq (1-2\eta)\sum_{j=1}^p |\xi_j|^2. \tag{6.2}$$

Then there are allowable constants C_0 *and* C_1 *so that for all* $\varphi,\, \psi \in \mathcal{C}^\infty_0(U)$,

$$\Big|\Im m\,\big(Y_0[\varphi],\varphi\big)_0\Big| \leq (1-\eta)\,\Big(\sum_{j=1}^p \big|\big|\, X_j[\varphi]\,\big|\big|_0^2\Big) + \eta^{-1}C_0\big|\big|\,\varphi\,\big|\big|_0^2; \tag{2.2}$$

$$\Big|\big(Y_0[\varphi],\psi\big)\Big| \leq C_1\,\Big[\sum_{j=1}^p \big(\big|\big|\, X_j[\varphi]\,\big|\big|_0^2 + \big|\big|\, X_j[\psi]\,\big|\big|_0^2\big) + \big|\big|\,\varphi\,\big|\big|_0^2 + \big|\big|\,\psi\,\big|\big|_0^2\Big]. \tag{2.3}$$

PROOF. It follows from equation (3.4) that for any vector field Y_0,

$$\big|\Re e\big(Y_0[\varphi],\varphi\big)_0\big| = \big|\big(\varphi, \nabla\cdot Y_0\,\varphi\big)_0\big| \leq C\big|\big|\,\varphi\,\big|\big|_0^2$$

where $C = \sup_{x\in U} |\mathrm{div}[Y_0](x)|$. Thus to establish equation (2.2) it suffices to estimate $\big|\big(Y_0[\varphi],\varphi\big)_0\big|$. We have

$$\begin{aligned}\big(Y_0[\varphi],\varphi\big)_0 &= \sum_{j=1}^p \big(b_j X_j[\varphi],\varphi\big)_0 + \frac{1}{2}\sum_{k,l}\big(c_{k,l}\,[X_k,\,X_l]\,[\varphi],\varphi\big)_0\\ &= I + II.\end{aligned}$$

We deal with the term I as follows. If $\varphi \in \mathcal{C}^\infty_0(U)$, then for any $\eta > 0$ we have

$$\begin{aligned} I &\leq \big|\big|\,\varphi\,\big|\big|_0 \sum_{j=1}^p \big|\big|\, b_j\,\big|\big|_{L^\infty(U)}\,\big|\big|\, X_j[\varphi]\,\big|\big|_0\\ &\leq \big|\big|\,\varphi\,\big|\big|_0\Big(\sum_{j=1}^p \big|\big|\, b_j\,\big|\big|^2_{L^\infty(U)}\Big)^{\frac{1}{2}}\,\Big(\sum_{j=1}^p \big|\big|\, X_j[\varphi]\,\big|\big|_0^2\Big)^{\frac{1}{2}}\\ &\leq \frac{\eta}{2}\Big(\sum_{j=1}^p \big|\big|\, X_j[\varphi]\,\big|\big|_0^2\Big) + \Big(\frac{1}{2\eta}\sum_{j=1}^p \big|\big|\, b_j\,\big|\big|^2_{L^\infty(U)}\Big)\,\big|\big|\,\varphi\,\big|\big|_0^2.\end{aligned}$$

To deal with II, use equation (3.5) and the fact that $c_{k,l}$ is real valued to obtain

$$\begin{aligned}\big(c_{k,l}\,[X_k,\,X_l]\,[\varphi],\varphi\big)_0 &= \big(X_kX_l[\varphi], c_{k,l}\varphi\big)_0 - \big(X_lX_k[\varphi], c_{k,l}\varphi\big)_0\\ &= -\big(X_l[\varphi], c_{k,l}X_k[\varphi]\big)_0 + \big(X_l[\varphi], X_k^*[c_{k,l}]\,\varphi\big)_0\\ &\qquad + \big(X_k[\varphi], c_{k,l}X_l[\varphi]\big)_0 - \big(X_k[\varphi], X_l^*[c_{k,l}]\,\varphi\big)_0\end{aligned}$$

Since $c_{k,l} = -c_{l,k}$ it follows that

$$II = \sum_{k,l=1}^p \big(c_{k,l}X_k[\varphi], X_l[\varphi]\big)_0 - \sum_{k,l=1}^p \big(X_k[\varphi], X_l^*[c_{k,l}]\,\varphi\big)_0,$$

and thus

$$\big|II\big| \leq \Big|\sum_{k,l=1}^p \big(c_{k,l}X_k[\varphi], X_l[\varphi]\big)_0\Big| + \sum_{k,l=1}^p \Big|\big(X_k[\varphi], \sum_{l=1}^p X_l^*[c_{k,l}]\,\varphi\big)_0\Big|.$$

As above, for any $\eta > 0$ we have

$$\Big|\big(X_k[\varphi], \sum_{l=1}^{p} X_l^*[c_{k,l}]\,\varphi\big)_0\Big| \leq \frac{\eta}{2}\sum_{j=1}^{p} \big|\big|\, X_j[\varphi]\,\big|\big|_0^2 + \Big[\frac{1}{2\eta}\sum_{k=1}^{p}\Big|\Big|\sum_{l=1}^{p} X_l^*[c_{k,l}]\Big|\Big|_{L^\infty(U)}^2\Big]\,\big|\big|\,\varphi\,\big|\big|_0^2$$

Putting the two estimates together, it follows that there is a constant $C_0 < \infty$ depending only on the supremums of $\{|b_j|(x)\}$ and $\{|X_l^*[c_{k,l}](x)|\}$ on Ω so that

$$\big|\big(Y_0[\varphi], \varphi\big)_0\big| \leq \Big|\sum_{k,l=1}^{p}\big(c_{k,l}X_k[\varphi], X_l[\varphi]\big)_0\Big| + \eta\Big(\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_0^2\Big) + \eta^{-1}C_0\big|\big|\,\varphi\,\big|\big|_0^2.$$

However by hypothesis

$$\Big|\sum_{k,l=1}^{p}\big(c_{k,l}X_k[\varphi], X_l[\varphi]\big)_0\Big| \leq \int_{\mathbb{R}^n}\Big|\sum_{k,l=1}^{p} c_{k,l}(x)X_k[\varphi](x)\,\overline{X_l[\varphi](x)}\Big|\,dx$$

$$\leq (1-2\eta)\int_{\mathbb{R}^n}\sum_{j=1}^{p}\big|X_j[\varphi](x)\big|^2\,dx = (1-2\eta)\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_0^2,$$

and so we have

$$\big|\big(Y_0[\varphi], \varphi\big)_0\big| \leq (1-\eta)\Big(\sum_{j=1}^{p}\big|\big|\,X_j[\varphi]\,\big|\big|_0^2\Big) + \eta^{-1}C_0\big|\big|\,\varphi\,\big|\big|_0^2.$$

This establishes estimate (2.2).

The proof of the estimate (2.3) is even easier. Using the same calculations as before we see that

$$\begin{aligned}\big|\big(Y_0[\varphi], \psi\big)\big| &\leq \sum_{k,l=1}^{p}\big|\big(c_{k,l}X_k[\varphi], X_l[\psi]\big)\big| + \sum_{k,l=1}^{p}\big|\big(X_k[\varphi], X_l^*[c_{k,l}]\psi\big)\big| \\ &\quad + \sum_{j=1}^{p}\big|\big(b_jX_j[\varphi], \psi\big)\big| \\ &\leq C_1\Big[\sum_{j=1}^{p}\big(\big|\big|\,X_j[\varphi]\,\big|\big|_0^2 + \big|\big|\,X_j[\psi]\,\big|\big|_0^2\big) + \big|\big|\,\varphi\,\big|\big|_0^2 + \big|\big|\,\psi\,\big|\big|_0^2\Big],\end{aligned}$$

where C_1 depends on the supremums of $|c_{k,l}(x)|$, $|X_l^*[c_{k,l}](x)|$, and $|b_j(x)|$ for $x \in \Omega$. This completes the proof. □

Suppose $u \in \mathcal{D}'(\Omega)$ and that u and $\mathcal{L}[u]$ are locally in $H^s(\mathbb{R}^n)$. This means that if $\varphi \in \mathcal{C}_0^\infty(\Omega)$, then both $\varphi\,u$ and $\varphi\,\mathcal{L}[u]$ belong to $H^s(\mathbb{R}^n)$.

Proposition 6.2. *Let $\zeta \in \mathcal{C}_0^\infty(\Omega)$, let $T_t[f] = f * \chi_t$, and let X be a vector field. There exists $\zeta' \in \mathcal{C}_0^\infty(\Omega)$ with $\zeta \prec \zeta'$ so that for all $s \in \mathbb{R}$ there is an allowable constant C_s so that for all distributions $u \in \mathcal{D}'(\Omega)$*

$$\big|\big|\,T_t\zeta u\,\big|\big|_s \leq C_s\big|\big|\,\zeta' u\,\big|\big|_s$$
$$\big|\big|\,[T_t\varphi,\, X]\,[u]\,\big|\big|_s \leq C_s\big|\big|\,\zeta' u\,\big|\big|_s.$$

PROPOSITION 6.3. *$X_j u$ is locally in $H^s(\mathbb{R}^n)$. If $\zeta \in \mathcal{C}_0^\infty(\Omega)$ there exists $\zeta' \in \mathcal{C}_0^\infty(\Omega)$ with $\zeta \prec \zeta'$ so that*

$$\big|\big|\,\zeta X_j u\,\big|\big|_s^2 \leq C_s\Big[\big|\big|\,\zeta'\mathcal{L}[u]\,\big|\big|_s^2 + \big|\big|\,\zeta' u\,\big|\big|_s^2\Big]. \tag{6.3}$$

PROOF. We need to show that if $\zeta \in \mathcal{C}_0^\infty(\Omega)$, then $\zeta X_j u \in H^s(\mathbb{R}^n)$, and it thus suffices to show that $\big|\big|\,T_t\zeta X_j u\,\big|\big|_s$ is uniformly bounded for small t. Let the support of ζ be the compact set $K \subset \Omega$. We have

$$\big|\big|\,T_t\zeta X_j u\,\big|\big|_s \leq \big|\big|\,X_j T_t\zeta u\,\big|\big|_s + \big|\big|\,[T_t\zeta,\, X_j]\,u\,\big|\big|_s \leq \big|\big|\,X_j T_t\zeta u\,\big|\big|_s + C_s\,\big|\big|\,\zeta' u\,\big|\big|_s$$

by Proposition 6.2. Since $T_t\zeta u \in \mathcal{C}_0^\infty(\Omega)$ for $0 < t < d_K$ we can apply Lemma 3.12 and write

$$\begin{aligned}
\frac{\eta}{8}\sum_{j=1}^{p}\big|\big|\,X_j T_t\zeta u\,\big|\big|_s^2 &\leq \Re e\Big(\mathcal{L}T_t\zeta u, T_t\zeta u\Big)_s + C_s\,\big|\big|\,T_t\zeta u\,\big|\big|_s^2 \\
&= \Big|\Big(T_t\zeta\mathcal{L}u, T_t\zeta u\Big)_s\Big| + \Big|\Big(\,[\mathcal{L},\, T_t\zeta]\,u, T_t\zeta u\Big)_s\Big| + C_s\,\big|\big|\,T_t\zeta u\,\big|\big|_s^2 \\
&\leq \Big|\Big(\,[\mathcal{L},\, T_t\zeta]\,u, T_t\zeta u\Big)_s\Big| + \big|\big|\,T_t\zeta\mathcal{L}u\,\big|\big|_s\big|\big|\,T_t\zeta u\,\big|\big|_s + C_s\big|\big|\,T_t\zeta u\,\big|\big|_s^2 \\
&\leq \Big|\Big(\,[\mathcal{L},\, T_t\zeta]\,u, T_t\zeta u\Big)_s\Big| + C_s\Big[\big|\big|\,\zeta'\mathcal{L}u\,\big|\big|_s^2 + \big|\big|\,\zeta' u\,\big|\big|_s^2\Big].
\end{aligned}$$

Now use Proposition 3.6 to write $[\mathcal{L},\, T_t\zeta] = \widetilde{B}_0 + \sum_{j=1}^p X_j^*\widetilde{B}_j$. Then

$$\begin{aligned}
\Big|\Big(\,[\mathcal{L},\, T_t\zeta]u, T_t\zeta u\Big)_s\Big| &\leq +\sum_{j=1}^{p}\Big|\Big(\Lambda^{2s}X_j^*\widetilde{B}_j u, T_t\zeta u\Big)_0\Big| + \Big|\Big(\Lambda^{2s}\widetilde{B}_0 u, T_t\zeta u\Big)_0\Big| \\
&\leq \sum_{j=1}^{p}\Big|\Big(X_j^*\Lambda^{2s}\widetilde{B}_j u, T_t\zeta u\Big)_0\Big| + \sum_{j=1}^{p}\Big|\Big(\Lambda^{-s}\,\big[\Lambda^{2s},\, X_j^*\big]\,\widetilde{B}_j u, \Lambda^s T_t\zeta u\Big)_0\Big| \\
&\qquad + \Big|\Big(\Lambda^{2s}\widetilde{B}_0 u, T_t\zeta u\Big)_0\Big| \\
&= \sum_{j=1}^{p}\Big|\Big(\widetilde{B}_j u, X_j T_t\zeta u\Big)_s\Big| + \sum_{j=1}^{p}\big|\big|\,\Lambda^{-2s}\,\big[\Lambda^{2s},\, X_j^*\big]\,\widetilde{B}_j u\,\big|\big|_s\,\big|\big|\,T_t\zeta u\,\big|\big|_s \\
&\qquad + \big|\big|\,\widetilde{B}_0 u\,\big|\big|_s\,\big|\big|\,T_t\zeta u\,\big|\big|_s \\
&\leq \delta\,\sum_{j=1}^{p}\big|\big|\,X_j T_t\zeta u\,\big|\big|_s^2 + C_s(1+\delta^{-1})\big|\big|\,\zeta' u\,\big|\big|_s^2.
\end{aligned}$$

Choosing $\delta = \frac{\eta}{16}$ it follows that

$$\sum_{j=1}^{p}\big|\big|\,T_t\zeta X_j u\,\big|\big|_s^2 \leq C_s\,\Big[\big|\big|\,\zeta'\mathcal{L}u\,\big|\big|_s^2 + \big|\big|\,\zeta' u\,\big|\big|_s^2\Big]$$

which completes the proof. □

LEMMA 6.4. *Let $\frac{1}{2}\epsilon = 2\cdot 4^{-m}$. Let $u \in \mathcal{D}'(\Omega)$ be a distribution, and suppose that for all $\zeta \in \mathcal{C}_0^\infty(\Omega)$, $\zeta\, u \in H^s(\mathbb{R}^n)$ and $\zeta\,\mathcal{L}[u] \in H^s(\mathbb{R}^n)$. Then $\zeta\, u \in H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$, and there is an allowable constant C_s such that*

$$\big|\big|\,\zeta\, u\,\big|\big|_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\,\zeta'\mathcal{L}[u]\,\big|\big|_s + \big|\big|\,\zeta' u\,\big|\big|_s\Big]. \tag{6.4}$$

PROOF. Suppose the support of ζ is the compact set $K \subset \Omega$. For $0 < t < d_K$, $T_t\zeta\, u \in \mathcal{C}_0^\infty(\Omega)$, so we can apply Theorem 4.1 to conclude that

$$\big|\big|\, T_t\zeta\, u\,\big|\big|_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\, \mathcal{L}T_t\zeta\, u\,\big|\big|_s + \big|\big|\, T_t\zeta\, u\,\big|\big|_s\Big].$$

As usual, we commute $\mathcal{L}$ and $T_t\zeta$ to obtain

$$\big|\big|\, \mathcal{L}T_t\zeta\, u\,\big|\big|_s \leq \big|\big|\, T_t\zeta\, \mathcal{L}[u]\,\big|\big|_s + \big|\big|\, [\mathcal{L},\, T_t\zeta\,]\, u\,\big|\big|_s.$$

Again using Proposition 3.6, we can write $[\mathcal{L},\, T_t\zeta\,] = B_0 + \sum_{j=1}^p B_j\, X_j$ where $\{B_0, \ldots, B_p\}$ are pseudodifferential operators of order zero with

$$\big|\big|\, B_j v\,\big|\big|_s \leq C_s\big|\big|\, \zeta' v\,\big|\big|_s$$

for any distribution $v \in \mathcal{D}'(\Omega)$. Thus by Proposition **??**,

$$\begin{aligned}\big|\big|\, [\mathcal{L},\, T_t\zeta\,]\, u\,\big|\big|_s &\leq \big|\big|\, B_0 u\,\big|\big|_s + \sum_{j=1}^p \big|\big|\, B_j X_j u\,\big|\big|_s\\ &\leq \big|\big|\, \zeta' u\,\big|\big|_s + \sum_{j=1}^p \big|\big|\, \zeta' X_j u\,\big|\big|_s\\ &\leq C_s\Big[\big|\big|\, \zeta''\mathcal{L}[u]\,\big|\big|_s + \big|\big|\, \zeta'' u\,\big|\big|_s\Big].\end{aligned}$$

It follows that $\big|\big|\, T_t\zeta\, u\,\big|\big|^2_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\, \zeta''\mathcal{L}[u]\,\big|\big|_s + \big|\big|\, \zeta'' u\,\big|\big|_s\Big]$, and the Lemma follows. □

The last step is to derive an improvement of Proposition 6.3.

LEMMA 6.5. *Let $\epsilon = 2\cdot 4^{-m}$. Let $u \in \mathcal{D}'(\Omega)$ be a distribution, and suppose that for all $\zeta \in \mathcal{C}_0^\infty(\Omega)$, $\zeta\, u \in H^s(\mathbb{R}^n)$ and $\zeta\, \mathcal{L}[u] \in H^s(\mathbb{R}^n)$. Then $X_j[u] \in H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$ for $1 \leq j \leq p$, and*

$$\sum_{j=1}^p \big|\big|\, \zeta\, X_j u\,\big|\big|^2_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\, \zeta'\mathcal{L}[u]\,\big|\big|^2_s + \big|\big|\, \zeta' u\,\big|\big|^2_s\Big]. \tag{6.5}$$

PROOF. The proof consists of repeating the argument for Proposition 6.3, except that we now can replace $\big|\big|\, \zeta\, u\,\big|\big|_s$ by $\big|\big|\, \zeta\, u\,\big|\big|_{s+\epsilon}$. We have $T_t\zeta\, u \in \mathcal{C}_0^\infty(\Omega)$ for $0 < t < d_K$, and we so can use Lemma 3.12 to conclude that

$$\begin{aligned}\frac{\eta}{8}\sum_{j=1}^p \big|\big|\, X_j T_t\zeta\, u\,\big|\big|^2_{s+\frac{1}{2}\epsilon} &\leq \Re e\Big(\mathcal{L}[T_t\zeta\, u], T_t\zeta\, u\Big)_s + C_s\big|\big|\, T_t\zeta\, u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\\ &= \Re e\Big(T_t\zeta\, \mathcal{L}[u], T_t u\Big)_{s+\frac{1}{2}\epsilon} + \Re e\Big(\,[\mathcal{L},\, T_t\zeta]\, u, T_t\zeta\, u\Big)_{s+\frac{1}{2}\epsilon}\\ &\qquad + C_s\big|\big|\, T_t\zeta\, u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\end{aligned}$$

But

$$\begin{aligned}\Re e\Big(T_t\zeta\, \mathcal{L}[u], T_t u\Big)_{s+\frac{1}{2}\epsilon} &\leq \big|\big|\, T_t\zeta\, \mathcal{L}[u]\,\big|\big|_s\big|\big|\, T_t\zeta\, u\,\big|\big|_{s+\epsilon}\\ &\leq \frac{1}{2}\big|\big|\, \zeta\, \mathcal{L}[u]\,\big|\big|^2_s + \frac{1}{2}\big|\big|\, \zeta\, u\,\big|\big|^2_{s+\epsilon},\end{aligned}$$

and

$$\Big(\,[\mathcal{L},\, T_t\zeta]\, [u], T_t u\Big)_{s+\frac{1}{2}\epsilon} = \Big(\Lambda^{2s+\epsilon}\, [\mathcal{L},\, T_t\zeta]\, [u], T_t u\Big)_0.$$

According to Proposition 3.6, we can write $[\mathcal{L}, T_t\zeta\,] = \widetilde{B}_0 + \sum_{j=1}^p X_j^*\widetilde{B}_j$ where $\{\widetilde{B}_0, \dots, \widetilde{B}_p\}$ are pseudodifferential operators of order zero, and hence

$$\Lambda^{2s}\,[\mathcal{L}, T_t\zeta\,] = \sum_{j=1}^p X_j^*\Lambda^{2s}\widetilde{B}_j + \sum_{j=1}^p \left[\Lambda^{2s}, X_j^*\right]\widetilde{B}_j + \Lambda^{2s}\widetilde{B}_0.$$

Hence for any $\delta > 0$ we have

$$\begin{aligned}\Big([\mathcal{L}, T_t\zeta\,]\,[u], T_t\zeta\,u\Big)_{s+\frac{1}{2}\epsilon} &= \sum_{j=1}^p \Big(\widetilde{B}_j[u], X_j[T_t\zeta\,u]\Big)_{s+\frac{1}{2}\epsilon} + C_s\big|\big|\,\zeta\,u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\\ &\leq \delta\sum_{j=1}^p \big|\big|\,X_jT_t\zeta\,u\,\big|\big|^2_{s+\frac{1}{2}\epsilon} + C_s(1+\delta^{-1})\big|\big|\,\zeta\,u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}.\end{aligned}$$

Putting these inequalities together, we get

$$\begin{aligned}\frac{\eta}{8}\sum_{j=1}^p \big|\big|\,X_j[T_t\zeta\,u]\,\big|\big|^2_{s+\frac{1}{2}\epsilon} \leq \delta\sum_{j=1}^p \big|\big|\,X_j[T_t\zeta\,u]\,\big|\big|^2_{s+\frac{1}{2}\epsilon}&\\ + C_s\Big[\big|\big|\,T_t\zeta\,\mathcal{L}[u]\,\big|\big|^2_s + \big|\big|\,T_t\zeta\,u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\Big]&.\end{aligned}$$

If we choose $\delta = \frac{\eta}{16}$, it follows that

$$\sum_{j=1}^p \big|\big|\,X_j[T_t\zeta\,u]\,\big|\big|^2_{s+\frac{1}{2}\epsilon} \leq C_s\Big[\big|\big|\,\zeta'\mathcal{L}[u]\,\big|\big|^2_s + \big|\big|\,\zeta' u\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\Big].$$

Finally,

$$\begin{aligned}\sum_{j=1}^p \big|\big|\,T_t\zeta\,X_ju\,\big|\big|^2_{s+\frac{1}{2}\epsilon} &\leq 2\sum_{j=1}^p \big|\big|\,X_j[T_t\zeta\,u]\,\big|\big|^2_{s+\frac{1}{2}\epsilon} + 2\sum_{j=1}^p \big|\big|\,[T_t\zeta\,, X_j]\,[u]\,\big|\big|^2_{s+\frac{1}{2}\epsilon}\\ &\leq C_s\Big[\big|\big|\,\zeta'\mathcal{L}[u]\,\big|\big|^2_s + \big|\big|\,\zeta' u\,\big|\big|^2_{s+\epsilon}\Big]\end{aligned}$$

which by Lemma 6.4 is

$$\leq C_s\Big[\big|\big|\,\zeta''\mathcal{L}[u]\,\big|\big|^2_s + \big|\big|\,\zeta'' u\,\big|\big|^2_s\Big].$$

This shows that $\zeta\,X_ju \in H^{s+\frac{1}{2}\epsilon}(\mathbb{R}^n)$ with the correct estimate for the sum of the norms. □

CHAPTER 5

Estimates for fundamental solutions

1. Introduction

Let $\Omega \subset \mathbb{R}^n$ be a connected open set, and let $\{X_1, \ldots, X_p\}$ be smooth real vector fields on Ω of finite type m. In this chapter, we construct a fundamental solution K for the second order partial differential operator

$$\mathcal{L} = \sum_{j=1}^{p} X_j^* X_j = -\sum_{j=1}^{p} X_j^2 - \sum_{j=1}^{p} (\nabla \cdot X_j) X_j \tag{1.1}$$

and obtain size estimates for K and its derivatives in terms of the control metric $\rho : \Omega \times \Omega \to [0, \infty)$ induced by the given vector fields. These estimates provide the first hint of the deep connection between the geometry of control metrics developed in Chapter A and analytic problems that arise in certain non-elliptic problems. And just as the properties of the Newtonian potential $N(x) = c_n |x|^{2-n}$, which is a fundamental solution for the elliptic Laplace operator $\triangle$, play an important role in the development of the calculus of pseudodifferential operators, in Chapter B we will use the estimates for the fundamental solution K to develop a calculus of non-isotropic smoothing (NIS) operators that are useful in the study of non-elliptic by still hypoelliptic operators such as $\mathcal{L}$.

Our construction of the fundamental solution for $\mathcal{L}$ proceeds in roughly four steps. We begin by studying the initial value problem for the heat operator associated to $\mathcal{L}$,

$$\mathcal{H} = \partial_t + \mathcal{L}_x = \partial_t - \sum_{j=1}^{p} X_j^2 - \sum_{j=1}^{p} (\nabla \cdot X_j) X_j, \tag{1.2}$$

which acts on functions and distributions on $\mathbb{R} \times \Omega$. Given $g \in L^2(\Omega)$, we want to find $u \in \mathcal{C}^\infty\big((0, \infty) \times \Omega\big)$ such that

$$\partial_t u(t, x) + \mathcal{L}_x[u](t, x) = 0 \qquad \text{for } (t, x) \in (0, \infty) \times \Omega, \text{ and} \tag{1.3}$$

$$\lim_{t \to 0^+} u(t, \cdot) = g(\,\cdot\,) \qquad \text{with convergence in } L^2(\Omega). \tag{1.4}$$

We show that the partial differential operator $\mathcal{L}$, defined initially on $\mathcal{C}_0^\infty(\Omega)$, has an extension to a self-adjoint (unbounded) operator on $L^2(\Omega)$. We then obtain a solution to (1.3) and (1.4) by setting $u(t, x) = e^{-t\mathcal{L}}[f](x)$ where the bounded operator $e^{-t\mathcal{L}} : L^2(\Omega) \to L^2(\Omega)$ is defined by the spectral theorem. At least formally, the function u will satisfy both conditions (1.3) and (1.4), since

$$\partial_t u = \frac{\partial}{\partial t}\big(e^{-t\mathcal{L}}[g]\big) = -\mathcal{L}_x\big(e^{-t\mathcal{L}}[g]\big) = -\mathcal{L}_x[u],$$

and

$$\lim_{t \to 0^+} u = \lim_{t \to 0^+} \big(e^{-t\mathcal{L}}[g]\big) = e^{0\mathcal{L}}[g] = g.$$

The family of operators $\{e^{-t\mathcal{L}}\}_{t\geq 0}$ form a strongly continuous semigroup of bounded operators on $L^2(\Omega)$, called the heat semigroup for $\mathcal{L}$. The second step of our program is to show that there is a function $H \in \mathcal{C}\big((0,\infty)\times\Omega\times\Omega\big)$ so that if $g \in L^2(\Omega)$,

$$e^{-t\mathcal{L}}[g](x) = \int_\Omega H(t,x,y)\,g(y)\,dy. \tag{1.5}$$

Our proof of the existence and smoothness of H relies on the subelliptic estimates established for the operator $\mathcal{L}$ in Chapter A. The function H is called the heat kernel for $\mathcal{L}$. Moreover, we show that if H is extended to $\mathbb{R}\times\Omega\times\Omega$ by setting $H(t,x,y) = 0$ for $t \leq 0$, then $\widetilde{H}\big((t,x),(s,y)\big) = H(s+t,x,y)$ is a fundamental solution for the heat operator $\partial_t + \mathcal{L}_x$ on $\mathbb{R}\times\Omega$. We now observe that the operator $\partial_t + \mathcal{L}_x$ is hypoelliptic, and since $[\partial_s + \mathcal{L}_y]\widetilde{H} = \delta_t \otimes \delta_x$, it follows that $\widetilde{H}$ is infinitely differentiable on $(\mathbb{R}\times\Omega)\times(\mathbb{R}\times\Omega)$ except when $t = s$ and $x = y$.

The third step is to obtain pointwise estimates for the function H and its derivatives in terms of the control metric ρ. For example, we show that for any integer $N \geq 0$ there is a constant C_N so that

$$H(t,x,y) \leq C_N \begin{cases} |B\big(x,\rho(x,y)\big)|^{-1}\,\left(\frac{t}{\rho(x,y)^2}\right)^N & \text{if } t \leq \rho(x,y)^2, \\ \\ |B\big(x,\sqrt{t}\big)|^{-1} & \text{if } t \geq \rho(x,y)^2. \end{cases}$$

The proof uses a scaling argument and the fact that the operators $e^{-t\mathcal{L}}$ are contractions on $L^2(\Omega)$.

The final step is to show that we can make sense of the integral $\int_0^\infty e^{-t\mathcal{L}}\,dt$ (which formally equals $\mathcal{L}^{-1}$), and that the resulting operator is indeed a fundamental solution K for $\mathcal{L}$. Explicitly,

$$K(x,y) = \int_0^\infty H(t,x,y)\,dt. \tag{1.6}$$

We can then use the estimates obtained for H to derive estimates for K. For example, we show that there is a constant C so that

$$|K(x,y)| \leq C\,\frac{\rho(x,y)^2}{|B\big(x,\rho(x,y)\big)|}.$$

2. Unbounded operators on Hilbert space

Hilbert space techniques are frequently used to reduce problems about the existence of solutions to linear partial differential operators to problems of establishing estimates. The basic idea is very simple. Let $T : V \to W$ be a linear operator from a Hilbert space V to a Hilbert space W, and suppose we want to show that for every $y_0 \in W$, we can solve

$$T[x] = y_0 \tag{2.1}$$

with $x \in V$. Let T^* denote the adjoint operator, and suppose we can prove that for all $z \in W$,

$$\big|\big|\,z\,\big|\big|_W \leq C\,\big|\big|\,T^*[z]\,\big|\big|_V. \tag{2.2}$$

It follows from equation (2.2) that T^* is one-to-one. Define a linear function L_{y_0} on the range of T^* by setting $L_{y_0}\big[T^*[z]\big] = (z,y_0)_W$. Then

$$\big|L_{y_0}\big[T^*[z]\big]\big| \leq \big|\big|\,y_0\,\big|\big|_W\,\big|\big|\,z\,\big|\big|_W \leq C\,\big|\big|\,y_0\,\big|\big|_W\,\big|\big|\,T^*[z]\,\big|\big|_V$$

and so L_{y_0} is bounded. Using the Hahn-Banach theorem[1], we can extend L_{y_0} to a bounded linear functional on all of V (with no increase in norm), and by the Riesz representation theorem[2], there exists $x_0 \in V$ so that $L_{y_0}[z] = (z, x_0)_V$. But this means that for all $z \in H$ we have

$$(z, y_0)_W = L_{y_0}\big[T^*[z]\big] = (T^*[z], x_0)_V = (z, T[x_0])_W,$$

and hence $T[x_0] = y_0$. The the estimate (2.2) implies the existence of a solution[3] to the equation (2.1).

Unfortunately, a linear differential operator T on a set $\Omega \subset \mathbb{R}^n$ will generally not be defined for all functions in the natural Hilbert space $L^2(\Omega)$ and will not be bounded on those functions for which it is defined. In particular, the symbol T^* in the inequality (2.2) is not defined. Thus in applying Hilbert space techniques to problems involving differential equations, it is important to consider operators on a Hilbert space H which are not bounded. To do this, it is convenient to consider linear mappings which are only defined on some (usually dense) subspace of H and which is not bounded on this subspace.

In contrast with the theory of bounded linear operators on a Hilbert space, it is a critical and often delicate issue to specify the domain of an unbounded linear operator such as a differential operator. As we will see, the choice of this subspace often encodes the appropriate boundary conditions for the differential operator. Different subspaces correspond to different boundary conditions. Although we do assume familiarity with the elementary aspects of the theory of bounded operators on Hilbert spaces, in this section we develop the theory of unbounded operators to the point where we can quote the spectral theorem for self-adjoint operators. We begin with the formal definition.

DEFINITION 2.1. *An unbounded operator T from a Hilbert space V to a Hilbert space W is a pair $(T, \mathrm{Dom}(T))$ where the subspace $\mathrm{Dom}(T) \subset V$ is the domain of the operator, and $T : \mathrm{Dom}(T) \to W$ is a linear mapping. An operator $(S, \mathrm{Dom}(S))$ is an* extension *of the operator $(T, \mathrm{Dom}(T))$ if $\mathrm{Dom}(T) \subset \mathrm{Dom}(S)$ and $T[x] = S[x]$ for every $x \in \mathrm{Dom}(T)$.*

2.1. Closed, densely defined operators and their adjoints.

For our purposes, the most important unbounded operators are *densely defined* and *closed*. These concepts are defined as follows. First, the operator $(T, \mathrm{Dom}(T))$ is densely defined if $\mathrm{Dom}(T)$ is a dense subspace of V. Next, recall that $V \oplus W$ is a Hilbert space with inner product given by

$$\big((x_1, y_1), (x_2, y_2)\big)_{V\oplus W} = (x_1, x_2)_V + (y_1, y_2)_W.$$

Then the operator $(T, \mathrm{Dom}(T))$ is closed if the graph

$$G_T = \Big\{(x, y) \in V \oplus W \,\Big|\, x \in \mathrm{Dom}(T) \text{ and } y = T[x]\Big\}$$

---

[1] HAHN-BANACH THEOREM: *Let X be a complex normed vector space, let $Y \subset X$ be a subspace, and let $\ell : Y \to \mathbb{C}$ be a linear functional such that $|\ell(y)| \leq C\big|\big|\, y \,\big|\big|$ for all $y \in Y$. Then there exists a linear functional $L : X \to \mathbb{C}$ such that $L[y] = \ell[y]$ for all $y \in Y$, and $|L[x]| \leq C\big|\big|\, x \,\big|\big|$ for all $x \in X$.* (See, for example, [**Fol84**], page 150.)

[2] RIESZ REPRESENTATION THEOREM: *If L is a bounded linear functional on a Hilbert space H, there exists a unique element $x \in H$ so that $L[y] = (y, x)_H$ for all $y \in H$.* (See, for example, [**Fol84**], page 166.)

[3] We shall use a variant of this argument in Theorem 3.1 below. See Remark 3.3.

is a closed subspace of $V \oplus W$. Equivalently, $(T, \mathrm{Dom}(T))$ is closed if whenever $\{x_n\}$ is a sequence in $\mathrm{Dom}(T)$ such that $\lim_{n\to\infty} x_n = x_0$ and $\lim_{n\to\infty} T[x_n] = y_0$, it then follows that $x_0 \in \mathrm{Dom}(T)$ and $T[x_0] = y_0$.

If $(T, \mathrm{Dom}(T))$ is a closed, densely defined operator on H, we can define the *Hilbert space adjoint* $(T^*, \mathrm{Dom}(T^*)$. This is intended to extend the concept of the adjoint B^* of a bounded operator $B : V \to W$ for which we have $(B[x], y)_W = (x, B^*[y])_V$. To motivate the definition, note that if $x \in \mathrm{Dom}(T)$ and $y \in \mathrm{Dom}(T^*)$, it is reasonable to expect that we should have

$$\big|\big(T[x], y\big)_W\big| = \big|\big(x, T^*[y]\big)_V\big| \leq \big|\big|\, T^*[y]\,\big|\big|_V \,\big|\big|\, x\,\big|\big|_V = C_y \,\big|\big|\, x\,\big|\big|_V.$$

In particular, the linear functional $L_y[x] = \big(T[x], y\big)$ defined on $\mathrm{Dom}(T)$ should be bounded.

Definition 2.2. *Set*

$$\mathrm{Dom}(T^*) = \Big\{ y \in W \,\Big|\, \big(\exists C_y\big)\,\big(\forall x \in \mathrm{Dom}(T)\big)\,\Big(\big|(T[x], y)_W\big| \leq C_y \big|\big|\, x\,\big|\big|_V\Big)\Big\}.$$

The linear functional $L_y[x] = \big(T[x], y\big)$ *defined for* $x \in \mathrm{Dom}(T)$ *extends by continuity to* V *which is the closure of* $\mathrm{Dom}(T)$. *If* $y \in Dom(T^*)$, *then* $T^*[y]$ *is the unique element of* V *(given by the Riesz representation theorem) such that* $\big(T[x], y\big)_W = L_y[x] = \big(x, T^*[y]\big)_V$.

Put slightly differently, if $x \in \mathrm{Dom}(T)$ and $y \in \mathrm{Dom}(T^*)$, then

$$(T[x], y)_W = (x, T^*[y])_V,$$

and if $y, z \in W$ have the property that $(T[x], y)_W = (x, z)_V$ for every $x \in \mathrm{Dom}(T)$, it follows that $y \in \mathrm{Dom}(T^*)$ and $T^*[y] = z$.

Remark 2.3. *It follows immediately from this last characterization of the adjoint that if* $V = W = H$ *and* $I : H \to H$ *is the identity operator, the standard formula* $(T + \lambda I)^* = T^* + \bar{\lambda} I$ *continues to be true. Precisely, suppose that* $(T, \mathrm{Dom}(T))$ *is a closed, densely defined operator on a Hilbert space* H *and that* $\lambda \in \mathbb{C}$. *Then* $(T + \lambda I, \mathrm{Dom}(T))$ *is a closed, densely defined operator with adjoint* $(T^* + \bar{\lambda} I, \mathrm{Dom}(T^*))$.

The basic properties of adjoints follow easily from a study of the graphs of the operators. We first introduce the following notation. Define J, $R : W \oplus V \to V \oplus W$ and J^*, $R^* : V \oplus W \to W \oplus V$by setting

$$J(x, y) = (-y, x) \qquad R(x, y) = (y, x),$$
$$J^*(y, x) = (x, -y) \qquad R(y, x) = (x, y).$$

The following is then the basic result about adjoints.

Lemma 2.4. *Let* $(T, \mathrm{Dom}(T))$ *be a closed, densely defined operator on a Hilbert space* H. *Then*

(1) The graphs of T *and* T^* *are related by*

$$G_{T^*} = \big(J(G_T)\big)^\perp = J\big((G_T)^\perp\big);$$
$$G_T = \big(J^*(G_{T^*})\big)^\perp = J^*\big((G_{T^*})^\perp\big).$$

(2) The operator $\big(T^*, \mathrm{Dom}(T^*)\big)$ *is closed and densely defined.*

(3) *The Hilbert space adjoint of the operator* $\big(T^*, \mathrm{Dom}(T^*)\big)$ *is the operator* $(T, \mathrm{Dom}(T))$.

PROOF. Let $\big(-T[x], x\big) \in J\big(G_T\big)$ and $\big(y, T^*[y]\big) \in G_{T^*}$. Then

$$\big((-T[x], x), (y, T^*[y]),\big)_{W\oplus V} = -(T[x], y)_W + (x, T^*[y])_V = 0,$$

so $G_{T^*} \subset \big(J\big(G_T\big)\big)^\perp$. On the other hand, if $(z, y) \in \big(J\big(G_T\big)\big)^\perp$, then for every $x \in \mathrm{Dom}(T)$ we have $0 = \big((z, y), (x, -T[x])\big)_{W\oplus V} = \big(z, x\big)_V - \big(y, T[x]\big)_{+W}$, or $(T[x], y)_W = (x, z)_V$. It follows that $y \in \mathrm{Dom}(T^*)$, and $T^*[y] = z$. Thus $G_{T^*} = \big(J\big(G_T\big)\big)^\perp$.

In a Hilbert space, the orthogonal complement of any set is closed. Since G_{T^*} is the orthogonal complement of $J(G_T)$ in $W \oplus V$, G_{T^*} is closed, and so T^* is a closed operator. Next, suppose that $z \in W$ is orthogonal to $\mathrm{Dom}(T^*)$. Then the pair $(z, 0)$ is clearly orthogonal in $W \oplus V$ to every pair $\big(y, T^*[y]\big)$ where $y \in \mathrm{Dom}(T^*)$. Thus $(z, 0) \in (G_{T^*})^\perp = J\big(G_T\big)$. Thus $(0, z) \in G_T$, and so $z = T(0) = 0$. It follows that $\mathrm{Dom}(T^*)$ is dense in H.

Finally, applying the first part of the lemma twice, we have

$$G_{(T^*)^*} = J^*\big((G_{T^*})^\perp\big) = J^*\big(J\big(G_T\big)\big) = G_T.$$

It follows that $\big((T^*)^*, D_{(T^*)^*}\big) = (T, \mathrm{Dom}(T))$. This completes the proof. □

If $(T, \mathrm{Dom}(T))$ is a closed, densely defined operator from V to W, the null space and range are denoted by

$$N(T) = \Big\{x \in \mathrm{Dom}(T)\,\Big|\,T[x] = 0\Big\};$$
$$R(T) = \Big\{y \in W\,\Big|\,\Big(\exists x \in \mathrm{Dom}(T)\Big)\,\Big(T[x] = y\Big)\Big\}.$$

Note that since $(T, \mathrm{Dom}(T))$ is closed, it follows that $N(T)$ is a closed subspace of V. There are the usual relationships between the null space and range of an operator and its adjoint.

PROPOSITION 2.5. *Let* $(T, \mathrm{Dom}(T))$ *be a closed, densely defined operator. Then*

$$R(T)^\perp = N(T^*) \qquad \textit{and} \quad R(T^*)^\perp = N(T).$$

PROOF. A vector $y \in R(T)^\perp$ if and only if the linear functional on $\mathrm{Dom}(T)$ given by $L_y[x] = (Tx, y)$ is identically zero and hence bounded. But this is equivalent to the condition that $y \in \mathrm{Dom}(T^*)$ and $T^*[y] = 0$. This establishes the first equality, and the second follows in the same way. □

We shall also need the following result.

LEMMA 2.6. *Let* $T : V \to W$ *be a closed, densely defined operator. Then the following statements are equivalent.*

(1) *There is a constant* $C > 0$ *so that for all* $x \in \mathrm{Dom}(T) \cap N(T)^\perp$ *we have*

$$\big|\big|\, x \,\big|\big|_V \leq C \,\big|\big|\, T(x) \,\big|\big|_W.$$

(2) *The range of* T *is a closed subspace of* W.

(3) *There is a constant* $C > 0$ *so that for all* $y \in \mathrm{Dom}(T^*) \cap N(T^*)^{\perp}$ *we have*

$$\big|\big|\, y \,\big|\big|_W \leq C \,\big|\big|\, T^*(y) \,\big|\big|_V.$$

(4) *The range of* T^* *is a closed subspace of* V.

Proof. Suppose that (1) holds, and let $\{y_n\}$ be a sequence in $R(T)$ which converges to a point $y_0 \in W$. For each n there exists $x_n \in \mathrm{Dom}(T) \cap N(T)^{\perp}$ with $y_n = T(x_n)$. Then by hypothesis, $\big|\big|\, x_m - x_n \,\big|\big|_V \leq C \,\big|\big|\, y_m - y_n \,\big|\big|$, and so $\{x_n\}$ is a Cauchy sequence in V, which converges to a point $x_0 \in V$. Since T is a closed operator, $y_0 = T[x_0] \in R(T)$, which shows that $R(T)$ is closed. Thus (1) implies (2), and a similar argument shows that (3) implies (4).

Next, suppose that (2) holds, so that $R(T) = N(T^*)^{\perp}$ is a Hilbert space. Let

$$E = \Big\{ y \in R(T) \,\Big|\, y \in \mathrm{Dom}(T^*) \text{ and } \big|\big|\, T^*(y) \,\big|\big|_V \leq 1 \Big\}.$$

Then statement (3) is equivalent to the statement that E is a bounded subset of $R(T)$. For any $z \in R(T)$ we can write $z = T[x]$, and so if $y \in E$ we have

$$|(z, y)_W| = |(T[x], y)_W| = |(x, T^*[y])_V| \leq \big|\big|\, x \,\big|\big|_V \,\big|\big|\, T^*[y] \,\big|\big|_V \leq \big|\big|\, x \,\big|\big|_V. \tag{2.3}$$

Let

$$V_N = \left\{ z \in R(T) \,\Big|\, \sup_{y \in E} |(z, y)_W| \leq N \big|\big|\, z \,\big|\big|_W \right\}$$

Then V_N is closed, and (2.3) shows that $z \in V_N$ as soon as $\big|\big|\, x \,\big|\big|_V \leq N \big|\big|\, z \,\big|\big|_W$. Thus $R(T) = \bigcup_{N=1}^{\infty} V_N$. By the Baire category theorem, there exists an integer N, a point $z_0 \in V_N$, and a constant $\epsilon > 0$ so that $\big|\big|\, z - z_0 \,\big|\big|_W \leq \epsilon$ implies $z \in V_N$. But then if $\big|\big|\, z \,\big|\big|_W \leq \epsilon$ and $y \in E$, both z_0 and $z - z_0$ belong to V_N so we have

$$\begin{aligned} |(z, y)_W| &\leq |(z_0, y)_W| + |(z - z_0, y)_W| \\ &\leq N \big|\big|\, z_0 \,\big|\big|_W + N \big|\big|\, z - z_0 \,\big|\big|_W \\ &\leq N\big(\big|\big|\, z_0 \,\big|\big|_W + \epsilon\big). \end{aligned}$$

But then if $y \in E$, we have

$$\begin{aligned} \big|\big|\, y \,\big|\big|_W &= \sup \Big\{ |(z, y)_W| \,\Big|\, \big|\big|\, z \,\big|\big|_W \leq 1 \Big\} \\ &= \epsilon^{-1} \sup \Big\{ |(z, y)_W| \,\Big|\, \big|\big|\, z \,\big|\big|_W \leq \epsilon \Big\} \\ &\leq \epsilon^{-1} N\big(\big|\big|\, z_0 \,\big|\big|_W + \epsilon\big). \end{aligned}$$

Thus E is bounded, and assertion (3) is established. The proof that assertion (4) implies assertion (1) is done the same way, and this completes the proof. □

2.2. The spectrum of an operator.

We now suppose that T is a linear transformation which maps a Hilbert space H to itself.

Definition 2.7. *Let* $(T, \mathrm{Dom}(T))$ *be a closed, densely defined operator on a Hilbert space* H. *The spectrum of* T, *denoted by* $\sigma(T)$, *is the complement in* $\mathbb{C}$ *of the set of numbers* λ *such that the operator* $T - \lambda I : \mathrm{Dom}(T) \to H$ *is one-to-one and onto, and such that the inverse operator* $(T - \lambda I)^{-1}$ *is a bounded operator.*

Remark 2.8. *It follows from the definition that for all* $\lambda \notin \sigma(T)$, *the range of the bounded operator* $(T - \lambda I)^{-1}$ *is the domain* $\mathrm{Dom}(T)$ *of the operator* T.

PROPOSITION 2.9. *Let* $(T, \mathrm{Dom}(T))$ *be a closed, densely defined operator on a Hilbert space* H. *Suppose that* $\lambda \notin \sigma(T)$. *If* $|\mu - \lambda| < \big|\big| (T - \lambda I)^{-1} \big|\big|^{-1}$, *then* $\mu \notin \sigma(T)$. *More generally, if neither* λ *nor* μ *belong to the spectrum of* T *then*

$$(T - \lambda I)^{-1}(T - \mu I)^{-1} = (T - \mu I)^{-1}(T - \lambda I)^{-1}$$

$$(T - \lambda I)^{-1} - (T - \mu I)^{-1} = -(\lambda - \mu)\,(T - \lambda I)^{-1}\,(T - \mu I)^{-1}$$

PROOF. We have the following identity of linear mappings on the domain $\mathrm{Dom}(T)$:

$$T - \mu I = T - \lambda I - (\mu - \lambda)I = \big[I - (\mu - \lambda)(T - \lambda I)^{-1}\big](T - \lambda I).$$

Our hypothesis is that $(T - \lambda I) : \mathrm{Dom}(T) \to H$ is one-to-one and onto, with bounded inverse. The operator $\big[I - (\mu - \lambda)(T - \lambda I)^{-1}\big]$ is defined on all of H and is bounded. If this operator is invertible, it follows that $T - \mu I : \mathrm{Dom}(T) \to H$ is one-to-one and onto, with inverse $(T - \lambda I)^{-1}\big[I - (\mu - \lambda)(T - \lambda I)^{-1}\big]^{-1}$. But if $|\mu - \lambda| < \big|\big| (T - \lambda I)^{-1} \big|\big|^{-1}$, it follows in the usual way that $\big[I - (\mu - \lambda)(T - \lambda I)^{-1}\big]$ is invertible, with the inverse given by the Neuman series

$$\big[I - (\mu - \lambda)(T - \lambda I)^{-1}\big]^{-1} = \sum_{j=0}^{\infty} (\mu - \lambda)^j [(T - \lambda I)^{-1}]^j.$$

We leave the verification of the two algebraic identities as an exercise. □

COROLLARY 2.10. *If* $(T, \mathrm{Dom}(T))$ *is a closed, densely defined operator on a Hilbert space* H, *then the spectrum* $\sigma(T)$ *is a closed subset of* $\mathbb{C}$.

2.3. Self-adjoint operators.

DEFINITION 2.11. *A closed, densely defined operator* $(T, \mathrm{Dom}(T))$ *on the Hilbert space* H *is self-adjoint if* $(T, \mathrm{Dom}(T)) = (T^*, \mathrm{Dom}(T^*))$.

PROPOSITION 2.12. *Suppose that* $(T, \mathrm{Dom}(T))$ *is a closed, densely defined self-adjoint operator on a Hilbert space* H.

(1) *If* $x \in \mathrm{Dom}(T)$, *it follows that* $(Tx, x) \in \mathbb{R}$.
(2) *The spectrum* $\sigma(T) \subset \mathbb{R}$, *and* $\big|\big| (T - \lambda I)^{-1} \big|\big| \leq \Im m[\lambda]^{-1}$.
(3) *If* $(Tx, x) \geq \lambda_0 \big|\big| x \big|\big|^2$ *for all* $x \in \mathrm{Dom}(T)$, *then* $\sigma(T) \subset [\lambda_0, \infty)$. *If* $\lambda < \lambda_0$, $\big|\big| (T - \lambda I) \big|\big| \leq (\lambda_0 - \lambda)^{-1}$.

PROOF. If $x \in \mathrm{Dom}(T)$, then $(Tx, x) = (x, Tx)$ because T is self-adjoint. But $(x, Tx) = \overline{(Tx, x)}$, which implies that $(Tx, x) \in \mathbb{R}$.

Next let $\lambda = a + ib$ with $b \neq 0$. To show that $\lambda \notin \sigma(T)$ is equivalent to showing that $bi \notin \sigma(T - a)$. However $T - a$ is another closed, densely defined self-adjoint operator. Thus in proving (2) it suffices to show that if $b \neq 0$ then $bi \notin \sigma(T)$. For $x \in \mathrm{Dom}(T)$ we have

$$\big|\big| (T \pm bi)[x] \big|\big|^2 = \big((T \pm bi)[x], (T \pm bi)[x]\big) = \big|\big| T[x] \big|\big|^2 + b^2 \big|\big| x \big|\big|^2.$$

and so

$$\big|\big| x \big|\big| \leq b^{-1} \big|\big| (T \pm bi)[x] \big|\big|. \tag{2.4}$$

It follows from this inequality that both operators $(T \pm bi)$ are one-to-one and have closed range. However if $y \in H$ is orthogonal to the range of $(T - bi)$, then by Proposition 2.5 and Remark 2.5 it follows that y is in the null space of

$(T-bi)^* = (T^*+bi) = (T+bi)$, which is one-to-one. Hence $y = 0$, and so the range of $(T-bi)$ is all of H. But then inequality (1.1) shows that $\big|\big| (T-bi)^{-1} \big|\big| \leq b^{-1}$.

Finally, suppose that $(Tx, x) \geq \lambda_0 \big|\big| x \big|\big|^2$ for all $x \in \mathrm{Dom}(T)$, and suppose that $\lambda < \lambda_0$. Then if $x \in \mathrm{Dom}(T)$,

$$(\lambda_0 - \lambda)\big|\big| x \big|\big|^2 \leq (T[x], x) - \lambda(x, x) = \big((T-\lambda I)[x], x\big) \leq \big|\big| (T-\lambda I)[x] \big|\big| \, \big|\big| x \big|\big|.$$

It follows that

$$\big|\big| x \big|\big| \leq (\lambda_0 - \lambda)^{-1}\big|\big| (T-\lambda I)[x] \big|\big|,$$

and the proof that $(T-\lambda I)$ is invertible then proceeds in the same way as the proof of (2). □

Proposition 2.13. *Suppose that $(T, \mathrm{Dom}(T))$ is a closed, densely defined self-adjoint operator on a Hilbert space H. Suppose also that T is one-to-one. Let $\mathrm{Dom}(S) = R(T)$ be the range of T, and let $S = T^{-1}$. Then $(S, \mathrm{Dom}(S))$ is also a closed, densely define sefl-adjoint linear operator on H.*

Proof. Since $G_S = R(G_T)$, it is clear that $(S, \mathrm{Dom}(S))$ is closed. Let $y \in H$ be orthogonal to $\mathrm{Dom}(S) = R(T)$ Then by Proposition 2.5, $y \in Null(T^*) = Null(T)$. Since T is one-to-one, it follows that $y = 0$. Thus $\mathrm{Dom}(S)$ is dense. Finally, by Lemma 2.4

$$\begin{aligned} G_{S^*} &= (J(G_S))^\perp = (J(R(G_T)))^\perp = (R(J(G_{T^*})))^\perp \\ &= (R(G_T)^\perp))^\perp = R(G_T^{\perp\perp}) = R(G_T) = G_S. \end{aligned}$$

and hence $(S, \mathrm{Dom}(S))$ is self-adjoint. This completes the proof. □

2.4. The spectral theorem for self-adjoint operators. We now state one version of the spectral theorem for self-adjoint operators on a Hilbert space H.

Denote by $\mathcal{B}$ the algebra of all bounded, complex-valued Borel measurable functions defined on $\mathbb{R}$. It is important to note that the elements of $\mathcal{B}$ are functions, and not equivalence classes of functions which differ only on a set of measure zero. If $f \in \mathcal{B}$, we write

$$\big|\big| f \big|\big| = \sup_{x \in \mathbb{R}} |f(x)|.$$

If $\{f_n\} \subset \mathcal{B}$ is a sequence, then $f_n \to f_0$ uniformly if $\lim_{n\to\infty} \big|\big| f_n - f_0 \big|\big| = 0$. We also say that a sequence $\{f_n\} \subset \mathcal{B}$ converges monotonically to $f_0 \in \mathcal{B}$ if for every $x \in \mathbb{R}$

$$f_n(x) \leq f_0(x) \qquad \text{and} \qquad \lim_{n\to\infty} f_n(x) = f_0(x).$$

If H is a Hilbert space with norm $\big|\big| \cdot \big|\big|_H$, let $\mathcal{L}(H)$ denote the algebra of bounded linear transformations from H to H. As usual, if $T \in \mathcal{L}(H)$,

$$\big|\big| T \big|\big| = \sup \left\{ \big|\big| T[x] \big|\big|_H \,\Big|\, \big|\big| x \big|\big|_H \leq 1 \right\}.$$

If $\{T_n\} \subset \mathcal{L}(H)$ is a sequence of bounded operators, we say that $T_n \to T_0$ in norm if $\lim_{n\to\infty} \big|\big| T_n - T_0 \big|\big| = 0$. We say that a sequence $\{T_n\} \subset \mathcal{L}(H)$ converges strongly to an operator $T_0 \in \mathcal{L}(H)$ for every $x \in H$ it follows that $\lim_{n\to\infty} \big|\big| T_n[x] - T_0[x] \big|\big|_H = 0$.

Theorem 2.14. *Let $(T, \mathrm{Dom}(T))$ be a closed, densely defined, self-adjoint operator on a Hilbert space H. There is a unique algebra homomorphism from $\mathcal{B}$ to $\mathcal{L}(H)$ which we write $f \to f(T)$ with the following properties:*

(1) *If* $\lambda \in \mathbb{C} - \mathbb{R}$ *and if* $f_\lambda(t) = (t-\lambda)^{-1}$, *then* $f_\lambda(T) = (T - \lambda I)^{-1}$.

(2) *If* $f \in \mathcal{B}$, *then* $\overline{f}(T) = \big(f(T)\big)^*$.

(3) *If* $f \in \mathcal{B}$ *then* $\big|\big| f(T) \big|\big| \leq \big|\big| F \big|\big|$.

(4) *If* $\{f_n\} \subset \mathcal{B}$ *is a sequence which converges monotonically to* $f_0 \in \mathcal{B}$, *then the sequence of bounded operators* $\{f_n(T)\} \subset \mathcal{L}(H)$ *converges strongly to* $f_0(T)$.

(5) *If* $f \in \mathcal{B}$ *and if suppt*$(f) \cap \sigma(T) = \emptyset$, *then* $f(T) = 0$.

3. The initial value problem for the heat operator $\partial_t + \mathcal{L}_x$

We now turn to the first step in the construction and estimation of a fundamental solution for $\mathcal{L}$, which is the study of the initial value problem for the heat operator $\partial_t + \mathcal{L}_x$ on $\mathbb{R} \times \Omega$ given in equations (1.3) and (1.4). We want to solve this problem by using the spectral theorem to define the operators $\{e^{-t\mathcal{L}}\}$. In order to do this, we need to construct a self-adjoint operator $(\mathcal{L}, D_\mathcal{L})$ on the Hilbert space $L^2(\Omega)$.

3.1. The Friedrich's construction.

In order to define a self-adjoint operator $(\mathcal{L}, D_\mathcal{L})$, , we begin by restricting our attention to the dense subspace $\mathcal{C}_0^\infty(\Omega) \subset L^2(\Omega)$. Note that for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$, we have

$$\big(\mathcal{L}[\varphi], \varphi\big) = \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|^2 \geq 0.$$

Using this positivity, we follow an argument of Friedrichs [**Fri34**] to show that there is a closed, densely defined self-adjoint extension of the operator $\big(\mathcal{L}, \mathcal{C}_0^\infty(\Omega)\big)$.

THEOREM 3.1. *There is an extension of the operator* $\big(\mathcal{L}, \mathcal{C}_0^\infty(\Omega)\big)$ *to a self-adjoint operator* $\big(\widetilde{\mathcal{L}}, D_\mathcal{L}\big)$ *with* $\mathcal{C}_0^\infty(\Omega) \subset D_\mathcal{L}$. *If* $f \in D_\mathcal{L}$, *then* $\mathcal{L}[f] = \widetilde{\mathcal{L}}[f]$ *in the sense of distributions. Moreover, if* $f \in D_\mathcal{L}$ *then:*

(1) $X_j[f] \in L^2(\Omega)$ *in the sense of distributions for* $1 \leq j \leq p$;

(2) $X_j[f] \in L^2(\Omega)$ *in the strong sense; that is, there is a sequence* $\{\varphi_n\} \subset \mathcal{C}_0^\infty(\Omega)$ *such that* $\varphi_n \to f$ *and* $X_j[\varphi] \to X_j[f]$ *in* $L^2(\Omega)$;

(3) $\big(\mathcal{L}[f], f\big) = \sum_{j=1}^{p} \big|\big| X_j[f] \big|\big|^2$.

REMARK 3.2. *Since* $\widetilde{\mathcal{L}}[f] = \mathcal{L}[f]$ *in the sense of distributions, we shall usually simplify notation and write* $\mathcal{L}$ *instead of* $\widetilde{\mathcal{L}}$ *unless it is critical to distinguish the two operators.*

PROOF. For $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$, put

$$Q(\varphi, \psi) = \sum_{j=1}^{p} (X_j[\varphi], X_j[\psi]) + (\varphi, \psi) = (\mathcal{L}[\varphi] + \varphi, \psi)$$

so that

$$Q(\varphi,\varphi) = \sum_{j=1}^{p} \big|\big| X_j[\varphi] \big|\big|^2 + \big|\big| \varphi \big|\big|^2 = \big(\mathcal{L}[\varphi], \varphi\big) + \big|\big| \varphi \big|\big|^2.$$

Q is a positive-definite Hermitian quadratic form on $\mathcal{C}_0^\infty(\Omega)$, and hence

$$|||\varphi||| = Q(\varphi,\varphi)^{\frac{1}{2}}$$

defines a norm, and makes $\mathcal{C}_0^\infty(\Omega)$ into a pre-Hilbert space. For $\varphi \in \mathcal{C}_0^\infty(\Omega)$ we clearly have

$$\big|\big| \varphi \big|\big| \leq |||\varphi|||.$$

Let W be the completion of $\mathcal{C}_0^\infty(\Omega)$ with respect to the norm $||| \cdot |||$, so that W is a Hilbert space. We will continue to write the norm in W as $||| \cdot |||$, and the inner product as $Q(\,\cdot\,,\cdot\,)$. We show that W can be identified with a subspace of H which contains $\mathcal{C}_0^\infty(\Omega)$.

Let $F \in W$. Since W is the completion of $\mathcal{C}_0^\infty(\Omega)$, there exists a sequence $\{\varphi_n\} \subset \mathcal{C}_0^\infty(\Omega)$ so that $\lim_{n\to 0} |||\varphi_n - F|||^2 = 0$. Thus $\{\varphi_n\}$ is a Cauchy sequence with respect to the norm $||| \cdot |||$. It follows from the definition that the sequence $\{\varphi_n\}$ and each of the sequences $\{X_j[\varphi_n]\}$ for $1 \leq j \leq p$ are Cauchy sequences in $L^2(\Omega)$. Let $i[F] = f = \lim_{n\to\infty} \varphi_n$ and $g_j = \lim_{n\to\infty} X_j[\varphi_n]$ be the limits of these sequences in $L^2(\Omega)$. It is easy to check that the functions $\{i[F] = f, g_1, \ldots, g_p\}$ depend only on $F \in W$ and not on the choice of approximating sequence $\{\varphi_n\}$. In particular, we have defined a mapping $i : W \to L^2(\Omega)$, which is clearly linear. Because we have convergence in $L^2(\Omega)$, if $\psi \in \mathcal{C}_0^\infty(\Omega)$ we have

$$\begin{aligned}
\int_\Omega g_j(x)\,\psi(x)\,dx &= \lim_{n\to\infty} \int_\Omega X_j[\varphi_n](x)\,\psi(x)\,dx \\
&= \lim_{n\to\infty} \int_\Omega \varphi_n(x)\,X_j^*[\psi](x)\,dx \\
&= \int_\Omega f(x)\,X_j^*[\psi](x)\,dx
\end{aligned}$$

But this means that $X_j[f] = g_j$ in the sense of distributions.

In particular, if $i[F] = f = 0$ we must have $g_j = X_j[f] = 0$ for each j. But then if $\{\varphi_n\} \subset \mathcal{C}_0^\infty(\Omega)$ is a sequence which converges to F in W, we have $\varphi_n \to 0$ and $X_j[\varphi_n] \to 0$ in $L^2(\Omega)$. Thus $|||\varphi_n||| \to 0$. But this means that $F = 0$, so the mapping $i : W \to L^2(\Omega)$ is one-to-one. We can thus identify W with its image $i[W]$ in $L^2(\Omega)$. We have shown that $\mathcal{C}_0^\infty(\Omega) \subset W$, and if $f \in W$, then $X_j[f] \in L^2(\Omega)$ strongly and in the sense of distributions.

For each $g \in L^2(\Omega)$, define a linear functional $L_g : W \to \mathbb{C}$ by setting $L_g[f] = \int_\Omega f(x)\,\overline{g(x)}\,dx$. We have

$$\big|L_g[f]\big| = \big|(f,g)\big| \leq \big|\big| f \big|\big|\,\big|\big| g \big|\big| \leq \big|\big| g \big|\big|\,|||f|||.$$

Thus L_g is a bounded linear functional on the Hilbert space W, and by the Riesz representation theorem there exists a unique element $B[g] \in W$ with $\Big|\Big| B[g] \Big|\Big| \leq \big|\big| g \big|\big|$ so that for all $f \in W$,

$$(f,g) = L_g[f] = Q(f, B[g]). \tag{3.1}$$

Clearly B is linear, and since $\big|\big|\,B[g]\,\big|\big| \leq \Big|\Big|\,B[g]\,\Big|\Big| \leq \big|\big|\,g\,\big|\big|$, it follows that $B : L^2(\Omega) \to W$ is a bounded linear transformation. Since for any $g_1, g_2 \in L^2(\Omega)$ we know that $B[g_j] \in W$, we can apply (3.1) to get

$$(B[g_1], g_2) = Q(B[g_1], B[g_2]) = \overline{Q(B[g_2], B[g_1])} = \overline{(B[g_2], g_1)} = (g_1, B[g_2]).$$

Thus B is a self-adjoint operator. Also $(B[g], g) = Q(B[g], B[g]) \geq 0$, so B is non-negative. Finally, if $B[g] = 0$, then for all $f \in W$ we have $0 = Q(f, B[g]) = (f, g)$, and so g is orthogonal to W. Since W contains $\mathcal{C}_0^\infty(\Omega)$ and therefore is dense, it follows that $g = 0$. Thus B is one-to-one.

Let $D_{\mathcal{L}} \subset W$ be the range of B. We show that $\mathcal{C}_0^\infty(\Omega) \subset D_{\mathcal{L}}$ and that

$$B\big[\mathcal{L}[\psi] + \psi\big] = \psi \tag{3.2}$$

for $\psi \in \mathcal{C}_0^\infty(\Omega)$. Put $g = \mathcal{L}[\psi] + \psi$. Let $\varphi \in \mathcal{C}_0^\infty(\Omega) \subset W$. Then from equation (3.1) we have

$$Q(\varphi, B[g]) = (\varphi, g) = (\varphi, \mathcal{L}[\psi] + \psi) = Q(\varphi, \psi).$$

Thus $B[g] - \psi$ is orthonogal in the Hilbert space W to the subspace $\mathcal{C}_0^\infty(\Omega)$. But since W was the completion of $\mathcal{C}_0^\infty(\Omega)$, it follows that $\mathcal{C}_0^\infty(\Omega)$ is dense in W, and so $B[g] = \psi$. Thus $\psi \in D_{\mathcal{L}}$ and we have verified equation (3.2).

On $\mathcal{C}_0^\infty(\Omega)$, the operator $\mathcal{L} + I$ is the same as the operator B^{-1}. Thus the operator $(B^{-1}, D_{\mathcal{L}})$ is an extension of the operator $(\mathcal{L} + I, \mathcal{C}_0^\infty(\Omega))$. Since B is self adjoint, it follows from Proposition 2.13 that $(B^{-1}, D_{\mathcal{L}})$ is self adjoint, and is an extension of the operator $\mathcal{L} + I$ on the space $\mathcal{C}_0^\infty(\Omega)$. It follows that $\widetilde{\mathcal{L}} = B^{-1} - I$ is a self-adjoint extension of the operator $(\mathcal{L}, \mathcal{C}_0^\infty(\Omega))$.

It remains to show that if $f \in D_{\mathcal{L}}$, then $\widetilde{\mathcal{L}}[f] = \mathcal{L}[f]$ in the sense of distributions. Let $\psi \in \mathcal{C}_0^\infty(\Omega)$. Then since $\mathcal{C}_0^\infty(\Omega) \subset D_{\mathcal{L}}$ and $(B^{-1} - I)$ is self-adjoint and equals $\mathcal{L}$ on $\mathcal{C}_0^\infty(\Omega)$ we have

$$\big(\widetilde{\mathcal{L}}[f], \psi\big) = \big((B^{-1} - I)[f], \psi\big) = \big(f, (B^{-1} - I)[\psi]\big) = \big(f, \mathcal{L}[\psi]\big).$$

This then completes the proof. □

REMARK 3.3. *As discussed at the beginning of Section* **??**, *this argument uses the Riesz representation theorem to produce a solution of the equation* $T[f] = \mathcal{L}[f] + f = g$ *as a consequence of the estimate* $\big|\big|\,\varphi\,\big|\big|^2 \leq \Big|\Big|\,\varphi\,\Big|\Big|^2 = \big(\mathcal{L}[\varphi], \varphi\big) + \big|\big|\,\varphi\,\big|\big|^2$.

The choice of a domain for a differential operator can encode boundary behavior of the functions in the domain. At least when Ω has smooth boundary, we now show that smooth elements of $\mathcal{D}_{\mathcal{L}}$ must vanish at non-characteristic points of $\partial\Omega$. Suppose that $\rho \in \mathcal{C}^\infty(\mathbb{R}^n)$, that $|\nabla\rho(x)| = 1$ when $\rho(x) = 0$, and that

$$\Omega = \Big\{x \in \mathbb{R}^n \,\Big|\, \rho(x) < 0\Big\}.$$

Suppose that $f \in \mathcal{C}^\infty(\overline{\Omega})$ and $X \in T(\overline{\Omega})$. (This means that f and the coefficients of the vector field X are infinitely differentiable in a neighborhood of the closure of Ω). Let $d\sigma$ denote surface area measure on $\partial\Omega$ The divergence theorem asserts that

$$\int_\Omega X[f](x)\,dx = \int_{\partial\Omega} X[\rho](y)\,f(y)\,d\sigma(y). \tag{3.3}$$

We say that a point $y \in \partial\Omega$ is non-characteristic for the vector field X if $X[\rho](y) \neq 0$.

LEMMA 3.4. *Suppose that $f \in \mathcal{C}^\infty(\overline{\Omega}) \cap D_{\mathcal{L}}$. Then $f(y) = 0$ for every point $y \in \partial\Omega$ which is non-characteristic for at least one of the vector fields $\{X_1, \dots, X_p\}$.*

PROOF. If $f \in \mathcal{C}^\infty(\overline{\Omega})$, then for any $g \in \mathcal{C}^\infty(\overline{\Omega})$ we have

$$\begin{aligned}
(X_j[f], g)_{L^2(\Omega)} &= \int_\Omega X_j[f](x)\,\overline{g(x)}\,dx \\
&= \int_\Omega X_j[f\,\overline{g}](x)\,dx + \int_\Omega f(x)\,\overline{X_j^*[g](x)}\,dx \\
&= (f, X_j^*[g]) + \int_{\partial\Omega} f(y)\,X_j[\rho](y)\,\overline{g(y)}\,d\sigma(y).
\end{aligned}$$

On the other hand, if $f \in D_{\mathcal{L}}$, there exists $\{\varphi_n\} \subset \mathcal{C}_0^\infty(\Omega)$ with $\varphi_n \to f$ and $X_j[\varphi_n] \to X_j[f]$ for $1 \leq j \leq p$, with convergence in $L^2(\Omega)$. Thus we have

$$\begin{aligned}
(X_j[f], g)_{L^2(\Omega)} &= \lim_{n\to\infty} \int_\Omega X_j[\varphi_n](x)\,\overline{g(x)}\,dx \\
&= \lim_{n\to\infty} \left[\int_\Omega X_j[\varphi_n\,\overline{g}](x)\,dx + \int_\Omega \varphi_n(x)\,\overline{X_j^*[g](x)}\,dx\right] \\
&= \lim_{n\to\infty} (\varphi_n, X_j^*[g]) \\
&= (f, X_j^*[g]).
\end{aligned}$$

It follows that

$$\int_{\partial\Omega} f(y)\,X_j[\rho](y)\,\overline{g(y)}\,d\sigma(y) = 0$$

for every $g \in \mathcal{C}^\infty(\overline{\Omega})$, and hence $f(y) = 0$ at every point at which $X_j[\rho](y) \neq 0$. □

COROLLARY 3.5. *If $f \in D_{\mathcal{L}}$ and $\mathcal{L}[f] = 0$, then $f = 0$.*

PROOF. If $\mathcal{L}[f] = 0$, then since $\mathcal{L}$ is hypoelliptic, it follows that $f \in \mathcal{C}^\infty(\Omega)$. On the other hand, since $f \in D_{\mathcal{L}}$ it follows that $0 = (\mathcal{L}[f], f) = \sum_{j=1}^p \big|\big|\, X_j[f] \,\big|\big|^2$, and hence $X_j[f] = 0$ for $1 \leq j \leq p$. Since the vector fields $\{X_1, \dots, X_p\}$ are of finite type, it follows that $\partial_{x_k}[f] = 0f$ on Ω for $1 \leq k \leq n$, and hence that f is a constant. But then $f \in \mathcal{C}^\infty(\overline{\Omega})$, and so $f = 0$ at every non-characteristic boundary point. Since this set is non-empty, it follows that $f \equiv 0$. □

3.2. The heat semigroup $\{e^{-t\mathcal{L}}\}$.

We can now apply the spectral theorem to the self-adjoint operator $(\mathcal{L}, D_{\mathcal{L}})$. Note that by part (3) of Theorem 3.1, $(\mathcal{L}[f], f) \geq 0$ for every $f \in D_{\mathcal{L}}$, and hence by part (3) of Proposition 2.12, $\sigma(\mathcal{L}) \subset [0, \infty)$. Thus the spectral theroem (Theorem 2.14) guarantees the existence of a unique algebra homomorphism $f \to f(\mathcal{L})$ from the algebra $\mathcal{B} = \mathcal{B}[0, \infty)$ of bounded Borel functions on the non-negative real axis to the algebra $\mathcal{L}\big(L^2(\Omega)\big)$ of bounded linear operators on $L^2(\Omega)$ with the following properties:

(1) If $\lambda \in \mathbb{C} - \mathbb{R}$ and if $f_\lambda(t) = (t - \lambda)^{-1}$, then $f_\lambda(\mathcal{L}) = (\mathcal{L} - \lambda I)^{-1}$.

(2) If $f \in \mathcal{B}$ is complex-valued, then $\overline{f}(\mathcal{L}) = \big(f(\mathcal{L})\big)^*$. In particular, if $f \in \mathcal{B}$ is real valued, then $f(\mathcal{L})$ is self-adjoint.

(3) If $f \in \mathcal{B}$ then $\big|\big|\, f(\mathcal{L}) \,\big|\big| \leq \sup_{x \geq 0} |f(x)|$.

(4) If $\{f_n\} \subset \mathcal{B}$ is a sequence which converges monotonically to $f_0 \in \mathcal{B}$, then the sequence of bounded operators $\{f_n(\mathcal{L})\} \subset \mathcal{L}(H)$ converges strongly to $f_0(\mathcal{L})$.

For each $t \geq 0$, let $e_t(x) = e^{-tx}$ for $x \geq 0$. Then each function e_t is a bounded continuous function on $[0, \infty)$. Let $e^{-t\mathcal{L}}$ denote the corresponding bounded linear operator on $L^2(\Omega)$ given by the spectral theorem. The family of operators $\{e^{-t\mathcal{L}}\}$ for $t \geq 0$ forms what is known as a strongly continuous semigroup of contractions, called the *heat semigroup* for $\mathcal{L}$. The use of the terms "strongly continuous", "semigroup", and "contractions" is justified in the following proposition.

PROPOSITION 3.6. *The family of bounded linear operators* $\left\{e^{-t\mathcal{L}}\right\}_{t\geq 0}$ *on* $L^2(\Omega)$ *have the following properties:*

(1) (Strongly continuous) For each $f \in L^2(\Omega)$*, the mapping* $t \to e^{-t\mathcal{L}}[f]$ *from* $[0, \infty)$ *to* $L^2(\Omega)$ *is continuous.*

(2) (Semigroup) For $t_1, t_2 \geq 0$*,* $e^{-t_1\mathcal{L}}\, e^{-t_2\mathcal{L}} = e^{-(t_1+t_2)\mathcal{L}}$*. In particular, the operator* $e^{-0\mathcal{L}} = I$ *is the identity operator.*

(3) (Contractions) For each $t \geq 0$ *and each* $f \in L^2(\Omega)$*,* $\big|\big|\, e^{-t\mathcal{L}}[f] \,\big|\big| \leq \big|\big|\, f \,\big|\big|$*.*

PROOF. The functions $\{e_t\}$ converge monotonically to e_{t_0} as $t \to t_0$ from above. Similarly, the functions $\{-e_t\}$ converge monotonically to $-e_{t_0}$ as $t \to t_0$ from below. Thus the continuity of $t \to e^{-t\mathcal{L}}[f]$ follows from statement (4) in Theorem 2.14. The semigroup property follows from the fact that the functions $\{e_t\}$ satisfy $e_{t_1}\, e_{t_2} = e_{t_1+t_2}$, and from the fact that the correspondence $e_t \to e^{-t\mathcal{L}}$ given by the spectral theorem is an algebra homomorphism. Finally the norm estimate follows from the fact that $|e_t(x)| \leq 1$ and statement (3) in Theorem 2.14. □

Before stating the important properties of the heat semigroup, we introduce the definition of the domain of the operator $\mathcal{L}^N$ when $N \geq 1$. Of course, the domain of $\mathcal{L}^1 = \mathcal{L}$ is the subspace $D_{\mathcal{L}}$. Assuming that we have defined the domain of $\mathcal{L}^N$ as a subspace $D_{\mathcal{L}^N} \subset D_{\mathcal{L}}$, the domain of $\mathcal{L}^{N+1}$ is the subspace of elements $f \in D_{\mathcal{L}^N}$ such that $\mathcal{L}[f] \in D_{\mathcal{L}^N}$. It is then easy to check that the domain of $\mathcal{L}^N$ is the range of the bounded linear operator $\left[(\mathcal{L} - \lambda I)^{-1}\right]^N$ for any $\lambda \notin \mathbb{R}$.

THEOREM 3.7. *The heat semigroup* $\left\{e^{-t\mathcal{L}}\right\}_{t\geq 0}$ *has the following properties:*

(1) For all $f \in L^2(\Omega)$*,*

$$\lim_{t\to 0^+} e^{-t\mathcal{L}}[f] = f,$$
$$\lim_{t\to +\infty} e^{-t\mathcal{L}}[f] = 0,$$

with convergence in the norm in $L^2(\Omega)$*.*

(2) If $f \in D_{\mathcal{L}}$ *then*

$$\big|\big|\, e^{-t\mathcal{L}}[f] - f \,\big|\big|_{L^2(\Omega)} \leq t \,\big|\big|\, \mathcal{L}[f] \,\big|\big|_{L^2(\Omega)}.$$

(3) *For all $t > 0$ and all positive integers N, the range of $e^{-t\mathcal{L}}$ is contained in the the domain of $\mathcal{L}^N$. The operator $\mathcal{L}^N\, e^{-t\mathcal{L}}$ is a bounded operator on $L^2(\Omega)$ and*

$$\big|\big|\, \mathcal{L}^n\, e^{-t\mathcal{L}}[f]\,\big|\big|_{L^2(\Omega)} \leq \left(\frac{N}{e}\right)^N\, t^{-N}\,\big|\big|\, f\,\big|\big|_{L^2(\Omega)}.$$

(4) *If $f \in D_{\mathcal{L}}$ then $e^{-t\mathcal{L}}\big[\mathcal{L}[f]\big] = \mathcal{L}\big[e^{-t\mathcal{L}}[f]\big]$.*

(5) *For $f \in L^2(\Omega)$, the mapping $t \to e^{-t\mathcal{L}}[f]$ is differentiable for $t > 0$, and $\dfrac{d}{dt}\left(e^{-t\mathcal{L}}[f]\right) = -\mathcal{L}\left[e^{-t\mathcal{L}}[f]\right]$. In other words, if $t > 0$ and $f \in L^2(\Omega)$,*

$$\lim_{h\to 0}\big|\big|\, h^{-1}\big[e^{-(t+h)\mathcal{L}}[f] - e^{-t\mathcal{L}}[f]\big] + \mathcal{L}\big[e^{-t\mathcal{L}}[f]\big]\,\big|\big| = 0.$$

PROOF. The first equality in statement (1) follows from Proposition 3.6. Let

$$e_\infty(x) = \begin{cases} 0 & \text{if } x > 0, \\ 1 & \text{if } x = 0. \end{cases}$$

Then the functions $\{-e_t\}$ converge monotonically to $-e_\infty$ as $t \to +\infty$, so by strong continuity, for each $f \in L^2(\Omega)$, $\lim_{t\to+\infty} e^{-t\mathcal{L}}[f] = e_\infty(\mathcal{L})[f]$. But $e_\infty(\mathcal{L})$ is the orthogonal projection onto the null space of $\mathcal{L}$. This completes the proof of statement (1).

According to Remark 2.8, the domain of $\mathcal{L}$ is the range of $(\mathcal{L}-\lambda I)^{-1}$ if $\lambda \notin \sigma(\mathcal{L})$. Thus if $f \in D_{\mathcal{L}}$ there exists $g \in L^2(\Omega)$ with $f = (\mathcal{L} + iI)^{-1}[g]$, or

$$\mathcal{L}[f] = g - if = \big(I - i(\mathcal{L} + iI)^{-1}\big)[g].$$

We also have

$$e^{-t\mathcal{L}}[f] - f = \big[e^{-t\mathcal{L}}(\mathcal{L} + iI)^{-1} - (\mathcal{L} + iI)^{-1}\big][g] = (e^{-t\mathcal{L}} - I)(\mathcal{L} + iI)^{-1}[g].$$

Put

$$\begin{aligned} F_1(x) &= (1 - i(x+i)^{-1}) = \frac{x}{x+i} \\ F_2(x) &= \frac{e^{-tx} - 1}{x+i} \\ G(x) &= t\left(\frac{e^{-tx} - 1}{tx}\right). \end{aligned}$$

These are all bounded Borel functions on the spectrum of $\mathcal{L}$, and since $F_2 = GF_1$, it follows from the spectral theorem that $F_2(\mathcal{L}) = G(\mathcal{L})\,F_1(\mathcal{L})$. Thus $\big(e^{-t\mathcal{L}} - I\big)[f] = G(\mathcal{L})\big[\mathcal{L}[f]\big]$. It follows that

$$\big|\big|\,\big(e^{-t\mathcal{L}} - I\big)[f]\,\big|\big|_{L^2(\Omega)} \leq \big|\big|\, G(\mathcal{L})\,\big|\big|\,\big|\big|\,\mathcal{L}[f]\,\big|\big|.$$

But

$$\big|\big|\, G(\mathcal{L})\,\big|\big| \leq \sup_{x\geq 0}|G(x)| = t,$$

and this establishes part (2).

The function $e^{-tx} = [(x+i)^{-1}]^N\big[(x+i)^N e^{-tx}\big]$, and since $F_N(x) = (x+i)^N e^{-tx}$ is a bounded Borel function on $[0, \infty)$, it follows that $e^{-t\mathcal{L}} = [(\mathcal{L} + iI)^{-1}]^N F_N(\mathcal{L})$. Thus the range of $e^{-t\mathcal{L}}$ is contained in the range of $[(\mathcal{L} + iI)^{-1}]^N$, which is the domain of $\mathcal{L}^N$.

The function $G_N(x) = x^N e^{-tx}$ is a bounded Borel function on $[0, \infty)$. We show by induction that $G_N(\mathcal{L}) = \mathcal{L}^N e^{-t\mathcal{L}}$. The case $N = 0$ is just the definition of $e^{-t\mathcal{L}}$. Assume that the statement is true for some $N \geq 0$. Then we have

$$G_N(x) = (x+i)^{-1}G_{N+1}(x) + i(x+i)^{-1}G_N(x)$$

and hence

$$\mathcal{L}^N e^{-t\mathcal{L}} = (\mathcal{L} + iI)^{-1}G_{N+1}(\mathcal{L}) + i(\mathcal{L} + iI)^{-1}\mathcal{L}^N e^{-t\mathcal{L}}.$$

We can apply $(\mathcal{L} + iI)$ to both sides, and conclude that $G_{N+1}(\mathcal{L}) = \mathcal{L}^{N+1}e^{-t\mathcal{L}}$. But then

$$\big|\big|\, \mathcal{L}^N e^{-t\mathcal{L}}[f]\, \big|\big|_{L^2(\Omega)} \leq \Big[\sup_{x\in[0,\infty)} G_N(x)\Big]\, \big|\big|\, f\, \big|\big|_{L^2(\Omega)} = \left(\frac{N}{e}\right)^N t^{-N}\, \big|\big|\, f\, \big|\big|_{L^2(\Omega)}.$$

This proves statement (3).

To establish statement (4), let $f \in D_{\mathcal{L}}$, and write $f = (\mathcal{L} + iI)^{-1}[g]$ with $g \in L^2(\Omega)$. We have already observed that if if $F_1(x) = x(x+i)^{-1}$, then $F_1(\mathcal{L})[g] = \mathcal{L}[f]$. But then since

$$\big[e^{-tx}\big]\,\big[\frac{x}{x+i}\big] = \big[x\, e^{-tx}\big]\,\big[\frac{1}{x+i}\big],$$

we have $e^{-t\mathcal{L}}\, F_1(\mathcal{L}) = \mathcal{L}\, e^{-t\mathcal{L}}(\mathcal{L} + iI)^{-1}$, and so

$$e^{-t\mathcal{L}}\big[\mathcal{L}[f]\big] = e^{-t\mathcal{L}}F_1(\mathcal{L})[g] = \mathcal{L}\, e^{-t\mathcal{L}}(L + iI)^{-1}[g] = \mathcal{L}\, e^{-t\mathcal{L}}[f].$$

Statement (5) follows from the observation that $\frac{d}{dt}\big(e^{-tx}\big) = -xe^{-tx}$,and for $t \neq 0$ the difference quotients which approximate the derivative converge monotonically to the derivative. Thus for each $f \in L^2(\Omega)$ we have

$$\lim_{h\to 0} \frac{e^{-(t+h)\mathcal{L}}[f] - e^{-t\mathcal{L}}[f]}{t} = -\mathcal{L}e^{-t\mathcal{L}}[f].$$

□

The heat semigroup gives us a solution to the initial-value problem posed in equations (1.3) and (1.4).

THEOREM 3.8. *For each $f \in L^2(\Omega)$, let $u(x,t) = e^{-t\mathcal{L}}[f](x)$. Then $u \in \mathcal{C}^\infty\big((0,\infty) \times \Omega\big)$, and satisfies the following two properties:*

(1) $\big[\partial_t + \mathcal{L}_x\big][u](t,x) = 0$ *for $t > 0$ and $x \in \Omega$;*

(2) $\displaystyle\lim_{t\to 0^+} \int_\Omega |u(x,t) - f(x)|^2\, dx = 0.$

PROOF. Define a distribution Ψ_f on $(0,\infty) \times \Omega$ as follows. If $\psi \in \mathcal{C}_0^\infty\big((0,\infty) \times \Omega\big)$, set

$$\Big\langle \Psi_f, \psi \Big\rangle = \int_0^\infty \left[\int_\Omega e^{-t\mathcal{L}}[f](x)\, \psi(t,x)\, dx\right] dt = \int_0^\infty \big(e^{-t\mathcal{L}}[f], \overline{\psi}_t\big)_{L^2(\Omega)}\, dt$$

where $\psi_t(x) = \psi(t,x)$. In fact, both $t \to e^{-t\mathcal{L}}[f]$ and $t \to \psi_t$ are differentiable functions with values in $L^2(\Omega)$, and so the function $t \to \big(e^{-t\mathcal{L}}[f], \overline{\psi}_t\big)_{L^2(\Omega)}$ is differentiable with compact support in $(0,\infty)$. If the support of ψ is contained in the set $\{(t,x)\,\big|\,0<\alpha\le t\le\beta<\infty\}$, the crude estimate

$$\left|\Big\langle \Psi_f,\psi\Big\rangle\right| \le (\beta-\alpha)\,\big|\big|\,f\,\big|\big|_{L^2(\Omega)}\,\sup_{\alpha\le t\le\beta}\big|\big|\,\psi_t\,\big|\big|_{L^2(\Omega)}$$

shows that this linear functional is continuous. Next, since $(e^{-t\mathcal{L}}[f],\overline{\psi}_t)$ is differentiable as a function of t and has compact support in $(0,\infty)$. Recall that $\partial_t[e^{-t\mathcal{L}}[f]] = -\mathcal{L}\big[e^{-t\mathcal{L}}[f]\big]$. Then using the fundamental theorem of calculus and the product rule, we have for every $\psi\in\mathcal{C}_0^\infty\big((0,\infty)\times\Omega\big)$

$$\begin{aligned}
0 &= -\int_0^\infty \frac{d}{dt}[(e^{-t\mathcal{L}}[f],\overline{\psi}_t)]\,dt\\
&= \int_0^\infty \big(\mathcal{L}_x\big[e^{-t\mathcal{L}}[f]\big],\overline{\psi}_t\big)\,dt - \int_0^\infty \big(e^{-t\mathcal{L}}[f],\overline{\partial_t\psi_t}\big)\,dt\\
&= \int_0^\infty \big(e^{-t\mathcal{L}}[f],\overline{\mathcal{L}_x\psi_t}\big)\,dt - \int_0^\infty \big(e^{-t\mathcal{L}}[f],\overline{\partial_t\psi_t}\big)\,dt\\
&= \int_0^\infty \big(e^{-t\mathcal{L}}[f],[-\partial_t+\mathcal{L}_x][\psi_t]\big)\,dt\\
&= \Big\langle \Psi_f,[-\partial_t+\mathcal{L}_x][\psi]\Big\rangle
\end{aligned}$$

Thus $[\partial_t+\mathcal{L}][\Psi_f]=0$ in the sense of distributions. However, the operator

$$\partial_t+\mathcal{L} = \partial_t - \sum_{j=1}^p X_j^2 - \sum_{j=1}^p (\nabla\cdot X_j)\,X_j$$

is of the type studied in Chapter A, and is hypoelliptic. It follows that $u(t,x) = e^{-t\mathcal{L}}[f](x)$ is infinitely differentiable, and satisfies the heat equation $\big[\partial_t+\mathcal{L}_x\big][u]=0$ in the classical sense. This establishes conclusion (1). Conclusion (2) is just the statement that $e^{-t\mathcal{L}}[f]\to f$ in $L^2(\Omega)$ as $t\to 0^+$. This completes the proof. □

4. The heat kernel and heat equation

4.1. The heat kernel.

In this section we prove the existence and study the basic properties of the heat kernel $H\in\mathcal{C}^\infty\big((0,\infty)\times\Omega\times\Omega\big)$ associated to the initial value problem solved in Section 3.

THEOREM 4.1. *There is a function $H:(0,\infty)\times\Omega\times\Omega\to\mathbb{C}$ with the following properties:*

1. *For fixed $(t,x)\in(0,\infty)\times\Omega$, the function $y\to H(t,x,y)$ belongs to $L^2(\Omega)$.*
2. *For each $f\in L^2(\Omega)$*

$$e^{-t\mathcal{L}}[f](x) = \int_\Omega f(y)\,H(t,x,y)\,dy. \tag{4.1}$$

3. *If $(t,x,y)\in(0,\infty)\times\Omega\times\Omega)$, then $H(t,x,y)=\overline{H(t,y,x)}$.*

(4) $H \in \mathcal{C}^\infty\big((0,\infty)\times\Omega\times\Omega\big)$*, and satisfies the heat equations*

$$\partial_t H(t,x,y) + \mathcal{L}_x H(t,x,y) = 0,$$
$$\partial_t H(t,x,y) + \mathcal{L}_y H(t,x,y) = 0,$$

for $(t,x,y) \in (0,\infty)\times\Omega\times\Omega$.

The proof of Theorem 4.1 will follow from a succession of lemmas. First, according to Theorem 3.8, the mapping $(t,x) \to e^{-t\mathcal{L}}[f](x)$ is infinitely differentiable. Hence for each integer $m \geq 0$ and each $x \in \Omega$, the mapping $f \to \mathcal{L}_x^m e^{-t\mathcal{L}}[f](x)$ is a linear functional on $L^2(\Omega)$. The key fact, which depends on the the subelliptic estimates for $\mathcal{L}$, is that this functional is bounded.

Lemma 4.2. *Let $m \geq 0$ and let $K \subset \Omega$ be a compact set. Choose an integer N so that $N\epsilon > n + 2m$. There is an allowable constant C so that for each $t > 0$, if $x \in K$ then for all $f \in L^2(\Omega)$,*

$$\big|\mathcal{L}_x^m e^{-t\mathcal{L}}[f](x)\big| \leq C\,(1+t^{-N})\,\big|\big|\,f\,\big|\big|_{L^2(\Omega)}.$$

Proof. Choose $\zeta \in \mathcal{C}_0^\infty(\Omega)$ with $\zeta(x) = 1$ for all $x \in K$. Then choose cut-off functions $\zeta \prec \zeta_1 \prec \cdots \prec \zeta_N = \zeta'$. By the Sobolev imbedding theorem, we have

$$\big|\mathcal{L}_x^m e^{-t\mathcal{L}}[f](x)\big| = \big|\mathcal{L}_x^m \zeta(x) e^{-t\mathcal{L}}[f](x)\big| \leq C\big|\big|\,\zeta e^{-t\mathcal{L}}[f]\,\big|\big|_{n+2m}.$$

Using the basic subelliptic estimate, we have

$$\big|\big|\,\zeta e^{-t\mathcal{L}}[f]\,\big|\big|_{n+2m} \leq C\big[\big|\big|\,\zeta_1 \mathcal{L} e^{-t\mathcal{L}}[f]\,\big|\big|_{n+2m-\epsilon} + \big|\big|\,\zeta_1 e^{-t\mathcal{L}}[f]\,\big|\big|_0\big].$$

If we repeat this argument N times we obtain

$$\big|\big|\,\zeta e^{-t\mathcal{L}}[f]\,\big|\big|_{n+2m} \leq C\sum_{j=0}^{N}\big|\big|\,\zeta'\mathcal{L}^j e^{-t\mathcal{L}}[f]\,\big|\big|_0 \leq C(1+t^{-N})\big|\big|\,f\,\big|\big|_{L^2(\Omega)},$$

which completes the proof. □

An application of the Riesz representation theorem gives the following result.

Corollary 4.3. *For $t > 0$, m a non-negative integer, and $x \in \Omega$, there exists a unique function $H_{t,x,m} \in L^2(\Omega)$ so that*

$$\mathcal{L}_x^m e^{-t\mathcal{L}}[f](x) = \int_\Omega f(y)\,H_{t,x,m}(y)\,dy.$$

Moreover, if $K \subset \Omega$ is compact and if C is the corresponding constant from Lemma 4.2, then if $x \in K$,

$$\int_\Omega |H_{t,x,m}(y)|^2\,dy \leq C^2\,(1+t^{-N})^2.$$

For each α we would like to regard $H_{t,x,m}(y)$ as a measurable function of three variables (t,x,y). We proceed as follows. Each element $H_{t,x,m}$ is by definition an equivalence class of measurable, square-integrable functions on Ω which differ only on sets of measure zero. For each t, x, m, choose one representative of this class, defined for all $y \in \Omega$, which we again call $H_{t,x,m}$. Then we can define a function $H_m(t,x,y)$ to be the value of $H_{t,x,m}$ at the point y.

Proposition 4.4. *The function H_m is measurable on $(0,\infty)\times\Omega\times\Omega$.*

Proof. Each function H_m satisfies the following properties:

(i) For each fixed $(t,x) \in (0,\infty)\times\Omega$, the function $y \to H_m(t,x,y)$ is measurable and in $L^2(\Omega)$.
(ii) For each fixed $f \in L^2(\Omega)$ the functions $(t,x) \to \int_\Omega f(y)\,H_m(t,x,y)\,dy$ is infinitely differentiable, and hence in particular is continuous.

Choose $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ supported in the unit ball with $\int_{\mathbb{R}^n} \varphi(u)\,du = 1$. For each positive integer k put $\varphi_k(u) = k^n \varphi(ku)$. Let $\Omega_N = \{u \in \Omega \,\big|\, B(u,N^{-1}) \subset \Omega\}$ be the set of points in Ω whose distance to the boundary is at least N^{-1}. For each $k > N$, consider the function defined on $(0,\infty)\times\Omega\times\Omega_N$ by

$$H_{m,k}(t,x,y) = \int_\Omega \varphi_k(u-y)\,H_m(t,x,u)\,du.$$

Then $H_{m,k}$ is continuous in (t,x,y), and hence is measurable. Also for each (t,x), $\lim_{k\to\infty} H_{m,k}(x,t,y) = H_m(x,t,y)$ for almost all y. The set of points at which $\lim_{k\to\infty} H_{m,k}(t,x,y)$ does not exist is measurable, and it follows that H_m is measurable. □

When $m = 0$, we shall write $H_0(t,x,y)$ as $H(t,x,y)$. This then establishes parts (1) and (2) of Theorem 4.1. Next, since the function $x \to e^{-tx}$ is real-valued, it follows from the spectral theorem that the operator $e^{-t\mathcal{L}}$ is self-adjoint. Thus if $\varphi, \psi \in L^2(\Omega)$ we have

$$(\mathcal{L}^m e^{-t\mathcal{L}}[\varphi],\psi) = \int_\Omega \mathcal{L}^m e^{-t\mathcal{L}}[\varphi](x)\,\overline{\psi(x)}\,dx = \iint_{\Omega\times\Omega} \varphi(y)\,\overline{\psi(x)}\,H_m(t,x,y)\,dx\,dy.$$

On the other hand,

$$\begin{aligned}(\mathcal{L}^m e^{-t\mathcal{L}}[\varphi],\psi) &= (e^{-t\mathcal{L}}[\varphi],\mathcal{L}^m\psi) && \text{(using integration by parts)},\\ &= \big(\varphi, e^{-t\mathcal{L}}\big[\mathcal{L}^m[\psi]\big]\big) && \text{(since } e^{-t\mathcal{L}} \text{ is self-adjoint)},\\ &= \big(\varphi, \mathcal{L}^m\, e^{-t\mathcal{L}}[\psi]\big) && \text{(since } e^{-t\mathcal{L}}\mathcal{L}^m[\psi] = \mathcal{L}^m e^{-t\mathcal{L}}[\psi]).\end{aligned}$$

But

$$(\varphi,\mathcal{L}^m e^{-t\mathcal{L}}[\psi]) = \int_\Omega \varphi(y)\,\overline{\mathcal{L}^m e^{-t\mathcal{L}}[\psi](y)}\,dy = \iint_{\Omega\times\Omega} \varphi(y)\,\overline{\psi(x)}\,\overline{H_m(t,y,x)}\,dx\,dy.$$

It follows that for each $t > 0$, we have $H_m(t,x,y) = \overline{H_m(t,y,x)}$ for almost all $(x,y) \in \Omega\times\Omega$. In particular, for each (t,y) the function $x \to H(t,x,y)$ belongs to $L^2(\Omega)$. Once we show that H is actually infinitely differentiable, this will establish part (3).

For every integer $m \geq 0$ and every $\varphi \in \mathcal{C}_0^\infty(\Omega)$, we have

$$\int_\Omega \varphi(y)\,H_m(t,x,y)\,dy = \mathcal{L}_x^m e^{-t\mathcal{L}}[\varphi](x) = \mathcal{L}_x^m\Big[\int_\Omega \varphi(y)\,H(t,x,y)\,dy\Big].$$

If we integrate this against $\psi \in \mathcal{C}_0^\infty(\Omega)$ we get

$$\iint_{\Omega\times\Omega} \varphi(y)\,\psi(x)\,H_m(t,x,y)\,dy\,dx = \iint_{\Omega\times\Omega} \varphi(y)\,\mathcal{L}_x^m\psi(x)\,H(t,x,y)\,dy\,dx,$$

and consequently

$$\int_\Omega \mathcal{L}^m\psi(x)\,H(t,x,y)\,dx = \int_\Omega \psi(x)\,H_m(t,x,y)\,dx = \int_\Omega \psi(x)\,\overline{H_{t,y,m}(x)}\,dx.$$

It follows that for each fixed (t,y), $\mathcal{L}_x^m H(t,x,y) = \overline{H_{t,y,m}(x)}$ in the sense of distributions, and since $H_m(t,x,y) = \overline{H(t,y,x)}$, we also have $\mathcal{L}_y^m H(t,x,y) = H_m(t,x,y)$ in the sense of distributions.

But since $\overline{H}_{t,y,m} \in L^2(\Omega)$, we can now use the basic subelliptic estimate again. For any $s > 0$, choose N so that $N\epsilon > s$. Choose $\zeta \prec \zeta_1 \prec \cdots \prec \zeta_N = \zeta' \in \mathcal{C}_0^\infty(\Omega)$. Then for fixed (t,y) we have

$$\begin{aligned}
\big|\big|\, \zeta\, H(t,\cdot,y)\,\big|\big|_s &\leq C\big[\big|\big|\,\zeta_1\,\mathcal{L}H(t,\cdot,y)\,\big|\big|_{s-\epsilon} + \big|\big|\,H(t,\cdot,y)\,\big|\big|_0 \\
&\leq C\big[\big|\big|\,\zeta_2\,\mathcal{L}^2 H(t,\cdot,y)\,\big|\big|_{s-2\epsilon} + \big|\big|\,\mathcal{L}H(t,\cdot,y)\,\big|\big|_0 + \big|\big|\,H(t,\cdot,y)\,\big|\big|_0\big] \\
&\leq \;\cdots \\
&\leq C \sum_{m=0}^{N} \big|\big|\,\zeta'\,\mathcal{L}^m H(t,\cdot,y)\,\big|\big|_0 < \infty.
\end{aligned}$$

It follows that the function $x \to H(t,x,y)$ is infinitely differentiable, and a similar argument shows that $y \to H(t,x,y)$ is infinitely differentiable.

It follows from Corollary 4.3 that the functions $(x,y) \to H_m(t,x,y)$ are locally square integrable on $\Omega \times \Omega$. Thus since all pure x derivatives and all pure y derivatives of $H(t,x,y)$ are square integrable, it follows from classical elliptic theory that the function $(x,y) \to H(t,x,y)$ is infinitely differentiable.

Finally, it follows from Theorem 3.8 that

$$[\partial_t + \mathcal{L}_x]\left[\int_\Omega H(t,x,y)\,\varphi(y)\,dy\right] = 0$$

for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$. Integrating against a test function in $\mathcal{C}_0^\infty\big((0,\infty)\times\Omega\big)$, it follows that, in the sense of distributions, $[\partial_t + \mathcal{L}_x]H(t,x,y) = 0$ for each $y \in \Omega$. But since $[\partial_t + \mathcal{L}_x]$ is hypoelliptic, it follows that $(t,x) \to H(t,x,y)$ is infinitely differentiable, and $\partial_t^m H(t,x,y) = (-1)^m \mathcal{L}_x^m H(t,x,y) = (-1)^m \mathcal{L}_y^m H(t,x,y)$. It follows that H is infinitely differentiable on $(0,\infty)\times\Omega\times\Omega$, and satisfies the heat equations specified in part (4) of Theorem 4.1. This completes the proof of the theorem.

4.2. The heat equation on $\mathbb{R} \times \Omega$.

In this section we study the operator $\partial_t + \mathcal{L}_x$ on the whole space $\mathbb{R}\times\Omega$. Extend the domain of the function H by setting $H(t,x,y) = 0$ for $t \leq 0$, and then define

$$\widetilde{H}\big((t,x),(s,y)\big) = H(s-t,x,y). \tag{4.2}$$

for $(t,x),(s,y) \in \mathbb{R}\times\Omega$. If $\psi \in \mathcal{C}_0^\infty(\mathbb{R}\times\Omega)$, set

$$\begin{aligned}
\Big\langle K_{(t,x)}, \psi \Big\rangle &= \lim_{\eta\to 0^+} \int_\eta^{+\infty}\int_\Omega K\big((s,x),(0,y)\big)\,\psi(s,y)\,dy\,ds \\
&= \lim_{\eta\to 0^+} \int_\eta^{+\infty}\int_\Omega H(s,x,y)\,\psi(s+t,y)\,dy\,ds.
\end{aligned} \tag{4.3}$$

Lemma 4.5. *The limit in equation (4.3) exists, and $K_{(t,x)}$ is a distribution on $\mathbb{R}\times\Omega$. Moreover, $[\partial_s + \mathcal{L}_y][K_{(t,x)}] = \delta_t \otimes \delta_x$ in the sense of distributions.*

Proof. Set $\psi_{t,s}(y) = \psi(s+t,y)$. Then $\psi_{t,s} \in \mathcal{C}_0^\infty(\Omega)$. Choose a positive integer N so that $N\epsilon > \frac{n}{2}$. Choose $\zeta \prec \zeta_1 \prec \cdots \prec \zeta_N = \zeta' \in \mathcal{C}_0^\infty(\Omega)$ with $\zeta(x) = 1$. Then by the Sobolev imbedding theorem and the basic subelliptic estimate applied N times we have

$$\begin{aligned}\left|\int_\Omega H(s,x,y)\,\psi(s+t,y)\,dy\right| &= \left|\zeta(x)\,e^{-s\mathcal{L}}[\psi_{t,s}](x)\right| \\ &\le C\,\big|\big|\,\zeta\,e^{-s\mathcal{L}}[\psi_{s,t}]\,\big|\big|_{N\epsilon} \\ &\le C\,\big[\big|\big|\,\zeta_1\mathcal{L}\big[e^{-s\mathcal{L}}[\psi_{s,t}]\big]\,\big|\big|_{(N-1)\epsilon}+\big|\big|\,\zeta_1\,e^{-s\mathcal{L}}[\psi_{s,t}]\,\big|\big|_0\big] \\ &\le \cdots \\ &\le C\sum_{j=0}^{N}\big|\big|\,\zeta'\,\mathcal{L}^j\big[e^{-s\mathcal{L}}[\psi_{s,t}]\big]\,\big|\big|_0\end{aligned}$$

But since $\mathcal{C}_0^\infty(\Omega)\subset D_{\mathcal{L}}$, it follows from Theorem 3.7, part (4), that $\mathcal{L}^j\big[e^{-s\mathcal{L}}[\psi_{s,t}]\big]=e^{-s\mathcal{L}}\big[\mathcal{L}^j[\psi_{s,t}]\big]$, and hence

$$\left|\int_\Omega H(s,x,y)\,\psi(s+t,y)\,dy\right|\le C\sum_{j=0}^{N}\big|\big|\,e^{-s\mathcal{L}}\big[\mathcal{L}^j[\psi_{s,t}]\big]\,\big|\big|_0\le C\sum_{j=0}^{N}\big|\big|\,\mathcal{L}^j[\psi_{s,t}]\,\big|\big|_0.$$

This last quantity is uniformly bounded in s, and hence

$$\left|\int_{\eta_1}^{\eta_2}\int_\Omega H(s,x,y)\,\psi(s+t,y)\,dy\,ds\right|\le C\,|\eta_2-\eta_1|\,\sup_{s,t}\sum_{j=0}^{N}\big|\big|\,\mathcal{L}^j[\psi_{s,t}]\,\big|\big|_0.$$

This shows that the limit in equation (4.3) exists, and that $K_{(t,x)}$ is a distribution on $\mathbb{R}\times\Omega$.

Again let $\psi\in\mathcal{C}_0^\infty(\mathbb{R}\times\Omega)$. We have

$$\begin{aligned}\Big\langle K_{(t,x)},\partial_y\psi\Big\rangle &= \lim_{\eta\to0^+}\int_\Omega\int_\eta^\infty H(s,x,y)\,(\partial_s\psi)(s+t,y)\,ds\,dx \\ &= \lim_{\eta\to0^+}\int_\Omega\left[H(s,x,y)\,\psi_{t,s}(y)\Big|_{s=\eta}^{\infty}-\int_\eta^\infty\partial_sH(s,x,y)\,\psi_{t,s}(y)\,ds\right]\,dy \\ &= -\lim_{\eta\to0^+}\left[\int_\Omega H(\eta,x,y)\,\psi_{t,\eta}(y)\,dy+\int_\eta^\infty\int_\Omega\mathcal{L}_yH(s,x,y)\,\psi_{t,s}(y)\,dy\,ds\right] \\ &= -\lim_{\eta\to0^+}\int_\Omega H(\eta,x,y)\,\psi_{t,\eta}(y)\,dy+\Big\langle K_{(t,x)},\mathcal{L}_y\psi\Big\rangle\end{aligned}$$

It follows that

$$\Big\langle K_{(t,x)},[-\partial_y+\mathcal{L}_y]\psi\Big\rangle=\lim_{\eta\to0^+}e^{-\eta\mathcal{L}}[\psi_{t,\eta}](x)=\psi(t,x).$$

This completes the proof. □

COROLLARY 4.6. *The distribution $K_{(t,x)}$ is given by a function which is infinitely differentiable on $\mathbb{R}\times\Omega-\{(t,x)\}$. In particular, for every multi-index α, if $x\neq y$ it follows that*

$$\lim_{t\to0^+}\partial_y^\alpha H(t,x,y)=0. \tag{4.4}$$

5. Pointwise estimates for H

5.1. Scaling maps.

We begin by recalling the definition and properties of scaling maps. Let $\{X_1, \ldots, X_p\}$ be smooth real vector fields defined in a neighborhood of the closure of a smoothly bounded open set $\Omega \subset \mathbb{R}^n$. Let $\rho : \Omega \times \Omega \to [0, \infty)$ be the control metric generated by these vector fields, and let $\{B(x, \delta)\}$ denote the corresponding family of control-metric balls. Let $\mathbb{B}(r)$ denote the Euclidean ball of radius r centered at the origin.

LEMMA 5.1. *There exists $\delta_0 > 0$, and for every $x_0 \in \Omega$ and every $0 < \delta < \delta_0$ there exists a diffeomorphism $\Phi = \Phi_{x_0,\delta} : \mathbb{B}(1) \to \mathbb{R}^n$ with the following properties:*

(1) $B(x_0, \frac{1}{2}\delta) \subset \Phi\big(\mathbb{B}(\frac{1}{2})\big) \subset B(x_0, \delta) \subset \big(\mathbb{B}(1)\big)$.

(2) Let $J\Phi$ denote the absolute value of the Jacobian determinant of the mapping Φ. There is a constant C so that for all $u \in \mathbb{B}(1)$ we have

$$C^{-1}\big|B(x, \delta)\big| \leq J\Phi(u) \leq C\,\big|B(x, \delta)\big|.$$

Let $\{Z_1, \ldots, Z_p\}$ be the vector fields on $\mathbb{B}(1)$ so that $d\Phi[Z_j] = \delta X_j$. If

$$\mathcal{L} = -\sum_{j=1}^{p} X_j^2 + \sum_{j=1}^{p} \alpha_j\, X_j,$$

and if

$$\mathcal{Z} = -\sum_{j=1} Z_j^p + \delta \sum_{j=1}^{p} (\alpha_j \circ \Phi^{-1})\, Z_j$$

then

$$d\Phi(\mathcal{Z}) = \delta^2\, \mathcal{L}.$$

Moreover, $\mathcal{Z}$ satisfies subelliptic estimates which are uniform in x and δ. If $\zeta \prec \zeta' \in \mathcal{C}_0^\infty\big(\mathbb{B}(1)\big)$, then if u is a distribution on $\mathbb{B}(1)$, we have

$$\big|\big|\,\zeta\, u\,\big|\big|_{s+\epsilon} \leq C\big[\big|\big|\,\zeta'\,\mathcal{Z}[u]\,\big|\big|_s + \big|\big|\,\zeta'\,u\,\big|\big|_0\big]$$

where the constant C can depend on the choice of ζ, ζ', but is independent of x and δ. The mappings $\{\Phi_{x_0,\delta}\}$ are called the scaling maps.

5.2. A bound for the scaled initial value problem.

Now consider the heat operator

$$\mathcal{H} = \partial_t + \mathcal{L} = \partial_t + \sum_{j=1}^{p} X_j^* X_j = \partial_t - \sum_{j=1}^{p} X_j^2 + \sum_{j=1}^{p} \alpha_j\, X_j.$$

For each $(x_0) \in \mathbb{R} \times \Omega$ and $0 < \delta < \delta_0$, define a mapping $\Psi = \Psi_{x_0,\delta}\big(\mathbb{R} \times \mathbb{B}(1)\big) \to \mathbb{R} \times \mathbb{R}^n$ by setting

$$\Psi_{x_0,\delta}(s, u) = \big(\delta^2 s, \Phi_{x_0,\delta}(u)\big).$$

Then

$$d\Psi_{x_0,\delta}[\partial_s + \mathcal{Z}_u] = \delta^2\big[\partial_t + \mathcal{L}_x\big].$$

Essentially, we can think of the operator $\partial_s + \mathcal{Z}_u$ as the normalized heat operator on $\mathbb{R} \times \mathbb{B}(1)$ which corresponds to the heat operator $\partial_t + \mathcal{L}_x$ on $\mathbb{R} \times B(x_0, \delta)$ under the change of variables given by the mapping $\Psi_{t_0,x_0,\delta}$. In particular, the operators $\pm\partial_s + \mathcal{Z}_u$ satisfy subelliptic estimates on $\mathbb{R}\times\mathbb{B}(1)$ which are uniform in the parameters (t_0, x_0, δ).

We can now pull back the heat kernel H using this same change of variables. If $s > 0$ and $u, v \in \mathbb{B}(1)$, put

$$W(s,u,v) = W_{x_0,\delta}(s,u,v) = H\big(\delta^2 s, \Phi_{x_0,\delta}(u), \Phi_{x_0,\delta}(v)\big).$$

It follows from the definition of $\mathcal{Z}$ and the chain rule that

$$\begin{aligned}\big[\partial_s + \mathcal{Z}_u\big][W](s,u,v) &= 0,\\ \big[\partial_s + \mathcal{Z}_v\big][W](s,u,v) &= 0\end{aligned}$$

For $s > 0$ and $f \in \mathcal{C}_0^\infty\big(\mathbb{B}(1)\big)$ define

$$W_s[f](u) = \int_{\mathbb{B}(1)} W(s,u,v)\, f(v)\, dv = \int_{\mathbb{B}(1)} H\big(\delta^2 s, \Phi(u), \Phi(v)\big)\, f(v)\, dv.$$

The key point is that we can bound the norm of the operator W_s on $L^2\big(\mathbb{B}(1)\big)$.

LEMMA 5.2. *There is a constant C which is independent of x_0, δ, and $s > 0$ so that*

$$\big|\big|\, W_s[f]\,\big|\big|_{L^2(\mathbb{B}(1))} \leq C\,\big|B(x_0,\delta)\big|^{-1}\,\big|\big|\, f\,\big|\big|_{L^2(\mathbb{B}(1))}.$$

PROOF. Let $\Theta = \Phi^{-1}$. Then if $x \in B(x_0,\delta)$, we have

$$\begin{aligned}W_s[f]\big(\Theta(x)\big) &= \int_{\mathbb{B}(1)} H(\delta^2 s, \Phi\big(\Theta(x)\big), \Phi(v)\big)\, f(v)\, dv\\ &= \int_{\Omega} H(\delta^2 s,\, x,\, y)\, (f\circ\Theta)(y)\, J\Theta(y)\, dy\\ &= e^{-\delta^2 s\mathcal{L}}\,\big[(f\circ\Theta)J\Theta\big](x)\end{aligned}$$

Since the operator $e^{-\delta^2 s\mathcal{L}}$ is a contraction on $L^2(\Omega)$, it follows that

$$\begin{aligned}\int_\Omega |W_s[f]\big(\Theta(x)\big)|^2\, dx &\leq \big|\big|\,(f\circ\Theta)J\Theta\,\big|\big|^2_{L^2(\Omega)}\\ &= \int_\Omega \big|f\big(\Theta(x)\big)\big|^2\, J\theta(x)^2\, dx\\ &= \int_{\mathbb{B}(1)} \big|f(u)\big|^2\, J\Theta\big(\Phi(u)\big)^2\, J\Phi(u)\, du\\ &\leq C\,|B(x_0,\delta)|^{-1} \int_{\mathbb{B}(1)} |f(u)|^2\, du,\end{aligned}$$

since $J\Theta\big(\Phi(u)\big) = J\Phi(u)^{-1}$, and $J\Phi(u) \geq C^{-1}\,|B(x_0,\delta)|$ according to Lemma 5.1. On the other hand, we also have

$$\begin{aligned}\int_\Omega \big|W_s[f]\big(\Theta(x)\big)\big|^2\, dx &= \int_{\mathbb{B}(1)} \big|W_s[f](u)\big|^2\, J\Phi(u)\, du\\ &\geq C^{-1}|B(x_0,\delta)| \int_{\mathbb{B}(1)} \big|W_s[f](u)\big|^2\, du.\end{aligned}$$

Putting the two inequalities together, we obtain

$$\int_{\mathbb{B}(1)} \big|W_s[f](u)\big|^2\, du \leq C^2\,\big|B(x_0,\delta)\big|^{-2} \int_{\mathbb{B}(1)} \big|f(u)\big|^2\, du,$$

which completes the proof. □

5.3. Pointwise estimates for H.

Let $\{e^{-t\mathcal{L}}\}_{t\geq 0}$ be the heat semigroup for the opertor $\mathcal{L} = \sum_{j=1}^{p} X_j^* X_j$, and let $H : (0,\infty) \times \Omega \times \Omega \to \mathbb{R}$ be the associated heat kernel so that

$$e^{-t\mathcal{L}}[f](x) = \int_\Omega H(t,x,y)\, f(y)\, dy$$

for every $f \in L^2(\Omega)$. The object of this section is to obtain local estimates for the function H and certain of its derivatives in terms of t and the control metric ρ. An expression X_x^α denotes the differential operator given by a product $X_{\alpha_1} \cdots X_{\alpha_k}$ in the variable x, and we write $|\alpha| = k$. The expression X_y^β is defined similarly.

THEOREM 5.3. *Let j, k, l be non-negative integers. For every positive integer N there is a constant $C_N = C_{N,j,k,l}$ so that if $|\alpha| = k$ and $|\beta| = l$*

$$\big|\partial_t^j\, X_x^\alpha\, X_y^\beta H(t,x,y)\big| \leq \begin{cases} C_N\, \rho(x,y)^{-2j-k-l}\, \big|B(x,\rho(x,y))\big|^{-1} \, \left(\frac{t}{\rho(x,y)^2}\right)^N & \text{if } t \leq \rho(x,y)^2, \\ C_0\; t^{-j-k/2-l/2}\, \big|B(x,\sqrt{t})\big|^{-1} & \text{if } t \geq \rho(x,y)^2, \end{cases}$$

for all (t,x,y) with $|t| + \rho(x,y) \leq 1$.

PROOF. We begin with the case $N = 0$. It suffices to establish the estimates for $|t| + \rho(x,y) \leq \delta_0$, since estimates for $\delta_0 \leq |t| + \rho(x,y) \leq 1$ then follow by compactness. Let (t_0,x_0), $(t,x) \in \mathbb{R} \times \Omega$ with $|t-t_0|^{\frac{1}{2}} + \rho(x,x_0) = \delta \leq \delta_0$. There is a unique point $(s_0,v_0) \in (-1,+1) \times \mathbb{B}(1)$ so that $(t,x) = \big(t_0 + \delta^2 s_0, \Phi_{x_0,\delta}(v_0)\big)$. There is an absolute constant $\tau > 0$ so that $|s_0|^{\frac{1}{2}} + |v_0| \geq \tau$.

For (r,u), $(s,v) \in \big(-1,+1\big) \times \mathbb{B}(1)$, put

$$W^{\#}\big((r,u),(s,v)\big) = H(\delta^2(s-r), \Phi_{x_0,\delta}(u), \Phi_{x_0,\delta}(v)\big).$$

Note that

$$\begin{aligned} \big[-\partial_r + \mathcal{Z}_u\big][W^{\#}] &= 0, \\ \big[+\partial_s + \mathcal{Z}_v\big][W^{\#}] &= 0, \end{aligned}$$

and

$$\begin{aligned} \big[\partial_t^j\, X_x^\alpha\, X_y^\beta H\big](\delta^2(s-r), &\Phi_{x_0,\delta}(u), \Phi_{x_0,\delta}(v)\big) \\ &= \delta^{-2j-k-l}\, \big[\partial_s^j Z_u^\alpha\, Z_v^\beta\, W^{\#}\big]\big((r,u),(s,v)\big). \end{aligned}$$

For $f \in \mathcal{C}_0^\infty\big((-1,+1) \times \mathbb{B}(1)\big)$, put

$$T^{\#}[f](r,u) = \iint_{\mathbb{R}\times\mathbb{B}(1)} W^{\#}\big((r,u),(s,v)\big)\, f(s,v)\, dv\, ds.$$

Put

$$\begin{aligned} B_1 &= \left\{(r,u) \,\middle|\, |r|^{\frac{1}{2}} + |u| < \frac{1}{3}\tau\right\}, \\ B_2 &= \left\{(s,v) \,\middle|\, |s-s_0|^{\frac{1}{2}} + |v-v_0| < \frac{1}{3}\tau\right\}. \end{aligned}$$

Then $B_1 \cap B_2 = \emptyset$. Choose functions $\zeta \prec \zeta' \prec \zeta'' \in \mathcal{C}_0^\infty(B_1)$ with $\zeta(0,0) = 1$, and $\eta \prec \eta' \in \mathcal{C}_0^\infty(B_2)$ with $\eta(s_0, v_0) = 1$. Then using the Sobolev inequality and the subelliptic estimate for the operator $[\partial_s + \mathcal{Z}_v]$, we have

$$\begin{aligned}
|\big[\partial_s^j Z_u^\alpha\, Z_v^\beta\, W^\#\big]\big((0,0),(s_0,v_0)\big)| &= |\zeta(s_0,v_0)\,\big[\partial_s^j Z_u^\alpha\, Z_v^\beta\, W^\#\big]\big((0,0),(s_0,v_0)\big)| \\
&\leq C\big|\big|\,\zeta'\,W^\#\big((r_0,u_0),(\cdot,\cdot)\big)\,\big|\big|_{n+1+j+k+l} \\
&\leq C\Big[\big|\big|\,\zeta''\,[\partial_s + \mathcal{Z}_v]W^\#\big((0,0),(\cdot,\cdot)\big)\,\big|\big|_{n+1+j+k+l-\epsilon} \\
&\qquad + \big|\big|\,\zeta''\,W^\#\big((0,0),(\cdot,\cdot)\big)\,\big|\big|_0\Big] \\
&= C\,\big|\big|\,\zeta''\,W^\#\big((0,0),(\cdot,\cdot)\big)\,\big|\big|_0
\end{aligned}$$

where the last equality follows since $[\partial_s + \mathcal{Z}_v]W^\#\big((0,0),(s,v)\big) = 0$ on B_2 which contains the support of ζ'. We estimate this last norm by duality. We have

$$\begin{aligned}
&\big|\big|\,\zeta'\,W^\#\big((0,0),(\cdot,\cdot)\big)\,\big|\big|_0 \\
&\qquad = \sup_\varphi\,\Big|\iint_{\mathbb{R}\times\mathbb{B}(1)} \zeta'(s,v)\,W^\#\big((0,0),(s,v)\big)\,\varphi(s,v)\,dv\,ds\Big| \\
&\qquad = \sup_\varphi\,\big|T^\#[\zeta'\varphi](0,0)\big|,
\end{aligned}$$

where the supremum is taken over all $\varphi \in \mathcal{C}_0^\infty(B_2)$ with $\big|\big|\,\varphi\,\big|\big| \leq 1$. Now use the Sobolev inequality and a subelliptic estimate again. We have

$$\begin{aligned}
\sup_\varphi\,\big|T^\#[\zeta'\varphi](0,0)\big| &= \sup_\varphi\,\big|\eta(0,0)T^\#[\zeta'\varphi](r_0,u_0)\big| \\
&\leq C\,\sup_\varphi\,\big|\big|\,\eta\,T^\#[\zeta'\varphi]\,\big|\big|_{n+1} \\
&\leq C\,\sup_\varphi\,\Big[\big|\big|\,\eta'[-\partial_r + \mathcal{Z}_u]T^\#[\zeta'\varphi]\,\big|\big|_{n+1-\epsilon} + \big|\big|\,\eta'\,T^\#[\zeta'\varphi]\,\big|\big|_0\Big] \\
&= C\,\sup_\varphi\,\big|\big|\,\eta'\,T^\#[\zeta'\varphi]\,\big|\big|_0 \\
&\leq C\,\sup_\varphi\,\big|\big|\,T^\#[\zeta'\varphi]\,\big|\big|_0 \\
&\leq C\,\big|\big|\,T^\#\,\big|\big|,
\end{aligned}$$

where the equality in the third to last line follows since $[\partial_r + \mathcal{Z}_u]T^\#[\zeta'\varphi](r,u) = 0$ on B_1 which includes the support of η'. Thus we have shown that

$$\big|\partial_t^j X_x^\alpha X_y^\beta H(\delta^2 s_0, x_0, x)\big| \leq C\,\delta^{-2j-k-l}\big|\big|\,T^\#\,\big|\big|,$$

the norm of the operator $T^\#$ on $L^2\big((-1,+1)\times\mathbb{B}(1)\big)$.

Let $f,\, g \in \mathcal{C}_0^\infty\big((-1,+1)\times\mathbb{B}(1)\big)$, and let $f_s(v) = f(s,v),\ g_r(u) = g(r,u)$. We then have

$$\begin{aligned}
&\left|\iint_{\mathbb{R}\times\mathbb{B}(1)} T^{\#}[f](u,r)\, g(u,r)\, du\, dr\right| \\
&\qquad= \left|\iint_{\mathbb{R}\times\mathbb{B}(1)}\iint_{\mathbb{R}\times\mathbb{B}(1)} W^{\#}\big((r,u),(s,v)\big)\, f(s,v)\, g(u,r)\, dv\, ds\, du\, dr\right| \\
&\qquad= \left|\iiiint H\big(\delta^2(s-r),\Phi(u),\Phi(v)\big)\, f(s,v)\, g(r,u)\, ds\, dr\, du\, dv\right| \\
&\qquad= \left|\iiiint H\big(\delta^2 s,\Phi(u),\Phi(v)\big)\, f(s+r,v)\, g(r,u)\, dv\, ds\, dr\, du\right| \\
&\qquad= \left|\iiint W_s[f_{s+r}](u)\, g(r,u)\, du\, ds\, dr\right| \\
&\qquad\le C\iint_{\mathbb{R}\times\mathbb{R}}\int_{\mathbb{B}(1)} \big|W_s[f_{s+r}](u)\, g_r(u)\big|\, du\, ds\, dr \\
&\qquad\le C\iint_{\mathbb{R}\times\mathbb{R}} \big|\big|\, W_s[f_{s+r}]\,\big|\big|_{L^2(\mathbb{B}(1))}\,\big|\big|\, g_r\,\big|\big|_{L^2(\mathbb{B}(1))}\, ds\, dr \\
&\qquad\le C\,|B(x_0,\delta)|^{-1}\iint_{\mathbb{R}\times\mathbb{R}} \big|\big|\, f_{s+r}\,\big|\big|_{L^2(\mathbb{B}(1))}\,\big|\big|\, g_r\,\big|\big|_{L^2(\mathbb{B}(1))}\, ds\, dr \\
&\qquad= C\,|B(x_0,\delta)|^{-1}\int_{\mathbb{R}} \big|\big|\, f_s\,\big|\big|_{L^2(\mathbb{B}(1))}\, ds\int_{\mathbb{R}} \big|\big|\, g_r\,\big|\big|_{L^2(\mathbb{B}(1))}\, dr \\
&\qquad\le C\,|B(x_0,\delta)|^{-1}\,\big|\big|\, f\,\big|\big|_{L^2(\mathbb{R}\times\mathbb{B}(1))}\,\big|\big|\, g\,\big|\big|_{L^2(\mathbb{R}\times\mathbb{B}(1))}.
\end{aligned}$$

In the last inequality, we have used the Schwarz inequality and the fact that $f_s = g_r \equiv 0$ unless $r, s, \in (-1. + 1)$. It follows that $\big|\big|\, T^{\#}\,\big|\big| \le C\,|B(x_0,\delta)|^{-1}$, and this completes the proof in case $N = 0$.

To deal with the case when $N > 0$, we use the fact that when $x \neq y$, the infinitely differentiable function $t \to H(t,x,y)$ and all its derivatives vanish when $t = 0$. Thus using Taylor's formula, we have, for example,

$$\begin{aligned}
|H(t,x,y)| &\le \frac{1}{(N-1)!}\int_0^t |\partial_s^N H(s,x,y)|\,(t-s)^{N-1}\, ds \\
&\le C_0\frac{1}{(N-1)!}\rho(x,y)^{-2N}|B(x,\rho(x,y))|^{-1}\int_0^t (t-s)^{N-1}\, ds \\
&= C_0\frac{1}{N!}\left(\frac{t}{\rho(x,y)^2}\right)^N\, |B(x,\rho(x,y))|^{-1}.
\end{aligned}$$

Estimates for other derivatives of $H(t,x,y)$ are handled in the same way. □

5.4. Action of $e^{-t\mathcal{L}}$ on bump functions.

THEOREM 5.4. *For each multi-index α there is an integer N_α and a constant C_α so that if $\varphi \in \mathcal{C}_0^\infty\big(B(x,\delta)\big)$, then*

$$\big|X_x^\alpha e^{-t\mathcal{L}}[\varphi](x)\big| \le C_\alpha \sup_{y\in\Omega}\sum_{|\beta|\le N_\alpha} \delta^\beta\big|X^\beta\varphi(y)\big|.$$

CHAPTER 6

Non-isotropic smoothing operators

Let $\Omega \subset \mathbb{R}^n$ be a connected open set, and let $\{X_1, \ldots, X_p\}$ be smooth real vector fields on Ω which are of finite type m. In this chapter we introduce and study a class of operators on Ω which stand in the same relationship to a control system of vector fields $\{X_1, \ldots, X_p\}$ as the more classical Calderón-Zygmund operators do to the standard coordinate vector fields $\{\partial_{x_1}, \ldots, \partial_{x_n}\}$.

1. Definitions and properties of NIS operators

We begin by recalling some standard notation. Let $\rho : \Omega \times \Omega \to [0, \infty)$ denote the control metric associated to the family of vector fields $\{X_1, \ldots, X_p\}$. Let $\mathbb{I}_k$ denote the set of ordered k-tuples of integers $(i_1, \ldots, i_k)$ where $1 \leq i_j \leq p$ for $1 \leq j \leq k$. If $I = (i_1, \ldots, i_k) \in \mathbb{I}_k$, then $X^I = X_{i_1} \cdots X_{i_k}$. The diagonal of $\Omega \times \Omega$ by $\triangle_\Omega$.

1.1. Definition.

Definition 1.1. *An operator $\mathcal{T} : \mathcal{C}_0^\infty(\Omega) \to \mathcal{D}'(\Omega)$ is a non-isotropic smoothing operator (NIS operator) of order m if the following conditions hold:*

(1) *There is a function $T_0 \in \mathcal{C}^\infty(\Omega \times \Omega \backslash \triangle_\Omega)$ so that if $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$ have disjoint supports,*

$$\Big\langle \mathcal{T}[\varphi], \psi \Big\rangle = \iint_{\Omega \times \Omega} \varphi(y)\, \psi(x)\, T_0(x, y)\, dx\, dy. \tag{1.1}$$

(2) *There exist functions $T_\epsilon \in \mathcal{C}^\infty(\Omega \times \Omega)$ for $\epsilon > 0$ so that if we set*

$$\mathcal{T}_\epsilon[\varphi](x) = \int_\Omega \varphi(y)\, T_\epsilon(x, y)\, dy \tag{1.2}$$

for $\varphi \in \mathcal{C}_0^\infty(\Omega)$, then for any $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$,

$$\big\langle \mathcal{T}[\varphi], \psi \big\rangle = \lim_{\epsilon \to 0} \iint_{\Omega \times \Omega} \varphi(y)\, \psi(x)\, T_\epsilon(x, y)\, dx\, dy. \tag{1.3}$$

In other words, $\lim_{\epsilon \to 0^+} \mathcal{T}_\epsilon[\varphi] = \mathcal{T}[\varphi]$ in the sense of distributions.

(3) *For each pair of non-negative integers k, l there is a constant $C_{k,l}$ so that if $K \in \mathbb{I}_k$, $L \in \mathbb{I}_l$, $(x, y) \neq 0$ and $\epsilon \geq 0$,*

$$\big| X_x^K X_y^L T_\epsilon(x, y) \big| \leq C_{k,l}\, \rho(x, y)^{m-k-l}\, V(x, y)^{-1}. \tag{1.4}$$

(4) *For each positive integer k there is a positive integer N_k and a constant C_k so that if $\varphi \in \mathcal{C}_0^\infty(B(x_0, \delta))$ and $K \in \mathbb{I}_k$,*

$$\sup_{x \in \Omega} \big| X^K T_\epsilon[\varphi](x) \big| \leq C_k\, \delta^{m-k} \sup_{y \in \Omega} \sum_{|J| \leq N_k} \delta^{|J|}\, \big| X^J[\varphi](y) \big|. \tag{1.5}$$

(5) *Conditions (1) through (4) also hold for the adjoint operator $\mathcal{T}^*$ which is defined by the requirement that*

$$\left\langle \mathcal{T}^*[\varphi], \psi \right\rangle = \left\langle \mathcal{T}[\psi], \varphi \right\rangle \tag{1.6}$$

for all $\varphi, \psi \in \mathcal{C}_0^\infty(\Omega)$.

We make some preliminary remarks about the various hypotheses in this definition.

(i) Condition (1) says that the distribution kernel for the operator $\mathcal{T}$ is given by integration against the function T_0 which is smooth away from the diagonal. Thus an NIS operator $\mathcal{T}$ is pseudo-local; the distribution $\mathcal{T}[u]$ can only be singular where u is singular. In particular, if $\varphi \in \mathcal{C}_0^\infty(\Omega)$, then away from the support of φ, $\mathcal{T}[\varphi]$ is a distribution given by the infinitely differentiable function $\mathcal{T}[\varphi](x) = \int_\Omega T_0(x, y)\, \varphi(y)\, dy$.

(ii) Condition (2) guarantees the existence of a regularization of the distribution kernel T_0 associated to an NIS operator $\mathcal{T}$. The important point is that the size estimates in conditions (3) and (4) are uniform in the parameter ϵ.

(iii) Condiiton (4) encodes the basic cancellation hypothesis needed to prove that NIS operators of order zero are bounded on $L^2(\Omega)$. Note that if $m \leq 0$ the function $y \to \rho(x, y)^m\, V(x, y)^{-1}$ is not locally integrable, and if $\mathcal{T}$ is an NIS operator of order $m \leq 0$, the integral $\int_\Omega T_0(x, y)\, \varphi(y)\, dy$ need not converge absolutely, even if $\varphi \in \mathcal{C}_0^\infty(\Omega)$. Condition (4) provides the required estimates for $\mathcal{T}[\varphi]$.

(iv) We shall often say that a function $\varphi \in \mathcal{C}_0^\infty(\Omega)$ is a *bump function* (relative to $B(x, \delta)$) if the support of φ is contained in $B(x, \delta)$. We say that φ is a *normalized bump function* if $\sup_{y \in \Omega} \left| \sum_{|J| \leq N} \delta^{|J|} X^J[\varphi](y) \right| \leq 1$. With this terminology, condition (4) says that if φ is a normalized bump function, then $\sup_{x \in \Omega} \left| X^K \mathcal{T}_\epsilon[\varphi](x) \right| \leq C_k\, \delta^{m-k}$.

1.2. Elementary properties.

Let $\mathcal{L} = \sum_{j=1}^p X_j^* X_j$, and let $e^{-t\mathcal{L}}$ be the semi-group of operators studied in Chapter 5. Let H be the corresponding heat kernel, so that

$$e^{-t\mathcal{L}}[\varphi](x) = \int_\Omega H(t, x, y)\, \varphi(y)\, dy. \tag{1.7}$$

PROPOSITION 1.2. *The identity operator is an NIS operator smoothing of order zero.*

PROOF. If we take $T_\epsilon(x, y) = H(\epsilon, x, y)$, then $T_\epsilon \in \mathcal{C}^\infty(\Omega \times \Omega)$, and Lemma **??** in Chapter 5 shows that the corresponding operators $\{\mathcal{T}_\epsilon = e^{-\epsilon\mathcal{L}}\}$ are a uniformly bounded family of NIS operators of order zero. Moreover, since $\lim_{\epsilon \to 0^+} e^{-\epsilon\mathcal{L}}[\varphi] = \varphi$ with convergence in $L^2(\Omega)$ it follows that if $\psi \in \mathcal{C}_0^\infty(\Omega)$, then

$$\lim_{\epsilon \to 0^+} \left\langle \mathcal{T}_\epsilon[\varphi], \psi \right\rangle = \lim_{\epsilon \to 0^+} \int_\Omega e^{-\epsilon\mathcal{L}}[\varphi](x)\, \psi(x)\, dx = \int_\Omega \varphi(x)\, \psi(x)\, dx = \left\langle I[\varphi], \psi \right\rangle.$$

This completes the proof. □

PROPOSITION 1.3. *Let $\mathcal{T}$ be an NIS operator smoothing of order k.*

(1) If $I \in \mathbb{I}_l$, then $X^I\,\mathcal{T}$ and $\mathcal{T}\,X^I$ are NIS operators smoothing of order $k-l$.

(2) If $m \in \mathcal{C}^\infty(\overline{\Omega})$ and if $M[\varphi](x) = m(x)\,\varphi(x)$, then $M\,\mathcal{T}$ and $\mathcal{T}\,M$ are NIS operators smoothing of order k.

PROOF. This is clear from the definition. □

In particular, it follows that the multiplication operator M is an NIS operator of order zero, and the differential operator X^I is an NIS operator of order l. Note that the distribution kernels of these operators are supported on the diagonal of $\Omega \times \Omega$.

DEFINITION 1.4. *The homogeneous dimension of the control system defined by the vector fields $\{X_1, \dots, X_p\}$ is*

$$M = \sup\Big\{\alpha > 0 \,\Big|\, \big|B(x,2\delta)\big| \geq 2^\alpha \,\big|B(x,\delta)\big|\Big\}. \tag{1.8}$$

We establish certain estimates for integrals that occur frequently.

PROPOSITION 1.5. *Suppose that for all $x \in \Omega$ and $\delta > 0$ we have*

$$\big|B(x,2\delta)\big| \geq 2^M \,\big|B(x,\delta)\big|$$
$$\big|B(x,\tfrac{1}{2}\delta)\big| \leq 2^{-m}\big|B(x,\delta)\big|.$$

Suppose that $\alpha, \beta \in \mathbb{R}$. There is a constant $C_{\alpha,\beta}$ such that

$$\int_{\rho(x,y)>\delta} \rho(x,y)^\alpha\, V(x,y)^\beta\, \frac{dy}{V(x,y)} \leq C_{\alpha,\beta}\,\delta^\alpha\big|B(x,\delta)\big|^\beta \quad \text{if } \beta \leq 0 \text{ and } \alpha + M\beta < 0,$$
$$\int_{\rho(x,y)<\delta} \rho(x,y)^\alpha\, V(x,y)^\beta\, \frac{dy}{V(x,y)} \leq C_{\alpha,\beta}\,\delta^\alpha\big|B(x,\delta)\big|^\beta \quad \text{if } \beta \geq 0 \text{ and } \alpha + m\beta > 0.$$

PROOF. Let $\Omega_j = \Big\{y \in \Omega \,\Big|\, 2^j\delta < \rho(x,y) \leq 2^{j+1}\delta\Big\}$. We have

$$\begin{aligned}
\int_{\rho(x,y)>\delta} \rho(x,y)^\alpha\, V(x,y)^{\beta-1}\,dy &= \sum_{j=0}^{\infty}\int_{\Omega_j} \rho(x,y)^\alpha\, V(x,y)^{\beta-1}\,dy \\
&\approx C\,\delta^\alpha \sum_{j=0}^{\infty} 2^{j\alpha}\,\big|B(x,2^j\delta)\big|^\beta \\
&\leq C\,\delta^\alpha\big|B(x,\delta)\big|^\beta \sum_{j=0}^{\infty} 2^{j(\alpha+M\beta)} \\
&\leq C\,\delta^\alpha\,\big|B(x,\delta)\big|^\beta
\end{aligned}$$

provided $\alpha + M\beta < 0$. A similar argument demonstrates the second inequality. □

Let $T_j \in \mathcal{C}^\infty(\Omega\times\Omega)$ for $j = 1, 2$. Define $\mathcal{T}_j$, $\mathcal{T}_j^* : \mathcal{C}_0^\infty(\Omega) \to \mathcal{C}^\infty(\Omega)$ by setting

$$\mathcal{T}_j[\varphi](x) = \int_\Omega T_j(x,y)\,\varphi(y)\,dy, \tag{1.9}$$
$$\mathcal{T}_j^*[\varphi](x) = \int_\Omega T_j(y,x)\,\varphi(y)\,dy. \tag{1.10}$$

Suppose

(1) For each $K \in \mathbb{I}_k$, $L \in \mathbb{I}_l$ there is a constant $C_{j,K,L}$ so that

$$\big|X_x^K X_y^L T_j(x,y)\big| \leq C_{j,K,L}\,\rho(x,y)^{m-k-l} V(x,y)^{-1}. \tag{1.11}$$

(2) For each $K \in \mathbb{I}_k$ there is a positive integer N_K and a constant $C_{j,K}$ so that for all $x \in \Omega$ and all $\delta > 0$, if $\varphi \in \mathcal{C}_0^\infty\big(B(x,\delta)\big)$, then

$$\sup_{x\in\Omega}\big|X^K \mathcal{T}_j[\varphi](x)\big| \leq C_{j,K}\,\delta^{m_j-k} \sup_{y\in\Omega}\Big[\sum_{j=0}^{N_K}\delta^j \sum_{J\in\mathbb{I}_j}\big|X^J[\varphi](y)\big|\Big], \tag{1.12}$$

$$\sup_{x\in\Omega}\big|X^K \mathcal{T}_j^*[\varphi](x)\big| \leq C_{j,K}\,\delta^{m_j-k} \sup_{y\in\Omega}\Big[\sum_{j=0}^{N_K}\delta^j \sum_{J\in\mathbb{I}_j}\big|X^J[\varphi](y)\big|\Big]. \tag{1.13}$$

LEMMA 1.6. *Suppose that $m_1 + m_2 < M$. If $x \neq y$ the integral*

$$S(x,y) = \int_\Omega T_1(x,z)\,T_2(z,y)\,dz. \tag{1.14}$$

converges absolutely, and

$$\big|S(x,y)\big| \leq C\,\rho(x,y)^{m_1+m_2} V(x,y)^{-1}. \tag{1.15}$$

PROOF. Let $\rho(x,y) = \delta > 0$. Choose $\eta \in \mathcal{C}^\infty\big((0,\infty)\big)$ such that $\eta(s) = 1$ if $0 < s \leq \frac{1}{4}$, and $\eta(s) = 0$ if $s \geq \frac{1}{3}$. Put

$$\chi_1(z) = \eta\left(\frac{\rho(x,z)}{\delta}\right) \qquad \varphi_y(z) = \chi_1(z)\,T_2(z,y)$$
$$\chi_2(z) = \eta\left(\frac{\rho(z,y)}{\delta}\right) \qquad \psi_x(x) = \chi_2(z)\,T_1(x,z)$$

and

$$\chi_3(z) = 1 - \chi_1(z) - \chi_2(z).$$

Note that χ_1 and φ_y are supported in $B(x,\frac{1}{3}\delta)$ while χ_2 and ψ_x are supported in $B(y,\frac{1}{3}\delta)$. Using elementary metric geometry it follows that if $z \in B(x,\frac{1}{3}\delta)$ then $\rho(z,y) \approx \rho(x,y)$, while if $z \in B(y,\frac{1}{3}\delta)$, then $\rho(x,z) \approx \rho(x,y)$. Also, any derivative of χ_1 is supported where $\rho(x,z) \approx \rho(x,y)$ and any derivative of χ_2 is supported where $\rho(z,y) \approx \rho(x,y)$. Using Lemma **??** it is easy to check that for each $J \in \mathbb{I}_j$ there is an admissible constant C_J so that

Lemma on smooth distance functions

$$\sup_{z\in\Omega}\big|X^J[\varphi_y](z)\big| \leq C_J\delta^{-m_2-j}$$
$$\sup_{z\in\Omega}\big|X^J[\psi_x](z)\big| \leq C_J\,\delta^{-m_1-j}.$$

But now

$$\int_\Omega T_1(x,z)\,T_2(z,y)\,\chi_1(z)\,dz = \int_\Omega T_1(x,z)\,\varphi_y(z)\,dz = \mathcal{T}_1[\varphi_y](x)$$
$$\int_\Omega T_1(x,z)\,T_2(z,y)\,\chi_2(z)\,dz = \int_\Omega T_2^*(y,z)\,\psi_x(z)\,dz = \mathcal{T}_2^*[\psi_x](y)$$

It follows from equation (1.12) that

$$\big|T_1[\varphi_y](x)\big| \leq C\,\delta^{m_1} \sup_{z\in\Omega} \Big[\sum_{j=0}^{N_0} \delta^j \sum_{J\in\mathbb{I}_j} \big|X^J[\varphi_y](z)\big|\Big] \leq C'\delta^{m_1+m_2}V(x,y)^{-1}$$

$$\big|T_2^*[\psi_x](y)\big| \leq C\,\delta^{m_2} \sup_{z\in\Omega} \Big[\sum_{j=0}^{N_0} \delta^j \sum_{J\in\mathbb{I}_j} \big|X^J[\psi_x](z)\big|\Big] \leq C'\delta^{m_1+m_2}V(x,y)^{-1}.$$

To complete the proof, we need to estimate $\int_\Omega T_1(x,z)\,T_2(z,y)\,\chi_3(z)\,dz$. Now $\chi_3(z) \neq 0$ implies that $z \in \Omega_1 \cup \Omega_2$ where

$$\Omega_1 = \left\{z\in\Omega \,\Big|\, \frac{1}{4}\rho(x,y) \leq \rho(x,z) \leq 2\rho(x,y) \quad \text{and} \quad \rho(z,y) \geq \frac{1}{4}\rho(x,y)\right\},$$
$$\Omega_2 = \big\{z\in\Omega \,\big|\, \rho(x,z) > 2\rho(x,y)\big\}.$$

For $z \in \Omega_1$ we have $\rho(x,z) \approx \rho(z,y) \approx \rho(x,y) = \delta$, and $V(x,z) \approx V(z,y) \approx V(x,y)$. Moreover, the volume of Ω_1 is bounded by the volume of $B(x,2\delta) \approx V(x,y)$. Then using equation (1.11), we have

$$\left|\int_{\Omega_1} T_1(x,z)\,T_2(z,y)\,\chi_3(z)\,dz\right| \leq C\int_{\Omega_1} \rho(x,z)^{m_1}\rho(z,y)^{m_2}V(x,z)^{-1}V(z,y)^{-1}dz$$
$$\leq C\,\delta^{m_1+m_2}V(x,y)^{-1}.$$

For $z \in \Omega_2$ we have $\rho(x,z) \approx \rho(z,y)$ and $V(x,z) \approx V(z,y)$. Thus

$$\left|\int_{\Omega_2} T_1(x,z)\,T_2(z,y)\,\chi_3(z)\,dz\right| \leq C\int_{\Omega_2} \rho(x,z)^{m_1}\rho(z,y)^{m_2}V(x,z)^{-1}V(z,y)^{-1}dz$$
$$\leq C\int_{\rho(x,z)>2\rho(x,y)} \rho(x,z)^{m_1+m_2}V(x,z)^{-2}\,dz$$
$$\leq \rho(x,y)^{m_1+m_2}\,V(x,y)^{-1}$$

provided that $m_1 + m_2 < d$. □

LEMMA 1.7. *Suppose that $\mathcal{T}_j$ is an NIS operator of order m_j for $j = 1, 2$. Then if $m_1 + m_2 < d$, the operator $\mathcal{T}_1\mathcal{T}_2$ is an NIS operator of order $m_1 + m_2$.*

PROOF. Let $\{\mathcal{T}_{j,\epsilon}\}_{\epsilon>0}$ be the regularized operators which approximate $\mathcal{T}_j$, and let $T_{j,\epsilon}(x,y)$ be the corresponding distribution kernels. If $\varphi \in \mathcal{C}_0^\infty(\Omega)$, then

$$\begin{aligned}\mathcal{T}_{1,\epsilon}\mathcal{T}_{2,\epsilon}[\varphi](x) &= \int_\Omega T_{1,\epsilon}(x,z)\,\mathcal{T}_{2,\epsilon}[\varphi](z)\,dz\\ &= \int_\Omega T_{1,\epsilon}(x,z)\,\left[\int_\Omega T_{2,\epsilon}(z,y)\,\varphi(y)\,dy\right]\,dz\\ &= \int_\Omega \varphi(y)\,\left[\int_\Omega T_{1,\epsilon}(x,z)\,T_{2,\epsilon}(z,y)\,dz\right]\,dy\end{aligned}$$

where the interchange of order of integration is justified since the inner integral on the last line converges uniformly. Thus the distribution kernel of the composition $\mathcal{T}_{1,\epsilon}\mathcal{T}_{2,\epsilon}$ is given by

$$S_\epsilon(x,y) = \int_\Omega T_{1,\epsilon}(x,z)\,T_{2,\epsilon}(z,y)\,dz.$$

Lemma 1.6 shows that S_ϵ and its derivatives satisfy the correct size estimates for an NIS operator of order $m_1 + m_2$. □

Finally, we check the action of the operator $T_1 T_2$ on a bump function φ supported in a ball $B(x,\delta)$. We can write

$$1 = \sum_{k=0}^{\infty} \psi_k(z)$$

with ψ_k is supported in $B(x, C\,2^k\,\delta)\backslash B(x, 2^k\,\delta)$, C a large constant, and $|X^J \psi_k| \lesssim (2^k \delta)^{-|J|}$. We study the function $\psi_k(z)\, T_2[\varphi](z)$. Using the bump function condition, we have

$$\left|X^J T_2[\varphi](z)\right| \lesssim \delta^{m_2 - |J|} \sum \delta^{|K|} ||X^K \varphi||_\infty$$

when $d(z,x) \leq 2\,\delta$. On the other hand, if $d(z,x) \geq 2^k\,\delta$ then using only size estimates we have

$$\left|X^J T_2[\varphi](z)\right| \lesssim (2^k\,\delta)^{m_2 - |I|} |B(x, 2^k\,\delta)|^{-1} |B(x,\delta)|.$$

Using Leibniz' rule, it follows that

$$\left|X^J \big(\psi_k T_2[\varphi]\big)(z)\right| \lesssim (2^k\,\delta)^{m_2 - |I|} |B(x, 2^k\,\delta)|^{-1} |B(x,\delta)| \sum \delta^{|K|} ||X^K \varphi||_\infty.$$

Now $T_1\, T_2[\varphi] = \sum_{k=0}^{\infty} T_1 \Big[\psi_k\, T_2[\varphi]\Big]$. Using the bump function condition, we have

$$\begin{aligned}\sup_{z\in\Omega} \left|T_1 \Big[\psi_k\, T_2[\varphi]\Big](z)\right| &\lesssim (2^k\,\delta)^{m_1}\, (2^k\,\delta)^{m_2} |B(x, 2^k\,\delta)|^{-1} |B(x,\delta)| \sum \delta^{|K|} ||X^K \varphi||_\infty \\ &= (2^k\,\delta)^{m_1+m_2} \frac{|B(x,\delta)|}{|B(x, 2^k\,\delta)|} \sum \delta^{|K|} ||X^K \varphi||_\infty.\end{aligned}$$

If we sum in k we get the correct estimate $\delta^{m_1+m_2} \sum \delta^{|K|} ||X^K \varphi||_\infty$.

CHAPTER 7

Algebras

In this chapter we present an account of material on free associative and Lie algebras that is needed elsewhere. This theory can also be found in many other places, including Bourbaki [**Bou89**], Humphreys [**Hum72**], and Jacobson [**Jac62**]. Our development is very close to this last reference. We have tried to make this material as concrete as possible. In particular, we shall only consider algebras over the field of real numbers $\mathbb{R}$.

1. Associative Algebras

An *associative algebra* is a vector space A over $\mathbb{R}$ with a product operation written $(x, y) \to xy$ which is distributive and associative. Thus if x, y, $z \in A$ and $\alpha \in \mathbb{R}$, we have

$$\begin{aligned} \alpha(xy) &= (\alpha x)y = x(\alpha y), \\ x(y+z) &= xy + xz, \\ (y+z)x &= yx + zx, \\ (xy)z &= x(yz). \end{aligned} \tag{1.1}$$

An element e is an *identity element* for A if $ex = xe = x$ for all $x \in A$. A *subalgebra* of A is a subspace which is closed under the product operation. The algebra A is *commutative* if $xy = yx$ for all $x, y \in A$. If A and B are associative algebras, a linear mapping $f : A \to B$ is an *algebra homomorphism* if for all $x, y \in A$ we have $f(xy) = f(x)\,f(y)$.

1.1. Free Associative Algebras.

If $S = \{x_i\}_{i \in I}$ is a set of elements, we want to construct the free associative algebra generated by the elements of S. Intuitively, this should be the smallest algebra $A(S)$ containing the elements of S in which the only identities in the algebra follow from the identities in the definition (1.1). Thus the algebra should contain all finite linear combinations of formal products of finitely many elements of S, and different formal products should give different elements of $A(S)$.

Let us be more precise. Define a non-commuting monomial of degree m in the elements of S to be a symbol $x_{i_1} x_{i_2} \cdots x_{i_m}$. Two such monomials $x_{i_1} x_{i_2} \cdots x_{i_m}$ and $x_{j_1} x_{j_2} \cdots x_{j_n}$ are the same if and only $m = n$ and the ordered m-tuples $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots, j_m\}$ of elements of the index set I are the same. In particular, if $x_i \neq x_j$ are two different elements of S, then $x_i x_j \neq x_j x_i$. The free associative algebra $A(S)$ generated by S is then the real vector space with basis given by the set of all symbols $\{x_{i_n} x_{i_2} \cdots x_{i_m}\}$ for $m \geq 1$, and an extra symbol e which will act as an identity element. This means that every element of $A(S)$ can be written uniquely as a finite linear combination of these symbols. The product structure in the algebra $A(S)$ is defined by first requiring that

$$\left(x_{i_1}\, x_{i_2}\, \cdots\, x_{i_m}\right)\left(x_{j_1}\, x_{j_2}\, \cdots\, x_{j_n}\right) = \left(x_{i_1}\, x_{i_2}\, \cdots\, x_{i_m}\, x_{j_1}\, x_{j_2}\, \cdots\, x_{j_n}\right)$$

and

$$e\left(x_{i_1}\, x_{i_2}\, \cdots\, x_{i_m}\right) = \left(x_{i_1}\, x_{i_2}\, \cdots\, x_{i_m}\right) e = \left(x_{i_1}\, x_{i_2}\, \cdots\, x_{i_m}\right),$$

Products of linear combinations of basis elements are then defined by using the distributive laws in equation 1.1. In other words, $A(S)$ is the space of polynomials in the non-commuting variables $\{x_i\}$. The general element of $A(S)$ has the form

$$P(x) = \alpha_0 e + \sum_i \alpha_i x_i + \sum_{i_1,i_2} \alpha_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_m} \alpha_{i_1,\ldots,i_m} x_{i_1} \cdots x_{i_m} \qquad (1.2)$$

where $\sum_{i_1,\ldots,i_k} \alpha_{i_1,\ldots,i_k} x_{i_1} \cdots x_{i_k}$ denotes the sum over all ordered k-tuples of elements of S, and only finitely many of the coefficients $\{\alpha_{i_1,\ldots,i_k}\}$ are non-vanishing. The *degree* of a polynomial $P(x)$ is the largest integer m such that $\alpha_{i_1,\ldots,i_m} \neq 0$ for some coefficient of P. If P is given as in (1.2), the sum

$$P_k(x) = \sum_{i_1,\ldots,i_k} \alpha_{i_1,\ldots,i_k} x_{i_1} \cdots i_k$$

is the component of P which is *homogeneous of degree* k. Every polynomial is thus a sum of homogeneous components.

The absence of extraneous identities in $A(S)$ should mean that if B is any associative algebra, and if $b_i \in B$ for $i \in I$, then there should not be any obstruction to the existence of an algebra homomorphism $\Phi : A(S) \to B$ such that $\Phi(x_i) = b_i$. Rather than rely on an explicit construction as we have just done, it is often convenient to define the free associative algebra by a corresponding universal mapping property. Thus we make the following formal definition.

Definition 1.1. *$A(S)$ is a* free associative algebra generated by *S if*

(a) $S \subset A(S)$;
(b) For every associative algebra B and every mapping $\varphi : S \to B$, there is a unique algebra homomorphism $\Phi : A(S) \to B$ such that $\Phi(x) = \varphi(x)$ for every $x \in S$.

We can also consider the real vector space V with basis S. Then any map $\varphi : S \to B$ extends to a unique linear mapping $\varphi : V \to B$. Thus we say that $A(V)$ is the free associative algebra generated by a vector space *V if*

(a′) $V \subset A(V)$;
(b′) For every associative algebra B and every linear mapping $\varphi : V \to B$, there is a unique algebra homomorphism $\Phi : A(V) \to B$ such that $\Phi(v) = \varphi(v)$ for every $v \in V$.

It is easy to see that if they exist, objects defined by this kind of universal mapping property are unique up to isomorphism. This is explored in the exercises.

An alternate description of an example of the free associative algebra $A(S)$ is given in terms of tensor products. Let $S = \{x_i\}_{i \in I}$ be a set, and let $V = V(S)$ denote the real vector space with basis S. We set $V^0 = \mathbb{R}$, $V^1 = V$, and for $k \geq 2$, $V^k = V \otimes V \cdots \otimes V$, the tensor product of k copies of V. A basis for V^k is provided by the set of all symbols $\{x_{i_1} \otimes \cdots \otimes x_{i_k}\}$.

The *tensor algebra* over V is then

$$T(V) = \bigoplus_{k=0}^{\infty} V^k = \mathbb{R} \oplus V \oplus V^2 \oplus \cdots \oplus V^k \oplus \cdots . \tag{1.3}$$

It is understood that every element in this direct sum has only finitely many non-vanishing components. Thus an element of $y \in T(V)$ can be written as a finite sum

$$y = \alpha_0 \oplus \sum_{i\in I} \alpha_i \, x_i \oplus \sum_{i_1,i_2\in I} \alpha_{i_1,i_2}(x_{i_1} \otimes x_{i_2}) \oplus \cdots \oplus \sum_{i_1,\ldots,i_m\in I} \alpha_{i_1,\ldots,i_m}(x_{i_1} \otimes \cdots \otimes x_{i_m})$$

We can make $T(V)$ into an algebra $A(S)$ as follows. The product of a basis element of V^k with a basis element of V^l is defined by setting

$$(x_{i_1} \otimes \cdots \otimes x_{i_k})(x_{j_1} \otimes \cdots \otimes x_{j_l}) = (x_{i_1} \otimes \cdots \otimes x_{i_k} \otimes x_{j_1} \otimes \cdots \otimes x_{j_l}) \in V^{k+l}.$$

This extends to a bilinear mapping $V^k \times V^l \to V^{k+l}$. We then define the product of two elements of $T(V)$ by requiring that the product be distributive in each entry. The identity element e of $T(V)$ is then identified with $1 \in \mathbb{R}$. It should be clear that the algebra $T(V)$ is the same (or is isomorphic to) the algebra $A(S)$ of non-commuting polynomials we considered earlier.

LEMMA 1.2. *Let $S = \{x_i\}_{i\in I}$, and let V be the real vector space with basis S. Let B be any associative algebra, and let $\varphi : S \to B$. Then there is a unique algebra homomorphism $\Phi : T(V) \to B$ such that $\Phi(x_i) = \varphi(x_i)$ for all $i \in I$. Thus the algebra $T(V)$ is a free associative algebra generated by the set S.*

PROOF. Let $\varphi(x_i) = b_i \in B$. Then define

$$\Phi\Big(\alpha_0 \oplus \sum_{i\in I} \alpha_i \, x_i \oplus \sum_{i_1,i_2\in I} \alpha_{i_1,i_2}(x_{i_1} \otimes x_{i_2}) \oplus \cdots\Big) = \alpha_0 + \sum_{i\in I} \alpha_i \, b_i + \sum_{i_1,i_2\in I} \alpha_{i_1,i_2} b_{i_1} b_{i_2} + \cdots .$$

It is clear that Φ is the unique algebra homomorphism extending φ. □

1.2. Algebras of formal series.

In addition to considering polynomials in a set S of non-commuting variables $\{x_i\}_{i\in I}$, it is often necessary to consider formal infinite series as well. Thus let $\widetilde{A}(S)$ denote the set of all formal infinite series

$$u = \alpha_0 + \sum_{i} \alpha_i x_i + \sum_{i_1,i_2} \alpha_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_m} \alpha_{i_1,\ldots,i_m} x_{i_1} \cdots x_{i_m} + \cdots .$$

As before, $\sum_{i_1,\ldots,i_k} \alpha_{i_1,\ldots,i_k} x_{i_1} \cdots x_{i_k}$ is a sum over all ordered k-tuples of elements of S with at most finitely many non-vanishing terms. However, the sum defining u is no longer finite. We allow terms with arbitrarily large homogeneity. Note that there is no requirement of convergence; the element u can be regarded simply as a collection of real numbers $\{\alpha_{i_1,\cdots,i_m}\}$ indexed by $m \geq 0$ and the collection of all finite ordered subsets of m elements of S. The real number α_0 is called the *constant term* of u. The *order* of u is the smallest integer $m \geq 0$ such that there exists a coefficient $\alpha_{i_1,\ldots,i_m} \neq 0$.

We shall also write $\widetilde{T}(V)$ for this set if V is the real vector space with basis S. Then $\widetilde{T}(V) = \widetilde{A}(S)$ is a real vector space. If u is given as above and

$$v = \beta_0 + \sum_i \beta_i x_i + \sum_{i_1,i_2} \beta_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_m} \beta_{i_1,\ldots,i_m} x_{i_1} \cdots x_{i_m} + \cdots .$$

then

$$u + v = (\alpha_0 + \beta_0) + \cdots + \sum_{i_1,\ldots,i_m} (\alpha_{i_1,\ldots,i_m} + \beta_{i_1,\ldots,i_m}) x_{i_1} \cdots x_{i_m} + \cdots$$

and if $\lambda \in \mathbb{R}$

$$\lambda\, u = \lambda\, \alpha_0 + \cdots + \sum_{i_1,\ldots,i_m} \lambda\, \alpha_{i_1,\ldots,i_m} x_{i_1} \cdots x_{i_m} + \cdots .$$

$\widetilde{A}(S)$ is also an algebra where the product is given by formal multiplication of series. Thus $u\, v = w$ where

$$w = \gamma_0 + \sum_{i\in I} \gamma_i x_i + \sum_{i_1,i_2\in I} \gamma_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_m\in I} \gamma_{i_1,\ldots,i_m} x_{i_1} \cdots x_{i_m} + \cdots$$

and

$$\gamma_{i_1,\ldots,i_m} = \alpha_0\, \beta_{i_1,\ldots,i_m} + \alpha_{i_1,\ldots,i_m}\, \beta_0 + \sum_{j=1}^{m-1} \alpha_{i_1,\ldots,i_j}\, \beta_{i_{j+1},\ldots,i_m}. \tag{1.4}$$

Proposition 1.3.

(1) If $u, v \in \widetilde{A}(S) = \widetilde{T}(V)$, then order$(u\, v)$ = order(u) + order(v).
(2) An element $u \in \widetilde{A}(S) = \widetilde{T}(V)$ is invertible if and only if the constant term of u is non-zero.

Proof. If order$(u) = m$ and order$(v) = n$, let $\alpha_{i_1,\ldots,i_m}$ and $\beta_{j_1,\ldots,j_n}$ be corresponding non-vanishing coefficients. Then (1.4) show that that $\gamma_{i_1,\ldots,i_m,j_1,\ldots,j_n} = \alpha_{i_1,\ldots,i_m}\beta_{j_1,\ldots,j_n} \neq 0$, so the order of uv is at most $m+n$. On the other hand, equation (1.4) also shows that all lower order coefficients are zero, since in each product, one factor must be zero. This establishes (1).

To prove (2), suppose that $u(x) = \alpha_0 + u_1(x)$ where $\alpha \neq 0$ and u_1 is of order at least 1. We show by induction on N that there is a non-commutative homogeneous polynomial $P_N(x)$ of degree N so that so that $u\left(\sum_{k=0}^N P_k(x)\right) - 1$ is of order at least $N+1$. It will then follow that $\sum_{k=0}^\infty P_k(x)$ is a right inverse for u. A similar argument gives a left inverse, which then must be the same as the right inverse.

We let $P_0(x) = \alpha^{-1}$. Then $u(x)\, P_0(x) - 1 = \alpha_0^{-1} u_1(x)$ is of order at least 1. Suppose we have constructed homogeneous non-commuting polynomials $\{P_0, \ldots, P_N\}$ so that

$$u(x) \Big(\sum_{k=0}^N P_k(x) \Big) - 1 = Q_{N+1}(x) + R_{N+2}(x)$$

where each $Q_{N+1}(x)$ is a non-commuting polynomial of degree $N+1$ and $R_{N+2}(x)$ has order at least $N+2$. Put $P_{N+1}(x) = -\alpha_0^{-1} Q_{N+1}(x)$. Then

$$\begin{aligned} u(x)\Big(\sum_{k=0}^{N+1} P_k(x)\Big) - 1 &= u(x)\Big(\sum_{k=0}^{N} P_k(x)\Big) - 1 + u(x)\, P_{N+1}(x) \\ &= Q_{N+1}(x) + R_{N+2}(x) - \alpha_0^{-1}\big(\alpha_0 + u_1(x)\big)\, Q_{N+1}(x) \\ &= R_{N+2}(x) - u_1(x)\, Q_{N+1}(x). \end{aligned}$$

This has order at least $N+2$, which completes the proof. □

Note that if $v(x) \in \widetilde{A}(S) = \widetilde{T}(V)$ has order at least 1, then $v(x)^n$ has order at least n. If $w(y) = \sum_{k=0}^{\infty} \alpha_k y^k$ is a formal infinite series in a single variable y, then the composition

$$w \circ v(x) = \sum_{k=0}^{\infty} \alpha_k [v(x)]^n$$

is an element of $\widetilde{A}(S)$. We shall be particularly concerned with two formal series which define the exponential and the logarithm. Thus if y is an indeterminate, we set

$$\exp(y) = 1 + y + \frac{1}{2!} y^2 + \cdots = \sum_{n=0}^{\infty} y^n.$$

If z is an indeterminate, we se

$$\log(1+z) = z - \frac{1}{2} z^2 + \frac{1}{3} z^3 - \cdots = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} z^m.$$

Note that $\exp(x) - 1$ has order 1 so it makes sense to consider the composition $\log\big(\exp(x)\big)$, and $\log(1+z)$ has order 1, so it makes sense to consider the composition $\exp\big(\log(1+z)\big)$. The following result then follows from standard calculations.

Proposition 1.4. *If $S = \{x\}$, then in the algebra $\widetilde{A}(S)$ of formal power series in x we have*

$$\log\big(\exp(x)\big) = x$$
$$\exp\big(\log(1+x)\big) = 1 + x$$

If $S = \{x, y\}$ and if x and y commute (so that $xy = yx$) then

$$\exp(x+y) = \exp(x)\, \exp(y)$$
$$\log(1+x) + \log(1+y) = \log\big((1+x)(1+y)\big)$$

2. Lie Algebras

A *Lie algebra* $\mathfrak{L}$ is a vector space over $\mathbb{R}$ with a product operator written $(x, y) \to [x, y]$ satisfying the following conditions. First, if $x,\ y,\ z \in \mathfrak{L}$, the analogues of the first three identities in (1.1) hold:

$$\begin{aligned} \alpha[x, y] &= [\alpha x, y] = [x, \alpha y] \\ [x, y+z] &= [x, y] + [x, z] \\ [y+z, x] &= [y, x] + [z, x] \end{aligned} \tag{2.1}$$

In addition, the associative property is replaced by the following identities for all $x, y, z \in \mathfrak{L}$:

$$\begin{aligned} 0 &= [x, y] + [y, x] \\ 0 &= [x, [y, z]] + [y, [z, x]] + [z, [x, y]] . \end{aligned} \tag{2.2}$$

The second of these is called the *Jacobi identity.* As with associate algebras, a subspace of $\mathfrak{L}$ which is closed under the bracket operation is a subalgebra. We say that a Lie algbera $\mathfrak{L}$ is generated by a subset S if there is no proper Lie subalgebra of $\mathfrak{L}$ which contains S. If $\mathfrak{L}$ and $\mathfrak{M}$ are Lie algebras, a linear mapping $g : \mathfrak{L} \to \mathfrak{M}$ is a *Lie algebra homomorphism* if for all $x, y \in \mathfrak{L}$ we have $g\big([x, y]\big) = [g(x),\, g(y)]$.

An important class of examples of Lie algebras arises as follows. If A is any associative algebra, we define

$$[x,\, y] = xy - yx.$$

Then the vector space A with this new product is a Lie algebra, which we write A_L. It is clear that $[x,\, y] = -[y,\, x]$, and the Jacobi identity follows from the computation

$$\begin{aligned} &[x,\, [y,\, z]] + [y,\, [z,\, x]] + [z,\, [x,\, y]] \\ &\qquad = x\,[y,\, z] - [y,\, z]\, x + y\,[z,\, x] - [z,\, x]\, y + z\,[x,\, y] - [x,\, y]\, z \\ &\qquad = xyz - xzy - yzx + zyx \\ &\qquad\quad + yzx - yxz - zxy + xzy \\ &\qquad\qquad + zxy - zyx - xyz + yxz = 0. \end{aligned}$$

For example, if V is a vector space and $\mathcal{L}(V)$ denotes the space of linear mapping from V to itself, then $\mathcal{L}(V)$ is an associative algebra where the product is composition of mappings. $\mathcal{L}(V)_L$ is then the corresponding Lie algebra.

If $\mathfrak{L}$ is a Lie algebra, then for every $x \in \mathfrak{L}$ we define a linear mapping $ad_x \in \mathcal{L}(\mathfrak{L})$ by settng

$$\mathrm{ad}_{\,x}(z) = [x, z]. \tag{2.3}$$

Note that

$$\begin{aligned} \mathrm{ad}_{\,[x,y]}(z) &= [[x,\, y]\,,\, z] \\ &= [x, [y, z]] - [y, [x, z]] \\ &= (\mathrm{ad}_{\,x}\mathrm{ad}_{\,y} - \mathrm{ad}_{\,y}\mathrm{ad}_{\,x})(z) \\ &= \big[\mathrm{ad}_{\,x}, \mathrm{ad}_{\,y}\big](z). \end{aligned}$$

Thus ad is a Lie algebra homomorphism of $\mathfrak{L}$ into the Lie algebra $\mathcal{L}(\mathfrak{L})_L$.

2.1. Iterated Commutators.

In analogy with the associative situation, we want to construct a *free Lie algebra* generated by a set of elements $S = \{x_i\}_{i \in I}$. This should be a Lie algebra $\mathfrak{L}(S)$ containing the set S with the property that if $\mathfrak{M}$ is any Lie algebra with $a_i \in \mathfrak{M}$, there is a unique Lie algebra homomorphism $\varphi : \mathfrak{L}(S) \to \mathfrak{M}$ with $\varphi(x_i) = a_i$ for all $i \in I$. The construction of $\mathfrak{L}(S)$ is more delicate that the construction of $A(S)$ because multiplication is not associative, and it is considerably more difficult to decide whether or not a particular relation between formal Lie products is a consequence of the Jacobi identity.

We begin by observing that the failure of associativity means that an ordered sequence of elements can be multiplied in many different ways. Thus if $\{x, y, z\}$ is an ordered triple of elements in a Lie algebra $\mathfrak{L}$, we can construct two possibly different products

$$[x,\, [y,\, z]] \qquad \text{and} \qquad [[x,\, y]\,,\, z]\,,$$

while if $\{x, y, z, w\}$ is an ordered 4-tuple, we can construct five *a priori* different products:

$$[x,\, [y,\, [z,\, w]]] \quad [x,\, [[y,\, z]\,,\, w]] \quad [[x,\, y]\,,\, [z,\, w]] \quad [[x,\, [y,\, z]]\,,\, w] \quad [[[x,\, y]\,,\, z]\,,\, w]$$

In order to deal with this plethora of products, we define certain special one. Thus the *iterated* commutator of length m of the elements $\{x_1, \dots, x_m\} \subset \mathfrak{L}$ is the element

$$[x_1,\, [x_2,\, [x_3,\, \cdots,[x_{m-1}, x_m] \cdots]]].$$

PROPOSITION 2.1. *Let $\{x_1, \dots, x_m, y_1, \dots, y_n\} \subset \mathfrak{L}$. Then the product of two iterated commutators*

$$\big[[x_1,\, [x_2,\, [x_3,\, \cdots,[x_{m-1}, x_m] \cdots]]]\,,\, [y_1,\, [y_2,\, [y_3,\, \cdots,[y_{n-1}, y_n] \cdots]]]\big]$$

can be written as a linear combination of iterated commutators.

PROOF. We argue first by induction on $m + n$ and then by induction on n. If $m + n = 2$, then $m = n = 1$ and $[x, y]$ is an iterated commutator. Thus suppose the assertion is true for any product with $m + n - 1$ terms, and consider a product of an iterated commutator of length m with one of length n. If $n = 1$ we write

$$\begin{aligned}\big[[x_1,\, [x_2,\, [x_3,\, \cdots,&[x_{m+n-2}, x_{m+n-1}] \cdots]]]\,, y_1\big] \\ &= -\big[y_1, [x_1,\, [x_2,\, [x_3,\, \cdots,[x_{m+n-2}, x_{m+n-1}] \cdots]]]\end{aligned}$$

which is an iterated commutator of length $m + n$.

Thus suppose the assertion is true for every product of an iterated commutator of length $m + 1$ with an iterated commutator of length $n - 1$. Using the Jacobi identity, we write

$$\begin{aligned}\big[[x_1,\, [x_2, [x_3,\, \cdots,&[x_{m-1}, x_m] \cdots]]]\,,\, [y_1,\, [y_2,\, [y_3,\, \cdots,[y_{n-1}, y_n] \cdots]]]\big] \\ &= -\big[y_1,\, \big[P,\, Q\big]\big] - \big[P,\, \big[Q, y_1\big]\big]\end{aligned}$$

where $P = [y_2,\, [y_3,\, \cdots,[y_{n-1}, y_n] \cdots]]$ and $Q = [x_1,\, [x_2,\, [x_3,\, \cdots,[x_{m-1}, x_m] \cdots]]]$. But $[P,\, Q]$ can be written as a linear combination of iterated commutators by the induction hypothesis on $m + n$, and so the same is true of the term $\big[y_1,\, \big[P,\, Q\big]\big]$. On the other hand

$$\big[P,\, \big[Q, y_1\big]\big] = \big[[y_1, Q],\, P\big],$$

and this last term can be written as a linear combination of iterated commutators by the induction hypothesis on n. □

COROLLARY 2.2. *If a Lie algebra is generated by a set $S = \{x_i\}_{i \in I}$, then every element of $\mathfrak{L}$ can be written as a linear combination of iterated commutators $[x_{i_1},\, [x_{i_2}, \dots,\, [x_{i_{m-1}}, x_{i_m}] \dots]]$, $1 \le m$.*

Even with this simplification, it is not immediately clear how to give a precise description of the free Lie algebra generated by elements $\{x_i\}_{i \in I}$. Consider the problem of describing a basis for the free Lie algebra generated by two elements x and y. There is only one linearly independent iterated commutator of length 2 since $[x,x] = [y,y] = 0$ and $[x,y] = -[y,x]$. There are two linearly independent iterated commutators of length 3. One choice is the set $\big\{[x,[x,y]],\ [y,[x,y]]\big\}$. It is then clear that the set of four iterated commutators of length four given by

$$\Big\{[x,[x,[x,y]]],\ [y,[x,[x,y]]],\ [x,[y,[x,y]]],\ [y,[y,[x,y]]]\Big\}$$

spans the set of all commutators of length four. However, these four are not linearly independent, since

$$[y,[x,[x,y]]] = -[x,[[x,y],y]] - [[x,y],[x,y]] = -[x,[[x,y],y]] + 0 = [x,[y,[x,y]]].$$

We will return to the problem of describing a basis for a free Lie algebra below in Section 6, after we have given a precise definition of this concept. And in order to do this, we first study the problem of imbedding Lie algebras into the Lie algebra structure of an associative algebra.

3. Universal Enveloping Algebras

Let $\mathfrak{L}$ be a Lie algebra.

Definition 3.1. *A universal enveloping algebra for $\mathfrak{L}$ is an associative algebra $\mathfrak{U}(\mathfrak{L})$ and a Lie algebra homomorphism $i : \mathfrak{L} \to \mathfrak{U}(\mathfrak{L})_L$ with the following universal property. If A is any associative algebra, and if $\varphi : \mathfrak{L} \to A_L$ is a Lie algebra homomorphism, then there exists a unique associative algebra homomorphism $\Phi : \mathfrak{U}(\mathfrak{L})_L \to A$ so that $\Phi \circ i = \varphi$.*

As with all universal definitions, if a universal enveloping algebra for $\mathfrak{L}$ exists, it is unique up to associative algebra isomorphism.

3.1. Construction of $\mathfrak{U}(\mathfrak{L})$.

To construct an enveloping algebra for an arbitrary Lie algebra $\mathfrak{L}$, consider the tensor algebra $T(\mathfrak{L})$ over the vector space $\mathfrak{L}$ given by

$$T(\mathfrak{L}) = \mathbb{R} \oplus \mathfrak{L}^1 \oplus \mathfrak{L}^2 \oplus \cdots \oplus \mathfrak{L}^k \oplus \cdots .$$

This is just the free associative algebra generated by $\mathfrak{L}$, and there is the standard inclusion $i : \mathfrak{L} \to T(\mathfrak{L})$ which identifies $\mathfrak{L}$ with $\mathfrak{L}^1$. Let $\mathcal{I}$ denote the two-sided ideal in $T(\mathfrak{L})$ generated by all elements of the form $x \otimes y - y \otimes x - [x,y]$ where $x, y, [x,y] \in \mathfrak{L}$. Thus $v \in \mathcal{I}$ if and only if $v = z\big(x \otimes y - y \otimes x - [x,y]\big)w$ where $x, y \in \mathfrak{L}$ and $z, w \in T(\mathfrak{L})$. In particular, $\mathcal{I} \cap \mathbb{R} = \emptyset$.

Let $\mathfrak{U}(\mathfrak{L}) = T(\mathfrak{L})/\mathcal{I}$ be the quotient algebra, and let $\pi : T(\mathfrak{L}) \to \mathfrak{U}(\mathfrak{L})$ be the standard quotient map. Define $\phi = \pi \circ i : \mathfrak{L} \to \mathfrak{U}(\mathfrak{L})$. Since $x \otimes y - y \otimes x - [x,y] \in \mathcal{I}$, it follows that $\phi\big([x,y]\big) = \big[\phi(x), \phi(y)\big]$, so ϕ is a Lie algebra homomorphism of $\mathfrak{L}$ into $\mathfrak{U}(\mathfrak{L})_L$.

Now let A be any associative algebra and let $\varphi : \mathfrak{L} \to A_L$ be a Lie algebra homomorphism. By the universal property of the free algebra $T(\mathfrak{L})$ given in Lemma 1.2, there is a unique associative algebra homomorphism $F : T(\mathfrak{L}) \to A$ so that $F = \varphi$ on $\mathfrak{L}^1$. If $x, y \in \mathfrak{L}$ we have

$$\begin{aligned} F\big(x \otimes y - y \otimes x - [x,y]\big) &= F(x)\,F(y) - F(y)\,F(x) - F\big([x,y]\big) \\ &= \big[\varphi(x), \varphi(y)\big]_A - \varphi\big([x,y]\big) = 0 \end{aligned}$$

since φ is a Lie algebra homomorphism. Thus we have shown

LEMMA 3.2. *The associative algebra* $\mathfrak{U}(\mathfrak{L})$ *is a universal enveloping algebra for the Lie algebra* $\mathfrak{L}$.

3.2. The Poincaré-Birkhoff-Witt Theorem.

Although the construction of the universal enveloping algebra $\mathfrak{U}(\mathfrak{L})$ is not difficult, several important questions remain open. In particular, it remains to prove that the mapping φ is one-to-one so that $\mathfrak{L}$ can be regarded as a subspace of $\mathfrak{U}(\mathfrak{L})$. This fact is a consequence of the Poincaré-Birkhoff-Witt theorem, which describes a basis for the algebra $\mathfrak{U}(\mathfrak{L})$ as a real vector space.

We begin by choosing a basis $S = \{x_i\}_{i\in I}$ for the real vector space $\mathfrak{L}$. For $k \geq 1$, the real vector space $T^k(\mathfrak{L})$ has a basis over $\mathbb{R}$ consisting of the "monomials" $\{x_{i_1} \otimes \cdots \otimes x_{i_k}\}$. The projection $\pi : T(\mathfrak{L}) \to \mathfrak{U}(\mathfrak{L})$ is surjective, so the images $\{\pi\big(x_{i_1} \otimes \cdots \otimes x_{i_k}\big)\}$, together with $e = \pi[1]$, certainly span the universal enveloping algebra.

To find a linearly independent subset of these elements, we give the index set I a total order. Then by definition, the monomial $\{x_{i_1} \otimes \cdots \otimes x_{i_k}\}$ is *standard* if $i_1 \leq i_2 \leq \cdots \leq i_k$. More generally, if $i, j \in I$, set

$$\eta_{i,j} = \begin{cases} 0 & \text{if } i \leq j, \\ 1 & \text{if } j < i. \end{cases}$$

Define the index of any monomial $x_{i_1} \otimes \cdots \otimes x_{i_k}$ to be

$$\operatorname{Ind}(x_{i_1} \otimes \cdots \otimes x_{i_k}) = \sum_{j<l} \eta_{i_j, i_l}.$$

The index measures the number of pairs of indices $\{j, l\}$ at which the monomial $x_{i_1} \otimes \cdots \otimes x_{i_k}$ fails to be standard.

LEMMA 3.3. *The set consisting of* e *and the images* $\{\pi\big(x_{i_1} \otimes \cdots \otimes x_{i_k}\big)\}$ *of all standard monomials spans* $\mathfrak{U}(\mathfrak{L})$.

PROOF. It suffices to prove that the image of every monomial $x_{i_1} \otimes \cdots \otimes x_{i_k}$ is a linear combination of the images of standard monomials. We prove this by induction on the length k of the monomial, and then by induction on the index of monomials of length k.

Observe that every monomial of length 1 is standard, so the case $k = 1$ is easy. Thus suppose that $k > 1$ and that the image of every monomial of length at most $k - 1$ is a linear combination of images of standard monomials. Let $x_{i_1} \otimes \cdots \otimes x_{i_k}$ be a monomial of length k. If this monomial has index zero, then it is standard, and again there is nothing to prove. Thus assume that $\operatorname{Ind}(x_{i_1} \otimes \cdots \otimes x_{i_k}) = \ell \geq 1$, and that the image of every monomial of length at most k and index at most $\ell - 1$ is a linear combination of images of standard monomials.

Since $\operatorname{Ind}(x_{i_1} \otimes \cdots \otimes x_{i_k}) \geq 1$, there is at least one integer j with $1 \leq j \leq k-1$ so that $i_{j+1} < i_j$. Then in $T(\mathfrak{L})$ we have the identity

$$\begin{aligned} x_{i_1} &\otimes \cdots \otimes x_{i_j} \otimes x_{i_{j+1}} \otimes \cdots \otimes x_{i_k} \\ &= x_{i_1} \otimes \cdots \otimes x_{i_{j+1}} \otimes x_{i_j} \otimes \cdots \otimes x_{i_k} \\ &\qquad + x_{i_1} \otimes \cdots \otimes [x_{i_j}, x_{i_{j+1}}] \otimes \cdots \otimes x_{i_k} \\ &\qquad\qquad + x_{i_1} \otimes \cdots \otimes \big[x_{i_j} \otimes x_{i_{j+1}} - x_{i_{j+1}} \otimes x_{i_j} - [x_{i_j}, x_{i_{j+1}}]\big] \otimes \cdots \otimes x_{i_k} \end{aligned}$$

The first of the three terms on the right hand side has index less than ℓ. The second term on the right hand side can be written as a linear combination of monomials of length $k-1$. The third term on the right hand side belong to the ideal $\mathcal{I}$. Applying the mapping π and using the induction hypotheses, we see that the image of $x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+1}}\otimes\cdots\otimes x_{i_k}$ can be written as a linear combination of images of standard monomials. □

THEOREM 3.4 (Poincaré-Birkhoff-Witt). *The images under π of $1\in\mathbb{R}$ and the standard monomials are linearly independent, and hence form a basis for $\mathfrak{U}(\mathfrak{L})$.*

PROOF. Following Jacobson [**Jac62**], we construct a linear mapping $L: T(\mathfrak{L})\to T(\mathfrak{L})$ with the following properties:

(1) $L(1)=1$.
(2) $L(x_{i_1}\otimes\cdots\otimes x_{i_k})=x_{i_1}\otimes\cdots\otimes x_{i_k}$ for every standard monomial $x_{i_1}\otimes\cdots\otimes x_{i_k}$.
(3) If $x_{i_1}\otimes\cdots\otimes x_{i_k}$ is not standard, and if $i_{j+1}<i_j$ then

$$\begin{aligned}L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+1}}\otimes\cdots\otimes x_{i_k})&\\ =L(x_{i_1}\otimes\cdots\otimes x_{i_{j+1}}\otimes x_{i_j}\otimes\cdots\otimes x_{i_k})&\\ +L(x_{i_1}\otimes\cdots\otimes[x_{i_j},x_{i_{j+1}}]\otimes\cdots\otimes\cdots\otimes x_{i_k})&\end{aligned}$$

Suppose we can show that such a linear transformation exists. It follows from (3) that L maps every element of the ideal $\mathcal{I}$ to 0, since every element of the ideal is a linear combination of elements of the form

$$x_{i_1}\otimes\cdots\otimes\left(x_{i_j}\otimes x_{i_{j+1}}-x_{i_{j+1}}\otimes x_{i_j}-\left[x_{i_j},x_{i_{j+1}}\right]\right)\otimes\cdots\otimes x_{i_k}.$$

It follows from (1) and (2) that L is the identity on the subspace $W\subset T(\mathfrak{L})$ spanned by 1 and the standard monomials. Thus it follows that $W\cap\mathcal{I}=(0)$. Hence the projection π is one-to-one on the space W, and hence the images of 1 and the standard monomials are linearly independent, since they are linearly independent in $T(\mathfrak{L})$. Thus to prove the theorem it suffices to construct L.

Since $T(\mathfrak{L})=\mathbb{R}\oplus\mathfrak{L}^1\oplus\cdots$, and since we know that 1 and the standard monomials do span $T(\mathfrak{L})$, it suffices to show that for every positive integer m there is a linear mapping

$$T:\mathbb{R}\oplus\mathfrak{L}^1\oplus\mathfrak{L}^2\oplus\cdots\oplus\mathfrak{L}^m\to\mathfrak{L}^1\oplus\mathfrak{L}^2\oplus\cdots\oplus\mathfrak{L}^m$$

which satisfies conditions (1), (2), and (3). We proceed by induction on m. When $m=1$, we can let T be the identity since each x_i is a standard monomial, and the hypotheses of condition (3) are never satisfied in this case.

Now suppose that we have constructed the mapping L on $\mathbb{R}\oplus\mathfrak{L}^1\oplus\mathfrak{L}^2\oplus\cdots\oplus\mathfrak{L}^{m-1}$ where $m>1$. In order to extend L to $\mathbb{R}\oplus\mathfrak{L}^1\oplus\mathfrak{L}^2\oplus\cdots\oplus\mathfrak{L}^m$, it suffices to define L on each monomial of length m. We proceed by induction on the index of the monomial. Of course, we let L be the identity on any standard monomial. Suppose that we have defined L on all monomials of length m and index less than $\ell>1$, and suppose that $x_{i_1}\otimes\cdots\otimes x_{i_m}$ is a monomial of index ℓ. We can then find an integer j so that $i_{j+1}<i_j$. We then try to define L on $x_{i_1}\otimes\cdots\otimes x_{i_m}$ by setting

$$\begin{aligned}L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+1}}\otimes\cdots\otimes x_{i_k})&\\ =L(x_{i_1}\otimes\cdots\otimes x_{i_{j+1}}\otimes x_{i_j}\otimes\cdots\otimes x_{i_k})&\\ +L(x_{i_1}\otimes\cdots\otimes[x_{i_j},x_{i_{j+1}}]\otimes\cdots\otimes\cdots\otimes x_{i_k})&\end{aligned}\tag{3.1}$$

The two terms on the right hand side are defined since the first monomial has index less than ℓ, and the product in the second term is a linear combination of monomials of length $m-1$. Thus if the tentative definition given in (3.1) is unambiguous, the extended L does satisfy conditions (1), (2), and (3).

The difficulty with (3.1) is that there may be two integers j and l so that $i_{j+1} < i_j$ and $i_{l+1} < i_l$. Without loss of generality, we may assume the $j < l$.

There are two cases to consider. The easier possibility is that $j+1 < l$. We then need to see whether

$$\begin{aligned}
&L(x_{i_1}\otimes\cdots\otimes x_{i_{j+1}}\otimes x_{i_j}\otimes\cdots\otimes x_{i_l}\otimes x_{i_{l+1}}\otimes\cdots\otimes x_{i_k})\\
&\qquad + L(x_{i_1}\otimes\cdots\otimes[x_{i_j},x_{i_{j+1}}]\otimes\cdots\otimes x_{i_l}\otimes x_{i_{l+1}}\otimes\cdots\otimes x_{i_k}) =\\
&L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+1}}\otimes\cdots\otimes x_{i_{l+1}}\otimes x_{i_l}\otimes\cdots\otimes x_{i_k})\\
&\qquad + L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+1}}\otimes\cdots\otimes[x_{i_l},x_{i_{l+1}}]\otimes\cdots\otimes x_{i_k}).
\end{aligned}$$

Since we are dealing with elements of smaller length or index, we can use condition (3) to evaluate each side. Let us write $x_{i_j}=a$, $x_{i_{j+1}}=b$, $x_{i_l}=c$, and $x_{i_{l+1}}=d$. Then we have

$$\begin{aligned}
&L(\cdots\otimes b\otimes a\otimes\cdots\otimes c\otimes d\otimes\cdots) + L(\cdots\otimes[a,b]\otimes\cdots\otimes c\otimes d\otimes\cdots)\\
&\quad = L(\cdots\otimes b\otimes a\otimes\cdots\otimes d\otimes c\otimes\cdots)\\
&\qquad + L(\cdots\otimes b\otimes a\otimes\cdots\otimes[c,d]\otimes\cdots)\\
&\qquad\quad L(\cdots\otimes[a,b]\otimes\cdots\otimes d\otimes c\otimes\cdots)\\
&\qquad\qquad L(\cdots\otimes[a,b]\otimes\cdots\otimes[c,d]\otimes\cdots)\\
&\quad = L(\cdots\otimes b\otimes a\otimes\cdots\otimes d\otimes c\otimes\cdots)\\
&\qquad + L(\cdots\otimes[a,b]\otimes\cdots\otimes d\otimes c\otimes\cdots)\\
&\qquad\quad + L(\cdots\otimes b\otimes a\otimes\cdots\otimes[c,d]\otimes\cdots)\\
&\qquad\qquad + L(\cdots\otimes[a,b]\otimes\cdots\otimes[c,d]\otimes\cdots)\\
&= L(\cdots\otimes a\otimes b\otimes\cdots\otimes d\otimes c\otimes\cdots) + L(\cdots\otimes a\otimes b\otimes\cdots\otimes[c,d]\otimes\cdots).
\end{aligned}$$

The more complicated case occurs when $j+1=l$. In this case we need to check that

$$\begin{aligned}
&L(x_{i_1}\otimes\cdots\otimes x_{i_{j+1}}\otimes x_{i_j}\otimes x_{i_{j+2}}\otimes\cdots\otimes x_{i_k})\\
&\qquad + L(x_{i_1}\otimes\cdots\otimes[x_{i_j},x_{i_{j+1}}]\otimes x_{i_{j+2}}\otimes\cdots\otimes x_{i_k}) =\\
&L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes x_{i_{j+2}}\otimes x_{i_{j+1}}\otimes\cdots\otimes x_{i_k})\\
&\qquad + L(x_{i_1}\otimes\cdots\otimes x_{i_j}\otimes[x_{i_{j+1}},x_{i_{j+2}}]\otimes\cdots\otimes x_{i_k}).
\end{aligned}$$

Let us write $x_{i_j}=a$, $x_{i_{j+1}}=b$, $x_{i_{j+2}}=c$. Then

$$\begin{aligned}
&L(\cdots\otimes b\otimes a\otimes c\otimes\cdots) + L(\cdots\otimes[a,b]\otimes c\otimes\cdots)\\
&\quad = L(\cdots\otimes b\otimes c\otimes a\otimes\cdots)\\
&\qquad + L(\cdots\otimes b\otimes[a,c]\otimes\cdots)\\
&\qquad\quad + L(\cdots\otimes[a,b]\otimes c\otimes\cdots)\\
&\quad = L(\cdots\otimes c\otimes b\otimes a\otimes\cdots)\\
&\qquad + L(\cdots\otimes[b,c]\otimes a\otimes\cdots)\\
&\qquad\quad + L(\cdots\otimes b\otimes[a,c]\otimes\cdots)
\end{aligned}$$

$$+ L(\cdots \otimes [a,b] \otimes c \otimes \cdots).$$

On the other hand

$$\begin{aligned}
L(\cdots \otimes a \otimes c \otimes b \otimes \cdots) + L(\cdots \otimes a \otimes [b,c] \otimes \cdots) \\
&= L(\cdots \otimes c \otimes a \otimes b \otimes \cdots) \\
&\quad + L(\cdots \otimes [a,c] \otimes b \otimes \cdots) \\
&\qquad + L(\cdots \otimes a \otimes [b,c] \otimes \cdots) \\
&= L(\cdots \otimes c \otimes b \otimes a \otimes \cdots) \\
&\quad + L(\cdots \otimes c \otimes [a,b] \otimes \cdots) \\
&\qquad + L(\cdots \otimes [a,c] \otimes b \otimes \cdots) \\
&\qquad\quad + L(\cdots \otimes a \otimes [b,c] \otimes \cdots).
\end{aligned}$$

Thus we need to show that

$$\begin{aligned}
&L(\cdots \otimes [b,c] \otimes a \otimes \cdots) + L(\cdots \otimes b \otimes [a,c] \otimes \cdots) + L(\cdots \otimes [a,b] \otimes c \otimes \cdots) \\
&\quad = L(\cdots \otimes c \otimes [a,b] \otimes \cdots) + L(\cdots \otimes [a,c] \otimes b \otimes \cdots) + L(\cdots \otimes a \otimes [b,c] \otimes \cdots).
\end{aligned}$$

But

$$\begin{aligned}
L(\cdots \otimes [b,c] \otimes a \otimes \cdots) &= L(\cdots \otimes a \otimes [b,c] \otimes \cdots) - L(\cdots \otimes [a,[b,c]] \otimes \cdots); \\
L(\cdots \otimes b \otimes [a,c] \otimes \cdots) &= L(\cdots \otimes [a,c] \otimes b \otimes \cdots) - L(\cdots \otimes [b,[c,a]] \otimes \cdots); \\
L(\cdots \otimes [a,b] \otimes c \otimes \cdots) &= L(\cdots \otimes c \otimes [a,b] \otimes \cdots) - L(\cdots \otimes [c,[a,b]] \otimes \cdots).
\end{aligned}$$

Thus the identity we seek follows from the Jacobi identity

$$[a,[b,c]] + [b,[c,a]] + [c,[a,b]] = 0,$$

and this completes the proof. □

4. Free Lie Algebras

Having constructed the universal enveloping algebra of a Lie algebra, we now turn to the problem of construction free Lie algebras.

4.1. The construction.

DEFINITION 4.1. *Let $S = \{x_i\}_{i \in I}$ be a set of elements. A Lie algebra $\mathfrak{F}(S)$ is the free Lie algebra generated by S if*

(a) S can be identified with a subset of $\mathfrak{F}(S)$;
(b) If $\mathfrak{L}$ is any Lie algebra and if $\varphi : S \to \mathfrak{L}$ is any mapping, then there is a unique Lie algebra homomorphism $\Phi : \mathfrak{F}(S) \to \mathfrak{L}$ such that $\Phi(x_i) = \varphi(x_i)$ for all $i \in I$.

Equivalently, we can consider the vector space V spanned by the elements of S. $\mathfrak{F}(V)$ is the free Lie algebra generated by V if

(a′) $V \subset \mathfrak{F}(V)$ is a subspace;
(b′) If $\mathfrak{L}$ is any Lie algebra and if $\varphi : V \to \mathfrak{L}$ is any linear map, there exists a unique Lie algebra homomorphism $\Phi : \mathfrak{F}(V) \to \mathfrak{L}$ so that $\Phi(v) = \varphi(v)$ for all $v \in V$.

We can construct the free Lie algebra generated by V as follows. Let $T(V) = \mathbb{R} \oplus V^1 \oplus V^2 \oplus \cdots$ be the tensor algebra generated by V. $T(V)$ is an associative algebra which can therefore be given a Lie algebra structure $T(V)_L$ by using the bracket product $[x, y] = xy - yx$. We then let $\mathfrak{F}(V)$ denote the smallest Lie subalgebra of $T(V)_L$ which contains V. We clearly have $V \subset \mathfrak{F}(V)$. It remains to check the universal mapping property.

LEMMA 4.2. *Let $\mathfrak{L}$ be a Lie algebra, and let $\varphi : V \to \mathfrak{L}$ be a linear map. Then there exists a unique Lie algebra homomorphism $\Phi : \mathfrak{F}(V) \to \mathfrak{L}$ so that $\Phi(v) = \varphi(v)$ for every $v \in V$.*

PROOF. Let $\mathfrak{U}(\mathfrak{L})$ be the universal enveloping algebra of $\mathfrak{L}$. By the Poincaré-Birkhoff-Witt theorem, (Theorem 3.4), we can regard $\mathfrak{L}$ as a subspace of $\mathfrak{U}(\mathfrak{L})$, and so we can think of φ as a linear mapping from V to $\mathfrak{U}(\mathfrak{L})$, an associative algebra. Since $T(V)$ is the free associative algebra generated by V, there is a unique algebra homomorphism $\Phi : T(V) \to \mathfrak{U}(\mathfrak{L})$ such that $\Phi(v) = \varphi(v)$ for $v \in V$. It follows that $\Phi : T(V)_L \to \mathfrak{U}(\mathfrak{L})_L$ is also Lie algebra homomorphism of the associated Lie algebras $T(V)_L$ and $\mathfrak{U}(\mathfrak{L})_L$. In particular, we can restrict Φ to $\mathfrak{F}(V)$, the Lie subalgebra of $T(V)_L$ generated by V.

Thus we have constructed a Lie algebra homomorphism $\Phi : \mathfrak{F}(V) \to \mathfrak{U}(\mathfrak{L})_L$ such that $\Phi(v) = \varphi(v)$ for all $v \in V$. The algebra $\mathfrak{L}$ is a subspace of $\mathfrak{U}(\mathfrak{L})$, and so we can study $W = \{w \in \mathfrak{F}(V) \,|\, \Phi(w) \in \mathfrak{L}\}$. Since $\mathfrak{L}$ is a Lie algebra, and since Φ is a Lie algebra homomorphism, it follows that W is also a Lie algebra. But $V \subset W$, so W is a Lie subalgebra of $T(V)_L$ containing V. Since $\mathfrak{F}(V)$ is the algebra generated by V it follows that $\mathfrak{F}(V) \subset W$. This shows that $\Phi : \mathfrak{F}(V) \to \mathfrak{L}$ is a Lie algebra homomorphism.

If $\widetilde{\Phi} : \mathfrak{F}(V) \to \mathfrak{L}$ is another Lie algebra homomorphism such that $\widetilde{\Phi}(v) = \varphi(v)$ for all $v \in V$, the set $W = \{w \in \mathfrak{F}(V) \,|\, \widetilde{\Phi}(w) = \Phi(w)\}$ is a Lie subalgebra containing V, and hence is all of $\mathfrak{F}(V)$. Thus the mapping Φ is unique, and this completes the proof. □

DEFINITION 4.3. *An element $w \in T(V)$ is a Lie element if $w \in \mathfrak{F}(V)$.*

An element of the algebra $T(V)$ is homogeneous of degree m if it is a sum of monomials of degree m. Equivalently, an element is homogeneous of degree m if it belongs to V^m. Clearly, every element $w \in T(V)$ can be written uniquely as a linear combination of homogeneous elements. We call these the homogeneous components of w.

PROPOSITION 4.4. *An element $w \in T(V)$ is a Lie element if and only if its homogeneous components are Lie elements.*

PROOF. Every iterated commutator $[x_{i_1}, [x_{i_1}, \cdots, [x_{i_{m-1}}, x_{i_m}] \cdots]]$ is homogeneous of degree m. Since every element of $\mathfrak{F}(S)$ is a linear combination of iterated commutators, it follows that every element of the free Lie algebra $\mathfrak{F}(S)$ is a sum of homogeneous elements which are Lie elements. The proof follows from the uniqueness of the decomposition. □

Thus we see that the free Lie algebra $\mathfrak{F}(V)$ can be written

$$\mathfrak{F}(V) = \mathfrak{F}(V)_1 \oplus \mathfrak{F}(V)_2 \oplus \cdots \oplus \mathfrak{F}(V)_m \oplus \cdots \tag{4.1}$$

where $\mathfrak{F}(V)_m$ is the subspace of Lie elements which are homogeneous of degree m. Since every element of V is certainly a Lie element, we have $V = \mathfrak{F}(V)_1$.

LEMMA 4.5. *Let V be a vector space, and let $\mathfrak{F}(V) \subset T(V)$ be the free Lie algebra generated by V and the tensor algebra over V. Then $T(V)$ is the universal enveloping algebra of $\mathfrak{F}(V)$.*

PROOF. Let A be an associative algebra, and let $\psi : \mathfrak{F}(V) \to A_L$ be a Lie algebra homomorphism. Then ψ restricts to a linear map of V to A. Since $T(V)$ is the free associative algebra generated by V, there is a unique associative algebra homomorphism $\Psi : T(V) \to A$ such that $\Psi(v) = \psi(v)$ for all $v \in V$. Thus Ψ is a Lie algebra homomorphism of $T(V)_L$ to A_L.

Let $W = \{w \in \mathfrak{F}(V) \mid \Psi(w) = \psi(w)\}$. Since Ψ and ψ are Lie algebra homomorphisms, it follows that W is a Lie algebra containing V. Since $\mathfrak{L}(V)$ is generated by V, it follows that $W = \mathfrak{L}(V)$. Thus Ψ is an algebra homomorphism from $T(V)$ to A_L and $\Psi(w) = \psi(w)$ for all $w \in \mathfrak{F}(V)$, and we have noted that it is the unique mapping with these properties. This completes the proof. □

COROLLARY 4.6. *Let $\{w_k\}_{k\geq 1}$ be a basis for the real vector space $\mathfrak{F}(V)$. Then the elements $\{w_1^{m_1} w_2^{m_2} \cdots w_k^{m_k}\}$ with $k \geq 0$ and $m_j \geq 0$ form a basis for $T(V)$.*

PROOF. We can give the basis for $\mathfrak{F}(V)$ a total order by declaring that $w_j < w_k$ if and only if $j < k$. It follows from the Poincaré-Birkhoff-Witt theorem that the standard monomials in the $\{w_j\}$ form a basis for the universal enveloping algebra of $\mathfrak{F}(V)$. The corollary thus follows from Lemma 4.5. □

In addition to the algebra $T(V)$ which can be identified with the space of all non-commuting polynomials in the elements of S, we have also defined the algebra $\widetilde{T}(V)$ which is the space of all formal series in the elements of S. Then $\widetilde{T}(V)_L$ is a Lie algebra containing $T(V)_L$, which in turn contains $\mathfrak{F}(V)$.

Now define $\widetilde{\mathfrak{F}}(V)$ as follows.

$$\widetilde{\mathfrak{F}}(V) = \Big\{ P(x) = \sum_{m=0}^{\infty} P_m(x) \in \widetilde{T}(V) \,\Big|\, P_m(x) \in \mathfrak{F}(V)_m \Big\}. \tag{4.2}$$

4.2. Characterizations of Lie elements.

Let $S = \{x_i\}_{i\in I}$ be a set and let V be the vector space with basis S. Let $\mathfrak{F}(S) = \mathfrak{F}(V)$ be the free Lie algebra generated by S or V. Then $\mathfrak{F}(S)$ can be realized as a Lie subalgebra of $T(V)$, the algebra of all non-commuting polynomials in the $\{x_i\}$. How can one recognize whether an element of $T(V)$ is a Lie element? It is clear that $x_i x_j - x_j x_i$ is a Lie element, but it is perhaps not so clear that the element $x_i x_j x_i - x_j x_i x_j$ (for example) is not a Lie element. The object of this section is to provide characterizations of such elements. In addition to its intrinsic interest, these will be needed later in the proof of the Campbell-Baker-Hausdorff theorem.

Recall from equation (2.3) that ad is a Lie algebra homomorphism of a Lie algebra $\mathfrak{L}$ into the Lie algebra of linear mapping from $\mathfrak{L}$ to itself. In particular if $\mathfrak{L} = \mathfrak{F}(S)$ is the free Lie algebra generated by S, the mapping ad extends to an associative algebra mapping θ of $T(S)$ to the linear mappings of $\mathfrak{F}(S)$ to itself.

Now let S be a set, and define a linear map $\sigma : T(S) \to \mathfrak{F}(S)$ by setting $\sigma(x_i) = x_i$ for $i \in I$, and

$$\sigma(x_{i_1} x_{i_2} \cdots x_{i_{m-1}} x_{i_m}) = \operatorname{ad}_{x_{i_1}} \operatorname{ad}_{x_{i_2}} \cdots \operatorname{ad}_{x_{i_{m-1}}}(x_{i_m}).$$

This is well-defined since the monomials form a basis for $T(S)$, and the image clearly belongs to the Lie subalgebra of $T(S)_L$ generated by S. Moreover, we have

$$\begin{aligned}\sigma(x_{i_1}\cdots x_{i_l}x_{i_{l+1}}\cdots x_{i_m}) &= \mathrm{ad}_{x_{i_1}}\cdots \mathrm{ad}_{x_{i_l}}\mathrm{ad}_{x_{i_{l+1}}}\cdots \mathrm{ad}_{x_{i_{m-1}}}x_{i_m})\\ &= \mathrm{ad}_{x_{i_1}}\cdots \mathrm{ad}_{x_{i_l}}\big(\sigma(x_{i_{l+1}}\cdots x_{i_m})\big)\\ &= \theta\big(x_{i_1}\cdots x_{i_l}\big)\big(\sigma(x_{i_{l+1}}\cdots x_{i_m})\big)\end{aligned}$$

Let $a, b \in T(S)$. It follows that

$$\sigma(ab) = \theta(a)\,\sigma(b).$$

Then if $a, b \in \mathfrak{F}(S)$

$$\begin{aligned}\sigma([a,b]) &= \sigma(ab - ba)\\ &= \theta(a)\sigma(b) - \theta(b)\sigma(a)\\ &= \mathrm{ad}_a\big(\sigma(b)\big) - \mathrm{ad}_b\big(\sigma(a)\big)\\ &= [a, \sigma(b)] + [\sigma(a), b].\end{aligned}$$

THEOREM 4.7 (Dynkin-Specht-Wever). *Let $y \in T(S)$ be homogeneous of degree m. Then $y \in \mathfrak{F}(S)$ if and only if $\sigma(y) = m\,y$.*

PROOF. If $\sigma(y) = m\,y$, then $y = m^{-1}\,\sigma(y) \in \mathfrak{F}(S)$. To prove the converse, we argue by induction on m. Every element of $\mathfrak{F}(S)$ is a linear combination of iterated commutators, and it suffices to see that $\sigma(y) = my$ if $y = [x_{i_1}, [x_{i_2}, \cdots [x_{i_{m-1}}, x_{i_m}]\cdots]]$. Let $z = [x_{i_2}, \cdots [x_{i_{m-1}}, x_{i_m}]\cdots]$, so that z is an iterated commutator of length $m-1$. By the last identity, we have

$$\begin{aligned}\sigma(y) &= \sigma\big([x_{i_1}, [x_{i_2}, \cdots [x_{i_{m-1}}, x_{i_m}]\cdots]]\big) = \sigma\big([x_{i_1}, z]\big)\\ &= \big[\sigma(x_{i_1}), z\big] + \big[x_{i_1}, \sigma(z)\big]\\ &= \big[x_{i_1}, z\big] + \big[x_{i_1}, (m-1)z\big] = m\big[x, z\big]\\ &= my\end{aligned}$$

as asserted. □

Let $S = \{x_i\}_{i\in I}$ be a set and let $A(S) = T(S)$ be the free algebra generated by S. We regard elements of $T(S)$ as non-commuting polynomials in the variables $\{x_i\}$. Then $T(S)\otimes T(S)$ is also an associative algebra. An element of $w \in T(S)\otimes T(S)$ can be written as $w = \sum_{j=1}^{m} P_j(x)\otimes Q_j(x)$ where P_j and Q_j are non-commuting polynomials in the elements of S. The product of two elements is given by

$$\Big(\sum_{j=1}^{m} P_j(x)\otimes Q_j(x)\Big)\Big(\sum_{l=1}^{n} R_l(x)\otimes S_l(x)\Big) = \sum_{j=1}^{m}\sum_{l=1}^{n}\big(P_j(x)R_l(x)\otimes Q_j(x)S_l(x)\big).$$

In the same way, we can define an algebra structure on $\widetilde{T}(S)\otimes\widetilde{T}(S)$.

Define $\delta : S \to T(S)\times T(S)$ by setting $\delta(x_i) = (x_i, 1) + (1, x_i)$ for $i \in I$. By the universal property of free algebras, there is a unique algebra homomorphism $D : T(S) \to T(S)\times T(S)$ such that $D(x_i) = \delta(x_i)$. It is clear that the mapping D extends to a mapping $D : \widetilde{T}(S) \to \widetilde{T}(S)\times\widetilde{T}(S)$.

Note that if $w \in T(S)$, the elements $w\otimes 1$ and $1\otimes w$ commute in $T(S)\otimes T(S)$, and hence

$$(w\otimes 1 + 1\otimes w)^m = \sum_{j=0}^{m}\frac{m!}{j!(n-j)!}w^j\otimes w^{n-j}.$$

THEOREM 4.8 (Friedrichs). *An element $w \in T(S)$ is a Lie element if and only if $D(w) = w \otimes 1 + 1 \otimes w$.*

PROOF. We first observe that the set $\{a \in T(S) \mid D(a) = (a \otimes 1) + (1 \otimes a)\}$ is a Lie subalgebra of $T(S)_L$ containing S, and hence containing $\mathfrak{F}(S)$. In fact, if $D(a) = (a \otimes 1) + (1 \otimes a)$ and $D(b) = (b \otimes 1) + (1 \otimes b)$, then

$$\begin{aligned} D\big([a,b]\big) &= D(ab - ba) \\ &= D(a)D(b) - D(b)D(a) \\ &= \big((a \otimes 1) + (1 \otimes a)\big)\big((b \otimes 1) + (1 \times b)\big) \\ &\qquad - \big((b \otimes 1) + (1 \otimes b)\big)\big((a \otimes 1) + (1 \otimes a)\big) \\ &= (ab \otimes 1) + (a \otimes b) + (b \otimes a) + (1 \otimes ab) - (ba \otimes 1) - (b \otimes a) \\ &\qquad - (a \otimes b) - (1 \otimes ba) \\ &= \big([a,b] \otimes 1\big) + \big(1 \otimes [a,b]\big). \end{aligned}$$

Let $\{y_1, y_2, \ldots\}$ be a basis for $\mathfrak{F}(S)$. We have just seen that for each j, $D(y_j) = (y_j \otimes 1 + 1 \otimes y_j)$. We order this basis so that $y_j < y_l$ if and only if $j < l$. The tensor algebra $T(S)$ is the universal enveloping algebra of $\mathfrak{F}(S)$, and by the Poincaré-Birkhoff-Witt theorem, the monomials $\{y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\}$ for $k \geq 0$ and $m_j \geq 0$ are a basis for $T(S)$ as a real vector space. Hence the products

$$\big\{y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k} \otimes y_1^{n_1} y_2^{n_2} \cdots y_l^{n_l}\big\}$$

are a basis for $T(S) \otimes T(S)$. Since D is an algebra homomorphism,

$$\begin{aligned} &D\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) \\ &\quad = D(y_1)^{m_1} D(y_2^{m_2}) \cdots D(y_k)^{m_k} \\ &\quad = \big(y_1 \otimes 1 + 1 \otimes y_1\big)^{m_1} \big(y_2 \otimes 1 + 1 \otimes y_2\big)^{m_2} \cdots \big(y_k \otimes 1 + 1 \otimes y_k\big)^{m_k} \\ &\quad = y_1^{m_1} y_2^{k_2} \cdots y_k^{m_k} \otimes 1 + 1 \otimes y_1^{m_1} y_2^{k_2} \cdots y_k^{m_k} + E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) \end{aligned}$$

where

$$E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) = \sum_{\substack{r_j + s_j = m_j \\ \sum_{j=1}^k r_j > 1 \\ \sum_{j=1}^k s_j > 1}} \Big[\prod_{j=1}^{k} \frac{m_j!}{r_j! s_j!}\Big] y_1^{r_1} y_2^{r_2} \cdots y_k^{r_k} \otimes y_1^{s_1} y_2^{s_2} \cdots y_k^{s_k}.$$

Note that

(1) If $k > 1$ then $E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) \neq 0$.
(2) If $k = 1$ and $m_1 > 1$ then $E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) \neq 0$.
(3) The elements $\big\{E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big)\big\}$ for which $k > 1$ or $k = 1$ and $m_1 > 1$ are linearly independent since the involve different elements of the basis of $T(S) \otimes T(S)$.

Now suppose $w = \sum \alpha_{m_1,\ldots,m_k} y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k} \in T(S)$ and $D(w) = w \otimes 1 + 1 \otimes w$. It follows that $\sum_{k>1} \alpha_{m_1,\ldots,m_k} E\big(y_1^{m_1} y_2^{m_2} \cdots y_k^{m_k}\big) = 0$, and hence $\alpha_{m_1,\ldots,m_k} = 0$ whenever $k > 1$ or $k = 1$ and $m_1 > 1$. Thus $w = \sum_j \alpha_j y_j$ is a linear combination of basis elements of $\mathfrak{F}(S)$, and hence w is a Lie element, as required. □

Since the property of being a Lie element depends only on whether or not each homogeneous component is a Lie element, we have

COROLLARY 4.9. *If $w \in \widetilde{T}(S)$ and if $D(w) = w \otimes 1 + 1 \otimes w$, then each homogeneous component of w is a Lie element.*

5. The Campbell-Baker-Hausdorff formula

We observed in Proposition 1.4 that if x and y are indeterminates which commute, then

$$\begin{gathered} \exp(x)\,\exp(y) = \exp(x+y), \\ \log\big((1+x)(1+y)\big) = \log(1+x) + \log(1+y). \end{gathered} \tag{5.1}$$

If x and y do not commute, we look for a formal power series

$$z(x,y) = x + y + P_2(x,y) + P_3(x,y) + \cdots \tag{5.2}$$

where $P_j(x,y)$ is a non-commuting polynomial in x and y such that

$$\exp(x)\,\exp(y) = \exp\big(z(x,y)\big). \tag{5.3}$$

The series for $z(x,y)$ is given by the Campbell-Baker-Hausdorff formula. It is a remarkable fact that that the polynomials $P_j(x,y)$ are actually Lie elements; that is, they are linear combinations of iteraged commutators of x and y.

THEOREM 5.1 (Campbell-Baker-Hausdorff). *As formal series we have*

$$\exp(x)\,\exp(y) = \exp\big(z(x,y)\big)$$

where

$$z(x,y) = x + y + \sum_{n=2}^{\infty} P_N(x,y)$$

where each P_N is a homogeneous non-commutative polynomial in x and y of degree N which can also be written as the Lie element

$$P_N(x,y) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{k}\sum_{j=1}^{k}\sum_{\substack{m_j+n_j\ge 1\\ \sum(m_j+n_j)=N}} \frac{(-1)^{k+1}}{m_1!\cdots m_k! n_1!\cdots n_k!}\; ad_x^{m_1} ad_y^{n_1}\cdots ad_x^{m_k} ad_y^{n_k}.$$

Here $ad_x^{m_1} ad_y^{n_1}\cdots ad_x^{m_k} ad_y^{n_k}$ means

$$ad_x^{m_1} ad_y^{n_1}\cdots ad_x^{m_k} ad_y^{n_k} = \begin{cases} ad_x^{m_1} ad_y^{n_1}\cdots ad_x^{m_k} ad_y^{n_k-1}[y] & \text{if } n_k \ge 1, \\ ad_x^{m_1} ad_y^{n_1}\cdots ad_y^{m_{k-1}} ad_x^{m_k-1}[x] & \text{if } n_k = 0. \end{cases}$$

PROOF. We work in the free associative algebra and free Lie algebra generated by a set S consisting of two elements x and y. We begin with an elementary computation that gives

$$z(x,y) = \log\big(\exp(x)\,\exp(y)\big)$$

as a formal series of non-commuting polynomials in x and y. We have

$$\exp(x)\,\exp(y) = \sum_{m,m=0}^{\infty}\frac{1}{m!\,n!}x^m y^n$$

and so

$$\exp(x)\,\exp(y) - 1 = \sum_{\substack{m,n\ge 0\\ m+n\ge 1}}\frac{1}{m!\,n!}x^m y^n.$$

Hence, as a formal series

$$\begin{aligned} z(x,y) &= \log\big(\exp(x)\,\exp(y)\big) \\ &= \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\Bigg[\sum_{m+n\geq 1}\frac{1}{m!\,n!}x^m y^n\Bigg]^k \\ &= \sum_{k=1}^{\infty}\sum_{\substack{1\leq j\leq k\\ m_j+n_j\geq 1}}\frac{(-1)^{k+1}}{k\,m_1!\cdots m_k!n_1!\cdots n_k!}x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}. \end{aligned}$$

Note that the component of this sum which is homogeneous of degree N is

$$P_N(x,y) = \sum_{k=1}^{N}\sum_{j=1}^{k}\sum_{\substack{m_j+n_j\geq 1\\ \sum(m_j+n_j)=N}}\frac{(-1)^{k+1}}{k\,m_1!\cdots m_k!n_1!\cdots n_k!}x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}.$$

If we can prove that the homogeneous polynomial $P_N(x,y)$ is a Lie element in $T(S)$, we can then use Theorem 4.7 write it explicitly as a Lie element, and this is the statement of the Theorem.

To prove that the homogeneous components of $\log\big(\exp(x)\,\exp(y)\big)$ are Lie elements, we use the criterion given in Theorem 4.8 and Corollary 4.9. Note that since the elements $(x\otimes 1)$ and $(1\otimes x)$ commute, we have

$$\begin{aligned} \exp\big(x\otimes 1 + 1\otimes x\big) &= \exp(x\otimes 1)\,\exp(1\otimes x) \\ &= \big(\exp(x)\otimes 1\big)\,\big(1\otimes\exp(x)\big). \end{aligned}$$

We have

$$\begin{aligned} D\big(\exp(x)\,\exp(y)\big) &= \exp\big(D(x)\big)\,\exp\big(D(y)\big) \\ &= \exp\big(x\otimes 1 + 1\otimes x\big)\,\exp\big(y\otimes 1 + 1\otimes y\big) \\ &= \exp\big(x\otimes 1\big)\,\exp\big(1\otimes x\big)\,\exp\big(y\otimes 1\big)\,\exp\big(1\otimes y\big) \\ &= \big(\exp(x)\otimes 1\big)\,\big(1\otimes\exp(x)\big)\,\big(\exp(y)\otimes 1\big)\,\big(1\otimes\exp(y)\big) \\ &= \big(\exp(x)\,\exp(y)\otimes 1\big)\big(1\otimes\exp(x)\,\exp(y)\big) \end{aligned}$$

Thus

$$\begin{aligned} D\big(\log\big(\exp(x)\,\exp(y)\big)\big) &= \log\big(D\big(\exp(x)\,\exp(y)\big)\big) \\ &= \log\big(\big(\exp(x)\,\exp(y)\otimes 1\big)\big(1\otimes\exp(x)\,\exp(y)\big)\big) \\ &= \log\big(\exp(x)\,\exp(y)\otimes 1\big) + \log\big(1\otimes\exp(x)\,\exp(y)\big)\big) \\ &= \log\big(\exp(x)\,\exp(y)\big)\otimes 1 + 1\otimes\log\big(\exp(x)\,\exp(y)\big). \end{aligned}$$

Thus $\log\big(\exp(x)\,\exp(y)\big)$ is a Lie element in $\widetilde{T}(S)$, and this completes the proof. □

6. Hall Bases for Free Lie Algebras

We now turn to the problem of giving an explicit description of a basis for a free Lie algebra. Let $S=\{x_1,\ldots,x_n\}$ be a finite set, let $\mathfrak{F}(S)$ be the free Lie algebra generated by S, and let $\mathfrak{F}(S)_m$ be the elements in $\mathfrak{F}(S)$ which are homogeneous of degree m.

6.1. Hall Sets.

We define an ordered set H of homogeneous monomials in $\mathfrak{F}(S)$, called Hall monomials[1]. The definition of H is given recursively. Thus let $\widehat{H}_k$ denote the elements of H which are homogeneous of exactly degree k, and let $H_k = \widehat{H}_1 \cup \cdots \cup \widehat{H}_k$ be the subset of elements which are homogeneous of degree at most k. We define $\widehat{H}_n$ inductively as follows.

(1) The subset $H_1 = \widehat{H}_1$ are exactly the generators $\{x_1, \ldots, x_n\}$.
(2) We give the set H_1 a total order (for example by requiring that $x_j < x_k$ if and only if $j < k$).
(3) Suppose that we have constructed the elements of H_{n-1} which are homogeneous of degree less than n where $n > 1$. Suppose also that the elements are given a total order in such a way that if $u \in \widehat{H}_j$ and $v \in \widehat{H}_k$ with $j < k$, then $u < v$.
(4) The elements of $\widehat{H}_n$ is the set of all elements of the form $y = [u, v]$ (called a standard form of y) where
    (a) $u \in \widehat{H}_r$, $v \in \widehat{H}_s$, and $r + s = n$.
    (b) $u < v$.
    (c) If $v = [a, b]$ is a standard form of v, then $a \leq u$.
(5) The order on H_{n-1} is extended to H_n in such a way that if $u \in H_{n-1}$ and $v \in \widehat{H}_n$, then $u < v$.

A set $H = \bigcup_{n=1}^{\infty} \widehat{H}_n$ is called a *Hall set* in $\mathfrak{F}(S)$.

THEOREM 6.1 (Hall). *If $H \subset \mathfrak{F}(S)$ is a Hall set, then the elements of H form a basis for $\mathfrak{F}(S)$.*

6.2. Two Examples.

Before giving the proof of Theorem 6.1, we calculate some low order terms in Hall set for the free Lie algebra generated by two elements x and y, and for the free Lie algebra generated by three elements x, y, z, We order these generators by requiring $x < y < z$.

Example 1 In the free Lie algebra $\mathfrak{L}(x, y)$ we have

$$\widehat{H}_1 = \{x, y\}$$
$$\widehat{H}_2 = \{[x, y]\}$$
$$\widehat{H}_3 = \{[x, [x, y]],\ [y, [x, y]]\}$$
$$\widehat{H}_4 = \{[x, [x, [x, y]]],\ [y, [x, [x, y]]],\ [y, [y, [x, y]]]\}$$

Note that $[y, [x, [x, y]]$ is eliminated bycasue of condition (4c). Also since there is only one element in $\widehat{H}_2$ there are no elements in $\widehat{H}_4$ of the form $[A, B]$ where $A = [u_1, v_1]$ and $B = [u_1, v_1]$. The set $\widehat{H}_5$ consists of six elements. The first four come from taking brackets of elements of $\widehat{H}_1$ and elements of $\widehat{H}_4$

$$\widehat{H}_{5a} = \{[x, [x, [x, [x, y]]]],\ [y, [x, [x, [x, y]]]],\ [y, [y, [x, [x, y]]]],\ [y, [y, [y, [x, y]]]]\},$$

[1]In his paper [**Hal50**], Marshall Hall calls these *standard* monomials, but we have already used this term in the proof of the Poincaré-Birkhoff-Witt theorem.

The next two come from taking brackets of elements of $\widehat{H}_2$ with elements of $\widehat{H}_3$

$$\widehat{H}_{5b} = \big\{ \big[[x,y],[x,[x,y]]\big],\ \big[[x,y],[y,[x,y]]\big] \big\}.$$

Example 2 In the free Lie algebra $\mathfrak{L}(x,y,z)$ we have

$$\begin{aligned}
\widehat{H}_1 &= \{x,y,z\}\\
\widehat{H}_2 &= \{[x,y],[x,z],[y,z]\}\\
\widehat{H}_3 &= \{[x,[x,y]],\ [x,[x,z]],\ [y,[x,y]],\ [y,[x,z]],\ [y,[y,z]],\ [z,[y,z]]\}\\
\widehat{H}_4 &= \{[x,[x,[x,y]]],\ [x,[x,[x,z]]],\ [x,[y,[x,y]]],\ [x,[y,[x,z]]],\ [x,[y,[y,z]]],\\
&\qquad [x,[z,[y,z]]],\ [y,[y,[x,y]]],\ [y,[y,[x,z]]],\ [y,[y,[y,z]]],\ [z,[z,[y,z]]]\\
&\qquad \big[[x,y],[x,z]\big],\ \big[[x,y],[y,z]\big],\ \big[[x,z],[y,z]\big]\}
\end{aligned}$$

6.3. Proof of Theorem 6.1.

Let $S = \{x_1,\ldots,x_n\}$, let V be the real vector space spanned by S, and let $H = \bigcup_{n=1}^{\infty} \widehat{H}_n$ be a Hall set for the free Lie algebra $\mathfrak{F}(S)$. Let $w \in \widehat{H}_n$. We say that an equation $w = \sum_k \alpha_k h_k$ gives a *Hall expansion* for the element w if $\alpha_k \in \mathbb{R}$, and each $h_k \in H$. If we can prove that every element $w \in \mathfrak{F}(S)$ has has a Hall expansion, and that such an expansion is unique, then we will have proved Theorem 6.1.

LEMMA 6.2. *Every element $w \in \mathfrak{F}(S)$ which is homogeneous of degree n has a Hall expansion $w = \sum_k \alpha_k h_k$ where each $h_k \in \widehat{H}_n$.*

PROOF. We argue by induction on n. Every element of $\mathfrak{F}(S)$ which is homogeneous of degree 1 belongs to V, and since $\widehat{H}_1$ consists of a basis for V, this finishes the case $n = 1$.

Now suppose that the assertion of the Lemma is true for every element of $\mathfrak{F}(S)$ which is homogeneous of degree less than n, and let $w \in \mathfrak{F}(S)$ be homogeneous of degree $n > 1$. Then we can write $w = \sum_j \alpha_j [u_j, v_j]$ where $\alpha_j \in \mathbb{R}$, and $u_j, v_j \in \mathfrak{F}(S)$ are elements which are homogeneous of degree strictly less than n. (This representation need not be unique.) We give an algorithm which replaces this expansion for w by an expansion which is a Hall expansion. Moreover, we will see that if we start with a Hall expansion, the algorithm does not change the expansion. The algorithm consists of four steps, which are then repeated, and after a finite number of iterations yields the desired Hall expansion.

Step 1: Since the degree of u_j and v_j are less than n, our induction hypothesis means that we can write $u_j = \sum_k \beta_{j,k}\, u_{j,k}$ and $v_j = \sum_l \gamma_l\, v_{j,l}$ where $\{u_{j,k}\}$ and $\{v_{j,l}\}$ are Hall elements. If either u_j and v_j are Hall elements to begin with, we do not replace them with different sums. Then $w = \sum_{j,k,l} \alpha_j \beta_{j,k} \gamma_{j.l} \big[u_{j,k}, v_{j,l}\big]$.

Step 2: For each term $\big[u_{j,k}, v_{j,l}\big]$, write

$$\big[u_{j,k}, v_{j,l}\big] = \begin{cases} 0 & \text{if } u_{j,k} = v_{j,l},\\ \big[u_{j,k}, v_{j,l}\big] & \text{if } u_{j,k} < v_{j,l},\\ -\big[u_{j,k}, v_{j,l}\big] & \text{if } v_{j,l} < u_{j,k}. \end{cases}$$

With this change, we have written $w = \sum_{j,k,l} \alpha_j \beta_{j,k} \gamma_{j.l} \bigl[u_{j,k}, v_{j,l}\bigr]$ where each of the terms $u_{j,k}$, $v_{j,l} \in H$ and $u_{j,k} < v_{j,l}$. Also if our original expansion was a Hall expansion, we have not changed it.

Step 3: Consider each pair of elements $\bigl\{u_{j,k}, v_{j,l}\bigr\}$.

(1) If the degree of $v_{j,l}$ is 1, then since $u_{j,k} < v_{j,l}$, the degree of $u_{j,k}$ is also 1, and it follows that $\bigl[u_{j,k}, v_{j,l}\bigr]$ is a Hall element. In this case we stop.

(2) If the degree of $v_{j,l}$ is greater than 1, we can write $v_{j,l} = [a_{j,l}, b_{j,l}]$ where $a_{j,l}$, $b_{j,l} \in H$ and $a_{j,l} < b_{j,l}$. Write

$$\bigl[u_{j,k}, v_{j,l}\bigr] = \begin{cases} \bigl[u_{j,k}, [a_{j,l}, b_{j,l}]\bigr] & \text{if } a_{j,l} \leq u_{j,k}, \\ \\ +\bigl[a_{j,l}, [u_{j,k}, b_{j,l}]\bigr] - \bigl[b_{j,l}, [u_{j,k}, a_{j,l}]\bigr] & \text{if } u_{j,k} < a_{j,l}. \end{cases}$$

(a) If $a_{j,l} \leq u_{j,k}$, then $\bigl[u_{j,k}, [a_{j,l}, b_{j,l}]\bigr]$ is a Hall element, and we stop.

(b) Suppose that $u_{j,k} < a_{j,l}$. If $[u_{j,l}, b_{j,l}]$ is a Hall element, it follows that $\bigl[a_{j,l}, [u_{j,k}, b_{j,l}]\bigr]$ is a Hall element since $u_{j,k} < a_{j,l}$. If $[u_{j,l}, a_{j,l}]$ is a Hall element, then $\bigl[b_{j,l}, [u_{j,k}, a_{j,l}]\bigr]$ is a Hall element since $u_{j,k} < a_{j,l} < b_{j,l}$. In either of these two cases, we stop.

Step 4: If either $[u_{j,l}, b_{j,l}]$ or $[u_{j,l}, a_{j,l}]$ is not a Hall element, then we return to Step 1 with the expressions $[u_{j,l}, b_{j,l}]$ and $[u_{j,l}, a_{j,l}]$, write them as a linear combination of Hall elements, we repeat the process.

To see that this procedure eventually stops, we argue as follows. If the degree of $u_{j,k}$ and the degree of $v_{j,k}$ are both greater than $\frac{1}{3}n$ then the degree of $v_{j,k}$ is less that $\frac{2}{3}n$. But then when we write $v_{j,l} = [a_{j,l}, b_{j,l}]$, we must have the degree of $a_{j,l}$ less than than or equal to one half of the degree of $v_{j,l}$, so the degree of $a_{j,l}$ is at most $\frac{1}{2}\frac{2}{3}n = \frac{1}{3}n$, which is less than the degree of $u_{j,k}$. It follows that $a_{j,l} < u_{j,k}$, and so in Step 3, we are in situation (2a), and the process stops.

If we do not stop at Step 3, then we go back to Step 1 with $u_{j,k}$ replaced by $a_{j,l}$ or $b_{j,l}$, and $v_{j,l}$ replaced by $[u_{j,l}, b_{j,l}]$ or $[u_{j,l}, a_{j,l}]$. However, since they are products, the degree of $[u_{j,l}, b_{j,l}]$ or $[u_{j,l}, a_{j,l}]$ is strictly greater than the degree of $u_{j,l}$. Also $u_{j,k} < a_{j,l}$. Thus we have replaced $[u_{j,k}, v_{j,l}]$ by a commutator where the first term appears later than $u_{j,k}$ in the total order of the Hall elements, and the degree of the second term is strictly larger than the degree of $u_{j,k}$. If we run through this process enough times, it follows that eventually the degrees of both terms will be greater than $\frac{1}{3}n$, and so the process will stop. This completes the proof. $\square$

CHAPTER 8

The $\overline{\partial}$ and $\overline{\partial}$-Neumann Problems

1. Differential forms and the d and $\overline{\partial}$-operators

Let $\Omega \subset \mathbb{R}^n$ be an open set. Recall that $\mathcal{E}(\Omega)$ denotes the space of infinitely differentiable real-valued differentiable functions on Ω, $T(\Omega)$ denotes the space of infinitely differentiable real vector fields on Ω, and for each $x \in \Omega$, T_x denotes the tangent space at x. We can identify T_x with the set of values of real vector fields at x.

1.1. Real differential forms.

We begin with the definition of differential forms and wedge product.

Definition 1.1.

(1) *For $m \geq 0$, Λ^m_x denotes the space of alternating m-linear mappings of the tangent space T_x to $\mathbb{R}$. In particular, we identify Λ^0_x with $\mathbb{R}$, and Λ^1_x is the dual space T^*_x, which is sometimes called the real cotangent space at x. An element $\omega \in \Lambda^m_x$ is called an m-form at x.*

(2) *If $\alpha \in \Lambda^j_x$ and $\beta \in \Lambda^k_x$, the wedge product $\alpha \wedge \beta \in \Lambda^{j+k}_x$ is defined by the equation*

$$\alpha \wedge \beta\big(v_1, \ldots, v_{j+k}\big) = \frac{1}{(j+k)!} \sum_{\sigma \in \mathfrak{S}_{j+k}} (-1)^\sigma \alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(j)}\big)\, \beta\big(v_{\sigma(j+1)}, \ldots v_{\sigma(j+k)}\big)$$

where $v_1, \ldots, v_{j+k} \in T_x$, $\mathfrak{S}_{j+k}$ is the group of permutations of the set $\{1, \ldots j+k\}$, and $(-1)^\sigma$ is the sign of the permutation σ.[1]

(3) *$\Lambda^m(\Omega)$ is the space of mappings $\omega : \Omega \to \bigcup_{x \in \Omega} \Lambda^m_x$ such that:*
   (a) *$\omega(x) = \omega_x \in \Lambda^m_x$;*
   (b) *if $X_1, \ldots, X_m \in T(\Omega)$, then the mapping $x \to \omega_x\big((X_1)_x, \ldots, (X_m)_x\big)$ is an infinitely differentiable function.*

   In particular, we identify the space $\Lambda^0(\Omega)$ with $\mathcal{E}(\Omega)$. Elements of $\Lambda^m(\Omega)$ are called smooth m-forms on Ω.[2]

(4) *If $\omega \in \Lambda^j(\Omega)$ and $\eta \in \Lambda^k(\Omega)$, set $(\omega \wedge \eta)_x = \omega_x \wedge \eta_x$. In particular, if $f \in \Lambda^0_x$, $(f\omega)_x = f(x)\omega_x \in \Lambda^k(\Omega)$. This makes $\Lambda^k(\Omega)$ into an $\mathcal{E}(\Omega)$-module.*

---

[1] If $\lambda \in \Lambda^0_x = \mathbb{R}$ and $\beta \in \Lambda^k_x$, it is traditional to write $\lambda\beta$ instead of $\lambda \wedge \beta$ since scalar multiplication gives the same value as the wedge product in this case.

[2] We are really considering smooth sections of the bundle Λ^m of alternating m-forms over Ω. However we will not develop the machinery of vector bundles here.

Proposition 1.2. *If $\alpha \in \Lambda_x^j$ and $\beta \in \Lambda_x^k$, then*

$$\alpha \wedge \beta = (-1)^{jk}\, \beta \wedge \alpha.$$

In particular, if $j = k = 1$, $\alpha \wedge \beta = -\beta \wedge \alpha$.

Proof. Define $\tau \in \mathfrak{S}_{j+k}$ by

$$\tau(1, \ldots, j, j+1, \ldots, j+k) = (j+1, \ldots, j+k, 1, \ldots j).$$

Then $(-1)^\tau = (-1)^{jk}$. We have

$$\begin{aligned}
\alpha \wedge \beta &= \frac{1}{(j+k)!} \sum_{\sigma \in \mathfrak{S}_{j+k}} (-1)^\sigma \alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(j)}\big)\, \beta\big(v_{\sigma(j+1)}, \ldots v_{\sigma(j+k)}\big) \\
&= \frac{1}{(j+k)!} \sum_{\sigma \in \mathfrak{S}_{j+k}} (-1)^\sigma \alpha\big(v_{\sigma\tau(k+1)}, \ldots, v_{\sigma\tau(k+j)}\big)\, \beta\big(v_{\sigma\tau(1)}, \ldots v_{\sigma\tau(k)}\big) \\
&= \frac{(-1)^{jk}}{(j+k)!} \sum_{\sigma \in \mathfrak{S}_{j+k}} (-1)^{\sigma\tau} \beta\big(v_{\sigma\tau(1)}, \ldots v_{\sigma\tau(k)}\big)\, \alpha\big(v_{\sigma\tau(k+1)}, \ldots, v_{\sigma\tau(k+j)}\big) \\
&= (-1)^{jk} \beta \wedge \alpha,
\end{aligned}$$

which completes the proof. □

1.2. Notation.

It is an unfortunate fact of life that the anti-commutativity established in Proposition 1.2 leads to unpleasant calculations with differential forms. The following notation, although at first sight excessive, actually helps to preserve sanity. To begin with, we fix the underlying dimension n, and define certain index sets:

(i) $\mathbb{I}_m$ denotes the set of all ordered m-tuples $K = (k_1, \ldots, k_m)$ with $1 \leq k_j \leq n$ for $1 \leq j \leq m$.

(ii) $\mathbb{I}_m^\#$ denotes the subset of $\mathbb{I}_m$ consisting of those m-tuples $K = (k_1, \ldots, k_m)$ such that $k_i \neq k_j$ for $i \neq j$. Thus $\mathbb{I}_m^\#$ is the set of m-tuples with distinct elements.

(iii) $\mathbb{I}_m^*$ denotes the subset of $\mathbb{I}_m^\#$ consisting of those q-tuples $K = (k_1, \ldots, k_m)$ such that $1 \leq k_1 < k_2 < \cdots < k_{m-1} < k_m \leq n$.

(iv) If $K \in \mathbb{I}_m^\#$ then $\{K\}$ denotes the unordered set of indices belonging to the ordered m-tuple K.

(v) If $K \in \mathbb{I}_m$ then $[K] \in \mathbb{I}_m$ be the m-tuple consisting of the elements of $\{K\}$ rearranged in (weakly) increasing order.

(vi) If $J = (j_1, \ldots, j_r) \in \mathbb{I}_r$ and $K = (k_1, \ldots, k_s) \in \mathbb{I}_s$, then

$$(J, K) = (j_1, \ldots, j_r, k_1, \ldots, k_s) \in \mathbb{I}_{r+s}.$$

Next, we introduce notation that reflects that anti-commutativity of quantities indexed by these sets. Thus if $J = (j_1, \ldots, j_r)$, $K = (k_1, \ldots, k_r) \in \mathbb{I}_r$ set

$$\epsilon_K^J = \begin{cases} 0 & \text{if either } J \notin \mathbb{I}_r^\# \text{ or } K \notin \mathbb{I}_r^\#, \\ 0 & \text{if } J \in \mathbb{I}_r^\#,\ K \in \mathbb{I}_r^\# \text{ and } \{J\} \neq \{K\}, \\ (-1)^\sigma & \text{if } J \in \mathbb{I}_r^\#,\ K \in \mathbb{I}_r^\#,\ \{J\} = \{K\}, \text{ and } \sigma(j_1, \ldots, j_r) = (k_1, \ldots, k_r). \end{cases}$$

For example,

$$\epsilon^{(2,1,3,5)}_{(1,2,3,5)} = -1, \qquad \epsilon^{(3,1,2,5)}_{(1,2,3,5)} = +1, \qquad \epsilon^{(4,1,3,5)}_{(1,2,3,5)} = 0, \qquad \epsilon^{(5,1,3,5)}_{(1,2,3,5)} = 0.$$

In particular, if $K = (k_1, \ldots, k_q) \in \mathbb{I}_q$, $L = (l_1, \ldots, l_{q+1}) \in \mathbb{I}_{q+1}$, and $1 \leq k \leq n$, we have

$$\epsilon^{(k,K)}_L = \begin{cases} 0 & \text{if } \{k\} \cup \{K\} \neq \{L\}, \\ \\ (-1)^\sigma & \text{if } \{k\} \cup \{K\} = \{L\} \text{ and } \sigma(k, k_1, \ldots, k_q) = (l_1, \ldots, l_{q+1}). \end{cases}$$

1.3. Differential forms with coordinates.

We now construct a basis for Λ^m_x. Let $e_1, \ldots, e_n$ be any basis of T_x, and let $\alpha_1, \ldots, \alpha_n$ be the dual basis of $\Lambda^1_x = T^*_x$, so that

$$\alpha_j[e_l] = \delta_{j,l} = \begin{cases} 1 & \text{if } j = l, \\ 0 & \text{if } j \neq l. \end{cases}$$

Let $J = (j_1, \ldots, j_m), K = (k_1, \ldots, k_m) \in \mathbb{I}_m$. It follows from Definition 1.1, (2) that

$$\begin{aligned}(m!)\, \alpha_{j_1} \wedge \cdots \wedge \alpha_{j_m}\big(e_{k_1}, \ldots, e_{k_m}\big) &= \sum_{\sigma \in \mathfrak{S}_m} (-1)^\sigma\, \alpha_{j_1}[e_{k_{\sigma(1)}}] \cdots \alpha_{j_m}[e_{k_{\sigma(m)}}] \\ &= \begin{cases} \epsilon^J_K & \text{if } J, K \in \mathbb{I}^\#_m \\ \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

On the other hand, an element $\alpha \in \Lambda^m_x$ is determined by its action on all m-tuples $(e_{k_1}, \ldots, e_{k_m})$. If we write $\alpha(e_{k_1}, \ldots, e_{k_m}) = \alpha_{k_1,\ldots,k_m} \in \mathbb{R}$, it follows from the anti-symmetry of α that

$$\alpha_{k_1,\ldots,k_k} = (-1)^\sigma \alpha_{k_{\sigma(1)},\ldots,k_{\sigma(m)}}.$$

Thus we can write

$$\alpha = \sum_{(k_1,\ldots,k_m) \in \mathbb{I}_m} \alpha(e_{k_1}, \ldots, e_{k_m})\, \alpha_{k_1} \wedge \cdots \wedge \alpha_{k_m}$$

since both sides have the same value when applied to $(e_{j_1}, \ldots, e_{j_m})$. Because of the anti-commutativity of the wedge products and the coefficients, we can write

$$\alpha = m! \sum_{(j_1,\ldots,j_m) \in \mathbb{I}^*_m} \alpha(e_{j_1}, \ldots, e_{j_m})\, \alpha_{j_1} \wedge \cdots \wedge \alpha_{j_m},$$

and the elements $\{\alpha_J = \alpha_{j_1} \wedge \cdots \wedge \alpha_{j_m}\}$ with $J = (j_1, \ldots, j_m) \in \mathbb{I}^*_m$ are a basis for Λ^m_x.

Now suppose that $X_1, \ldots, X_n \in T(\Omega)$ are vector fields such that for each $x \in \Omega$ the vectors $\{(X_1)_x, \ldots, (X_n)_x\}$ form a basis for for T_x. Define $\omega_j \in \Lambda^1(\Omega)$ by

$$\omega_j[X_k] = \delta_{j,k}.$$

The above discussion establishes the following result.

PROPOSITION 1.3. *The space $\Lambda^m(\Omega)$ is generated as a module over $\mathcal{E}(\Omega)$ by the differential forms*

$$\omega_J = \omega_{j_1} \wedge \cdots \wedge \omega_{j_m}, \qquad 1 \leq j_1 < \cdots < j_k \leq n.$$

The number of generators is $\binom{n}{m}$. Every $\eta \in \Lambda^m(\Omega)$ can be written uniquely as

$$\eta = \sum_{J \in \mathbb{I}_m^*} \eta_J \, \omega_J$$

where each $\eta_J \in \mathcal{E}(\Omega)$.

As a consequence of the notation we introduced, note that if $J \in \mathbb{I}_m^*$ and $1 \leq j \leq n$, we have

$$\omega_j \wedge \omega_J = \sum_{L \in \mathbb{I}_{m+1}^*} \epsilon_L^{(j,J)} \, \omega_L.$$

In later formulas, it will be convenient to consider the coefficients ω_J of a m-form $\omega \in \Lambda^m(\Omega)$ for multi-indices J which are not in $\mathbb{I}_m^*$. Thus for any $J \in \mathbb{I}_m$, set

$$\omega_J = \begin{cases} 0 & \text{if } J \notin \mathbb{I}_m^{\#}, \\ \\ \epsilon_{[J]}^J \, \omega_{[J]} & \text{if } J \in \mathbb{I}_m^{\#}. \end{cases}$$

In particular, we have

$$\epsilon_{[(j,J)]}^{(j,J)} \, \omega_{[(j,J)]} = \omega_{(j,J)}.$$

Since $\Omega \subset \mathbb{R}^n$, we can always choose a basis for the vector fields by choosing $X_j = \dfrac{\partial}{\partial x_j}$. In this case we write the corresponding elements of Λ_x^1 as $\{dx_1, \ldots, dx_n\}$, and every element $\omega \in \Lambda^m(\Omega)$ can be written

$$\omega = \sum_{J \in \mathbb{I}_m^*} \omega_J \, dx_J \qquad \text{with } \omega_J \in \mathcal{E}(\Omega)$$

where, if $J = (j_1, \ldots, j_m)$ we write

$$dx_J = dx_{j_1} \wedge \cdots \wedge dx_{j_m}.$$

1.4. The exterior derivative d.

The exterior derivative d maps m-forms to $(m+1)$-forms and is usually defined in terms of coordinates:

$$df = \sum_{j=1}^n \frac{\partial f}{\partial x_j} \, dx_j \qquad \text{if } f \in \Lambda^0(\Omega),$$

$$d\omega = \sum_{J \in \mathbb{I}_m^*} \sum_{j=1}^n \frac{\partial \omega_J}{\partial x_j} \, dx_j \wedge dx_J \qquad \text{if } \omega = \sum_{J \in \mathbb{I}_m^*} \omega_J \, dx_J \in \Lambda^m(\Omega).$$

However, it will be important to know that this operator has an intrinsic meaning. Thus we begin with a coordinate free definition.

DEFINITION 1.4. *The exterior derivative $d : \Lambda^m(\Omega) \to \Lambda^{m+1}(\Omega)$ is defined as follows:*

(1) *For $m = 0$, if $f \in \mathcal{E}(\Omega) = \Lambda^0(\Omega)$ and if $X \in T(\Omega)$, then*

$$df(X) = X[f].$$

(2) *For $m \geq 1$, if $\omega \in \Lambda^m(\Omega)$ and if $X_1, \ldots, X_{m+1} \in T(\Omega)$, then*

$$\begin{aligned}(m+1)\, d\omega\big(X_1, \ldots, X_{m+1}\big) \\ &= \sum_{j=1}^{m+1} (-1)^{j+1} X_j \big[\omega(X_1, \ldots, \widehat{X_j}, \ldots, X_{m+1})\big] \\ &\quad + \sum_{j<k} (-1)^{j+k} \omega\big(\,[X_j,\, X_m]\,, X_1, \ldots, \widehat{X_j}, \ldots, \widehat{X_k}, \ldots, X_{m+1}\big)\end{aligned}$$

For example, when $m = 1$, if $\eta \in \Lambda^1(\Omega)$ and $X, Y \in T(\Omega)$, then

$$2d\eta(X, Y) = X\big[\eta[Y]\big] - Y\big[\eta[X]\big] - \eta\big[\,[X,\, Y]\,\big].$$

PROPOSITION 1.5. *The exterior derivative d has the following properties:*

(1) *If $\omega \in \Lambda^j(\Omega)$ and $\eta \in \Lambda^k(\Omega)$, then*

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^j\, \omega \wedge d\eta.$$

(2) *If $\omega \in \Lambda^j(\Omega)$ then $d^2(\omega) = 0$.*

(3) *In terms of coordinates,*

(a) *If $f \in \mathcal{E}(\Omega)$ then*

$$df = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\, dx_j.$$

(b) *If $\omega = \sum_{K \in \mathbb{I}_k^*} \omega_J\, dx_J$, then*

$$d\omega = \sum_{J \in \mathbb{I}_k^*} d\omega_J \wedge dx_K = \sum_{J \in \mathbb{I}_k^*} \sum_{j=1}^{n} \frac{\partial \omega_J}{\partial x_j}\, dx_j \wedge dx_J.$$

GIVE PROOF

Let $\Omega \subset \mathbb{R}^n$ be an open set (or more generally a real manifold of dimension n). The operator d gives us the following sequence of mappings:

$$\mathcal{E}(\Omega) = \Lambda^0(\Omega) \xrightarrow{d_0} \Lambda^1(\Omega) \xrightarrow{d_1} \Lambda^2(\Omega) \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} \Lambda^n(\Omega).$$

Since $d^2 = 0$, at each point the image of one mapping is contained in the null space of the next. This sequence is called the deRham complex for Ω. The quotient groups

$$H^m(\Omega, \mathbb{R}) = \text{Kernel}(d_m)/\text{Image}(d_{m-1})$$

are the deRham cohomology groups of Ω.

1.5. Complex coordinates and holomorphic functions.

We now turn to a discussion of differential forms in complex spaces. We denote points of $\mathbb{C}^n$ by $z = (z_1, \ldots, z_n)$, where $z_j = x_j + iy_j$. With these coordinates we can identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ via the mapping

$$\mathbb{C}^n \ni z = (z_1, \ldots, z_n) \longleftrightarrow (x_1, y_1, \ldots, x_n, y_n) \in \mathbb{R}^{2n}.$$

However, complex analysis is not just real analysis in twice as many variables. The key point is that for open sets in $\Omega \subset \mathbb{C}^n$ there is a distinguished class of complex-valued smooth functions $\mathcal{O}(\Omega)$ called *holomorphic functions*. There are many equivalent characterizations of this class. We shall define them as the solutions of a system of complex first order partial differential operators.

We proceed as follows. In addition to the usual real vector fields

$$X_j = \frac{\partial}{\partial x_j} \qquad \text{and} \qquad Y_j = \frac{\partial}{\partial y_j}$$

for $1 \leq j \leq n$, we introduce certain special *complex-valued* vector fields:

$$\begin{aligned} Z_j &= \frac{\partial}{\partial z_j} = \frac{1}{2}\Big[\frac{\partial}{\partial x_j} - i\frac{\partial}{\partial y_j}\Big] = \frac{1}{2}\big[X_j - iY_j\big] \qquad \text{and} \\ \overline{Z}_j &= \frac{\partial}{\partial \bar{z}_j} = \frac{1}{2}\Big[\frac{\partial}{\partial x_j} + i\frac{\partial}{\partial y_j}\Big] = \frac{1}{2}\big[X_j + iY_j\big]. \end{aligned} \tag{1.1}$$

It follows that

$$\begin{aligned} X_j &= \ (Z_j + \overline{Z}_j), \\ Y_j &= i(Z_j - \overline{Z}_j). \end{aligned} \tag{1.2}$$

Definition 1.6. *Let $\Omega \subset \mathbb{C}^n$ be open. A complex-valued function $f \in \mathcal{E}_{\mathbb{C}}(\Omega)$ is holomorphic if it satisfies $\overline{Z}_j[f](z) = 0$ for all $z \in \Omega$ and all $1 \leq j \leq n$. These equations,*

$$\frac{\partial f}{\partial \bar{z}_j}(z) = \frac{\partial f}{\partial x_j}(z) + i\frac{\partial f}{\partial y_j}(z) = 0, \quad \textit{for} \quad 1 \leq j \leq n,$$

are called the (homogeneous) Cauchy-Riemann equations.

The property of being holomorphic is independent of the choice of a coordinate system. In fact, we have the following result which follows easily from the chain rule.

Proposition 1.7. *Let $\Omega_1 \subset \mathbb{C}^n$ and $\Omega_2 \subset \mathbb{C}^m$ be open sets, and let $F = (f_1, \ldots, f_m) : \Omega_1 \to \Omega_2$ be a mapping such that each component function f_j is holomorphic on Ω_1.*

(1) If $h : \Omega_2 \to \mathbb{C}$ is a holomorphic function, then $H(z) = h\big(F(z)\big)$ is a holomorphic function on Ω_1.

(2) If $m = n$ and F is a diffeomorphism (so that F is one-to-one, onto, and has a smooth inverse F^{-1}), then the components of $F^{-1} : \Omega_2 \to \Omega_1$ are holomorphic functions.

1.6. Splitting the complexified tangent space.

The existence of the class of holomorphic functions on an open set $\Omega \subset \mathbb{C}^n$ leads to a natural splitting of the complexified tangent space. Recall that for any point $z \in \mathbb{C}^n$, the real tangent space T_z is a real $2n$-dimensional vector space over $\mathbb{R}$. T_z is spanned by the real vector fields $\{X_1, Y_1, \ldots, X_n, Y_n\}$. If $\Omega \subset \mathbb{C}^n$ is an open set, $T(\Omega)$ is the space of smooth, *real-valued* vector fields on Ω, and if $z \in \Omega$, T_z is just the restriction to the point z of elements of $T(\Omega)$.

The complexified tangent space $T_z \otimes \mathbb{C}$ is obtained by taking all complex linear combinations of $\{X_1, Y_1, \ldots, X_n, Y_n\}$. This space has complex dimension $2n$, is denoted by $\mathbb{C}T_z$, and is called the *complex tangent space* at z. The real tangent space T_z is a subspace of $\mathbb{C}T_z$, but it is not closed under multiplication by i and hence is not a complex subspace. It is clear from equations (1.1) and (1.2) that the set of real vector fields $\{X_1, Y_1, \ldots, X_n, Y_n\}$ and the set of complex vector fields $\{Z_1, \ldots, Z_n, \overline{Z}_1, \ldots, \overline{Z}_n\}$ are both bases for the vector space $\mathbb{C}T_z$ over $\mathbb{C}$.

DEFINITION 1.8. *The complex structure mapping J is the real-linear transformation $J : T_z \to T_z$ uniquely determined by setting*

$$J[X_j] = Y_j \qquad \text{and} \qquad J[Y_j] = -X_j$$

for $1 \leq j \leq n$. It follows that $J^2 = -I$, so the eigenvalues of J are $\pm i$. J extends to the complex vector space $\mathbb{C}T_z$ as a complex-linear mapping. Let

$$T_z^{1,0} = \left\{ v \in \mathbb{C}T_z \,\middle|\, J[v] = +iv \right\},$$
$$T_z^{0,1} = \left\{ v \in \mathbb{C}T_z \,\middle|\, J[v] = -iv \right\}$$

be the subspaces of corresponding eigenvectors.

It follows that we have the following decomposition of the complexified tangent space:

$$\mathbb{C}T_z = T_z^{1,0} \oplus T_z^{0,1}. \tag{1.3}$$

Note that we have

$$J[Z_j] = J\big[(X_j - iY_j)\big] = J[X_j] - iJ[Y_j] = Y_j + iX_j = (+i)(X_j - iY_j)$$
$$J[\overline{Z}_j] = J\big[(X_j + iY_j)\big] = J[X_j] + iJ[Y_j] = Y_j - iX_j = (-i)(X_j + iY_j).$$

Thus $\{Z_1, \ldots, Z_n\}$ are a basis for $T_z^{1,0}$ and $\{\overline{Z}_1, \ldots, \overline{Z}_n\}$ are a basis for $T_z^{0,1}$.

DEFINITION 1.9. *A complex-valued vector field X on Ω is of type $(1,0)$ (respectively, of type $(0,1)$) if for every $z \in \Omega$ we have $X_z \in T_z^{1,0}$ (respectively $X_z \in T_z^{0,1}$).*

PROPOSITION 1.10. *If Z, W are complex vector fields of type $(1,0)$ then $[Z, W]$ is again a complex vector field of type $(1,0)$. If $\overline{Z}, \overline{W}$ are complex vector fields of type $(0,1)$ then $\big[\overline{Z}, \overline{W}\big]$ is again a complex vector field of type $(0,1)$.*

PROOF. If Z, W are vector fields of type $(1,0)$, we can write $Z = \sum_{j=1}^n a_j\, Z_j$ and $W = \sum_{k=1}^n b_k\, Z_k$. Then

$$[Z, W] = \sum_{j=1}^n \big(Z[a_j] - W[b_j]\big)\, Z_j$$

is again of type $(1,0)$. A similar calculation works for vector fields of type $(0,1)$. $\square$

It is not immediately clear from Definition 1.8 that the mapping J is independent of the choice of holomorphic coordinates. However, the following result shows that there is an intrinsic characterization of J in terms of the class of holomorphic functions.

PROPOSITION 1.11. *Let $z \in \mathbb{C}^n$ and let $v \in T_z$. Then $J[v] \in T_z$ is the unique vector such that $\big[v + iJ[v]\big][f] = 0$ for every function f which is holomorphic in a neighborhood of z.*

PROOF. Write $v = \sum_{k=1}^n \big(\alpha_k X_k + \beta_k Y_k\big)$. Then $J[v] = \sum_{k=1}^n \big(\alpha_k Y_k - \beta_k X_k\big)$, and so

$$v + iJ[v] = \sum_{k=1}^n (\alpha_k - i\beta_k)(X_k + iY_k) = 2\sum_{k=1}^n (\alpha_k - i\beta_k)\overline{Z}_k.$$

Since $\overline{Z}_k[f] = 0$ for every holomorphic function f, the same is true for $v + iJ[v]$. To prove uniqueness, observe that if $w \in T_z$ and $(v + iw)[f] = 0$ for every holomorphic function f, then $(w - J[v])[f] = 0$ for all holomorphic functions. However $w - J[v] \in T_z$ so we can write $w - J[v] = \sum_{k=1}^n (a_j X_j + b_j Y_j)$ with $a_j, b_j \in \mathbb{R}$. If we apply this to $z_j = x_j + iy_j$, it follows that $a_j + ib_j = 0$, so $w = J[v]$. □

1.7. Complex differential forms.

We now study the complexifications of the spaces Λ_z^m. Thus $\mathbb{C}\Lambda_z^m = \Lambda_z^m \otimes \mathbb{C}$ is the space of alternating complex-valued m-linear mappings of $\mathbb{C}T_z$ to $\mathbb{C}$. In particular, when $m = 1$, the space $\mathbb{C}\Lambda_z^1$ is just the (complex) dual space of the (complex) vector space $\mathbb{C}T_z$, and is sometimes called the complex cotangent space. This space also has a natural decomposition as the direct sum of two subspaces.

DEFINITION 1.12. *Let*

$$\Lambda_z^{1,0} = \Big\{\omega \in (\mathbb{C}T_z)^* \,\Big|\, \omega[v] = 0 \text{ for all } v \in T_z^{0,1}\Big\} = (T_z^{0,1})^\perp,$$
$$\Lambda_z^{0,1} = \Big\{\omega \in (\mathbb{C}T_z)^* \,\Big|\, \omega[v] = 0 \text{ for all } v \in T_z^{1,0}\Big\} = (T_z^{1,0})^\perp.$$

It then follows from equation (1.3) that

$$\mathbb{C}\Lambda_z^1 = \Lambda_z^{1,0} \oplus \Lambda_z^{0,1}. \tag{1.4}$$

There is a second way of obtaining this decomposition. The dual of the structure map $J : \mathbb{C}T_z \to \mathbb{C}T_z$ is a complex linear mapping $J^* : \mathbb{C}\Lambda_z^1 \to \mathbb{C}\Lambda_z^1$, and it is easy to check that

$$J^*[dx_j] = dy_j \qquad \text{and} \qquad J^*[dy_j] = -dx_j,$$

since $\{dx_1, \ldots, dx_n, dy_1, \ldots, dy_n\}$ is the dual basis to $\{X_1, \ldots, X_n, Y_1, \ldots, Y_n\}$. Now since $J^2 = -I$, it follows that $(J^*)^2 = -I$. Thus the space $\mathbb{C}\Lambda_z^1$ is the direct sum of the two subspaces which are the eigenspaces of J^* corresponding to eigenvalues $+i$ and $-i$.

PROPOSITION 1.13. *For $z \in \mathbb{C}^n$, $\Lambda_z^{1,0}$ is the subspace of $\mathbb{C}\Lambda_z^1$ consisting of eigenvectors for J^* with eigenvalue $+i$, and $\Lambda_z^{0,1}$ is the subspace of $\mathbb{C}\Lambda_z^1$ consisting of eigenvectors for J^* with eigenvalue $-i$.*

PROOF. We have already observed that $Z_1, \ldots, Z_n \in \mathbb{C}T_z$ are eigenvectors of J with eigenvalues $+i$, and $\overline{Z}_1, \ldots, \overline{Z}_n \in \mathbb{C}T_z$ are eigenvectors of J with eigenvalues $-i$. The dual basis of $\mathbb{C}\Lambda^1_z$ are the elements

$$dz_j = dx_j + i dy_j \qquad \text{and} \qquad d\bar{z}_j = dx_j - i dy_j,$$

with $1 \le j \le n$. We have $J^*[dz_j] = i\, dz_j$ and $J^*[d\bar{z}_j] = -i\, d\bar{z}_j$. Thus the space of eigenvectors of J^* with eigenvalue $+i$ is spanned by $\{dz_1, \ldots, dz_n\}$, and the space of eigenvectors of J^* with eigenvalue $-i$ is spanned by $\{d\bar{z}_1, \ldots, d\bar{z}_n\}$. It is now easy to check that these are, respectively, the spaces $(T^{1,0}_z)^\perp$ and $(T^{0,1}_z)^\perp$. □

The decomposition (1.3) of $\mathbb{C}T_z$ also leads to a decomposition of the spaces $\mathbb{C}\Lambda^m_z$ for $m > 1$.

DEFINITION 1.14. *Let $p + q = m$. Then*

$$\Lambda^{p,q}_z = \Big\{\omega \in \mathbb{C}\Lambda^m_z \;\Big|\; \omega(v_1, \ldots, v_m) = 0 \quad \textit{if } v_1, \ldots, v_j \in T^{1,0}_z,\ v_{j+1}, \ldots, v_m \in T^{0,1}_z \textit{ and } j \neq p\Big\}.$$

It is not hard to check that

$$\mathbb{C}\Lambda^m_z = \bigoplus_{p+q=m} \Lambda^{p,q}_z, \tag{1.5}$$

and $\Lambda^{p,q}_z$ is spanned by the set of all forms $dz_J \wedge d\bar{z}_K$ where

$$dz_J = dz_{j_1} \wedge \cdots \wedge dz_{j_p},$$
$$d\bar{z}_K = d\bar{z}_{k_1} \wedge \cdots \wedge d\bar{z}_{k_q},$$

and $J \in \mathbb{I}^*_p$, $K \in \mathbb{I}^*_k$ so that $1 \le j_1 < \cdots < j_p \le n$ and $1 \le k_1 < \cdots < k_q \le n$.

If $\Omega \subset \mathbb{C}^n$ is open, we let $\mathbb{C}\Lambda^m(\Omega)$ denote the space of complex valued m-forms on Ω, amd $\Lambda^{p,q}(\Omega)$ denote the subspace of complex valued $(p+q)$-forms whose value at z lies in $\Lambda^{p,q}_z$. We have

$$\mathbb{C}\Lambda^m(\Omega) = \bigoplus_{p+q=m} \Lambda^{p,q}(\Omega). \tag{1.6}$$

An element $\omega \in \mathbb{C}\Lambda^{p,q}(\Omega)$ can be written

$$\omega = \sum_{J \in \mathbb{I}^*_p,\, K \in \mathbb{I}^*_q} \omega_{J,K}\, dz_J \wedge d\bar{z}_K$$

where $\omega_{J,K}$ is now a *complex-valued* infinitely differentiable function on Ω.

1.8. The $\overline{\partial}$-operator.

The exterior derivative extends to a mapping on complex-valued differential forms $d : \mathbb{C}\Lambda^m(\Omega) \to \mathbb{C}\Lambda^{m+1}(\Omega)$ with the properties established in Proposition 1.5. Moreover, the decomposition given in (1.6) leads to a decomposition of the opertor d into the sum of two operators ∂ and $\overline{\partial}$. The following is the key relationship between d and the decomposition given in (1.6).

PROPOSITION 1.15. *Let $\omega \in \Lambda^{p,q}(\Omega)$. Then $d\omega \in \Lambda^{p+1,q}(\Omega) \oplus \Lambda^{p,q+1}(\Omega)$.*

PROOF. Let $p+q=m$, let $\omega \in \Lambda^{p,q}(\Omega)$, and let $\{X_1,\ldots,X_{m+1}\}$ be complex vector fields with $X_1,\ldots,X_j$ of type $(1,0)$ and $X_{j+1},\ldots,X_{m+1}$ of type $(0,1)$. We need to show that $d\omega(X_1,\ldots,X_{m+1})=0$ if $j\notin\{p,p+1\}$.

There are two kinds of terms in the definition of $d\omega$. The first arise the sum

$$\sum_{r=1}^{m+1}(-1)^{r+1}X_l\big[\omega(X_1,\ldots,\widehat{X_r},\ldots,X_{m+1})\big].$$

Since $\omega\in\Lambda^{p,q}(\Omega)$, we have $\omega(X_1,\ldots,\widehat{X_r},\ldots,X_{m+1})=0$ unless there are p vectors of type $(1,0)$ and q vectors of type $(0,1)$ in the set $\{X_1,\ldots,\widehat{X_r},\ldots,X_{m+1}\}$. But since X_r is itself either of type $(1,0)$ or $(0,1)$, this would imply that $j\in\{p,p+1\}$. Hence all these terms must vanish.

The other kind of terms arise in the sum

$$\sum_{r<s}(-1)^{r+s}\omega\big([X_r,\,X_s]\,,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\big).$$

Suppose that X_r and X_s are both of type $(1,0)$. Then according to Proposition 1.10, $[X_r,\,X_s]$ is also of type $(1,0)$. If

$$\omega\big([X_r,\,X_s]\,,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\big)\neq 0,$$

it follows that the set $\{[X_r,\,X_s]\,,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\}$ contains p vectors of type $(1,0)$. It follows that $j=p+1$. Similarly, if X_r and X_s are both of type $(0,1)$, it follows that $j=p$.

Finally if the vectors X_r and X_s are of different type, we can write $[X_r,\,X_s]=U+V$ where U is of type $(1,0)$ and V is of tyep $(0,1)$. But if

$$\omega\big(U,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\big)\neq 0,$$

it follows that the set $\{U,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\}$ must contain p vectors of type $(0,1)$, in which case $j=p+1$. If

$$\omega\big(V,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\big)\neq 0,$$

it follows that the set $\{V,X_1,\ldots,\widehat{X_r},\ldots,\widehat{X_s},\ldots,X_{m+1}\}$ must contain p vectors of type $(0,1)$, in which case $j=p$. Thus all these terms must vanish as well, and this completes the proof. □

DEFINITION 1.16. *Let $\pi_{p,q}$ be the projection from $\mathbb{C}\Lambda^{p+q}(\Omega)$ to $\Lambda^{p,q}(\Omega)$. Then*

$$\partial:\Lambda^{p,q}(\Omega)\to\Lambda^{p+1,q}(\Omega)\qquad\textit{and}\qquad\overline{\partial}:\Lambda^{p,q}(\Omega)\to\Lambda^{p,q+1}(\Omega)$$

are defined by $\partial=\pi_{p+1,q}\,d$ and $\overline{\partial}=\pi_{p,q+1}\,d$.

It follows from Proposition 1.15 that $d=\partial+\overline{\partial}$. Since $d^2=0$, it follows that $\partial^2+\partial\overline{\partial}+\overline{\partial}\partial+\overline{\partial}^2=0$. But if $\omega\in\Lambda^{p,q}(\Omega)$, then $\partial^2[\omega]\in\Lambda^{p+2,q}$, $(\partial\overline{\partial}+\overline{\partial}\partial)[\omega]\in\Lambda^{p+1,q+1}$, and $\overline{\partial}^2[\omega\in\Lambda^{p,q+2}$. Since these spaces intersect only at (0), it follows that

$$\partial^2=0,\qquad\partial\overline{\partial}+\overline{\partial}\partial=0,\qquad\overline{\partial}^2=0. \tag{1.7}$$

The other basic properties of the operators ∂ and $\overline{\partial}$ now follow from Proposition 1.5.

PROPOSITION 1.17.

(1) For any differential forms $\omega \in \mathbb{C}\Lambda^l(\Omega)$ and $\eta \in \mathbb{C}\Lambda^m(\Omega)$, we have

$$\partial(\omega \wedge \eta) = \partial\omega \wedge \eta + (-1)^l \omega \wedge \partial\eta,$$
$$\overline{\partial}(\omega \wedge \eta) = \overline{\partial}\omega \wedge \eta + (-1)^l \omega \wedge \overline{\partial}\eta.$$

(2) In local coordinates,

(a) if $f \in \mathcal{E}_{\mathbb{C}}(\Omega)$ then

$$\partial f = \sum_{j=1}^{n} \frac{\partial f}{\partial z_j}\, dz_j \qquad \text{and} \qquad \overline{\partial} f = \sum_{j=1}^{n} \frac{\partial f}{\partial \bar{z}_j}\, d\bar{z}_j;$$

(b) If $\omega = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \omega_{J,K}\, dz_J \wedge d\bar{z}_K \in \Lambda^{p,q}(\Omega)$, then

$$\partial\omega = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \partial[\omega] \wedge dz_J \wedge d\bar{z}_K = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \sum_{l=1}^{n} \frac{\partial \omega_{J,K}}{\partial z_l}\, dz_l \wedge dz_J \wedge d\bar{z}_K,$$

$$\overline{\partial}\omega = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \overline{\partial}[\omega] \wedge dz_J \wedge d\bar{z}_K = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \sum_{l=1}^{n} \frac{\partial \omega_{J,K}}{\partial \bar{z}_l}\, d\bar{z}_l \wedge dz_J \wedge d\bar{z}_K.$$

1.9. The $\overline{\partial}$-problem.

In analogy with the real situation, we see that if $\Omega \subset \mathbb{C}^n$ is an open set (or more generally a complex manifold of dimension n), then for each integer $0 \leq p \leq n$, the operator $\overline{\partial}$ gives us the following sequence of mappings:

$$\Lambda^{p,0}(\Omega) \xrightarrow{\overline{\partial}_0} \Lambda^{p,1}(\Omega) \xrightarrow{\overline{\partial}_1} \Lambda^{p,2}(\Omega) \xrightarrow{\overline{\partial}_2} \cdots \xrightarrow{\overline{\partial}_{n-1}} \Lambda^{p,n}(\Omega).$$

Since $\overline{\partial}_{k+1}\overline{\partial}_k = 0$, at each point the image of one mapping is contained in the null space of the next. This is called the Dolbeault-complex. The corresponding quotients

$$H^{p,q}(\Omega, \mathbb{C}) = \text{Kernel}(\overline{\partial}_q)/\text{Image}(\overline{\partial}_{q-1})$$

are called the Dolbeault cohomology groups. See, for example [**GR65**].

Note that a function $u \in \Lambda^{0,0}(\Omega)$ is holomorphic if and only if $\overline{\partial}[u] = 0$, and these are the homogeneous Cauchy-Riemann equations. For many reasons, it is also important to consider the inhomogeneous version

$$\overline{\partial}[u] = g$$

where $g \in \Lambda^{0,1}(\Omega)$ is a given $(0,1)$-form. Since $\overline{\partial}^2 = 0$, a necessary condition for finding a solution for the unknown function u is that $\overline{\partial}[g] = 0$. The $\overline{\partial}$-*problem* can be stated rather vaguely as follows:

> Determine if the equation $\overline{\partial}[u] = g$ has a solution u when g is a given $(p, q+1)$-form such that $\overline{\partial}[g] = 0$. If a solution does exist, find a solution u_g with good regularity properties that reflect the regularity of g.

2. The $\overline{\partial}$-Neumann problem

Part of the difficulty of the $\overline{\partial}$-problem is that a solution to $\overline{\partial}[u] = g$, if it exists, is not unique. Thus if u is a function, then $\overline{\partial}[u+h] = \overline{\partial}[u]$ for any function h which is holomorphic. We now turn to one classical approach to the lack of uniqueness, which is to choose the particular solution which is orthogonal, in an appropriate sense, to the null space of $\overline{\partial}$.

From an algebraic point of view, we are given linear transformations $\overline{\partial}_{q-1}$ and $\overline{\partial}_q$ between certain vector spaces of differential forms, with $\overline{\partial}_q\overline{\partial}_{q-1} = 0$. We want to know if the range of $\overline{\partial}_{q-1}$ is equal to the null space of $\overline{\partial}_q$. In order to talk about orthogonality, we need to give a Hilbert space structure to the vector spaces of forms, and find a way of establishing the existence of solutions. We start with an easy finite dimensional analogue of this problem. This shows that the existence of a certain estimate (equation (2.1) below) implies that solutions exist. We will later attempt to imitate this procedure in the infinite dimensional situation.

2.1. A finite dimensional analogue.

PROPOSITION 2.1. *Suppose U, V, W are finite dimensional Hilbert spaces, and that $S : U \to V$ and $T : V \to W$ are linear maps with $TS = 0$. Let $S^* : V \to U$ and $T^* : W \to V$ be the adjoint mappings. Put*[3]

$$L = S\,S^* + T^*\,T : V \to V.$$

(1) *If L is invertible with inverse N, and if $v \in V$ with $T[v] = 0$, then $u = S^*N[v] \in U$ is the (unique) solution to $S[u] = v$ which is orthogonal to the null space of S.*

(2) *The following statements are equivalent:*
  (a) *There is a constant $C > 0$ so that for every $v \in V$,*

$$\big|\big|\, v \,\big|\big|_V^2 \leq C^2 \Big[\big|\big|\, S^*[v] \,\big|\big|_U^2 + \big|\big|\, T[v] \,\big|\big|_W^2\Big]. \tag{2.1}$$

  (b) *The mapping L is one-to-one and onto.*
  (c) *The range of the mapping S equals the null space $N(T)$ of the mapping T.*

PROOF. If $T[v] = 0$ and N is the inverse of L it follows that

$$v = (SS^* + T^*T)\big[N[v]\big].$$

Applying T to both sides we get

$$0 = T[v] = TSS^*N[v] + TT^*TN[v] = TT^*TN[v]$$

since $TS = 0$. Thus $TT^*TN[v] = 0$, and so

$$0 = (TT^*TN[v], TN[v])_W = (T^*TN[v], T^*TN[v])_V = \big|\big|\, T^*TN[v] \,\big|\big|_V.$$

It follows that $T^*TN[v] = 0$. But then

$$v = SS^*N[v] + T^*TN[v] = S\big[S^*N[v]\big].$$

[3]In our later examples, the operator L will be the Laplace operator acting on components of a differential form. The inverse which satisfies the required boundary conditions is called the Neumann operator.

Thus $u = S^*N[v]$ is a solution to $S[u] = v$. Moreover, the range of S^* is always orthogonal to the null space $N(S)$ of S, so we have established the first assertion.

Next, suppose that assertion (a) is true. If $v \in V$ with $L[v] = 0$, we have

$$\begin{aligned} 0 &= (L[v], v)_V = (SS^*[v], v)_V + (T^*T[v], v)_V \\ &= (S^*[v], S^*[v])_U + (T[v], T[v])_W = \big|\big| S^*[v] \big|\big|_U^2 + \big|\big| T[v] \big|\big|_W^2. \end{aligned}$$

But then

$$\big|\big| v \big|\big|_V^2 \leq C^2 \left[\big|\big| S^*[v] \big|\big|_U^2 + \big|\big| T[v] \big|\big|_W^2\right] = 0,$$

so $v = 0$, and hence L is one-to-one. Since V is finite dimensional, this implies that L is onto, and hence assertion (b) is true. (Note that we will need a more involved argument in the infinite dimensional case).

Now if assertion (b) is true, then assertion (1) shows that the range of S equals the null space of T, so assertion (c) is true. Finally, if (c) holds, then since orthogonal complement of the null space of S^* is the (closure of the) range of S, it follows that $V = N(T) \oplus N(S^*)$. Thus if $v \in V$ we can write $v = v_1 + v_2$ where $v_1 \in N(T)$ and $v_2 \in N(S^*)$. We have $\big|\big| v \big|\big|_V^2 = \big|\big| v_1 \big|\big|_V^2 + \big|\big| v_2 \big|\big|_V^2$. Now S^* restricted to $N(S^*)^\perp = N(T)$ is one-to-one, and hence $\big|\big| v_1 \big|\big|_V \leq C \big|\big| S^*[v_1] \big|\big|_U = C \big|\big| S^*[v] \big|\big|_U$. Similarly, we have $\big|\big| v_2 \big|\big|_V \leq C \big|\big| T[v_2] \big|\big|_U = C \big|\big| T[v] \big|\big|_U$, and this establishes assertion (a). □

2.2. Hilbert spaces.

In trying to develop an analogue of Proposition 2.1 for use in the $\overline{\partial}$-problem, there are several difficulties. After putting a Hilbert space structure on the space of differential forms $\mathcal{E}(\overline{\Omega})_0^{0,q}$, we need to establish the analogue of the inequality in equation (2.1), and we also must deal with the fact that differential operators like $\overline{\partial}_q$ are not bounded operators on L^2-spaces. We defer the first problem, which is specific to the $\overline{\partial}$-problem, to Section 2.5. In this section we continue to work more abstractly, and develop an analogue of Proposition 2.1 for closed, densely defined operators in infinite dimensional Hilbert spaces.

Thus suppose that U, V, and W are Hilbert spaces and $S : U \to V$ and $T : V \to W$ are closed, densely defined linear mappings. It follows that the Hilbert space adjoints $S^* : V \to U$ and $T^* : W \to V$ are also closed and densely defined. (See Chapter 5, Section 2 for background information on unbounded operators). Thus as before, we have

$$U \xrightarrow{S} V \xrightarrow{T} W,$$

but of course S and T are now not defined on all of U and V. Let $N(S)$ and $N(T)$ denote the null spaces of S and T, and let $R(S)$ and $R(T)$ denote the ranges of these operators. Let

$$H = \text{Dom}(T) \cap \text{Dom}(S^*) \subset V.$$

For $h_1, h_2 \in H$, put

$$Q(h_1, h_2) = (S^*[h_1], S^*[h_2])_U + (T[h_1], T[h_2])_W.$$

Put

$$\text{Dom}(L) = \left\{ v \in V \,\middle|\, v \in H,\ S^*[v] \in \text{Dom}(S), \text{ and } T[v] \in \text{Dom}(T^*) \right\},$$

and for $v \in \text{Dom}(L)$, put

$$L[v] = S\,S^*[v] + T^*.T[v].$$

We make three basic assumptions:

(i) $R(S) \subset N(T)$ so that $TS = 0$.

(ii) H is dense in V.

(iii) There is a constant C so that for all $h \in H$

$$\big|\big|\, h \,\big|\big|_V^2 \leq C^2 \Big[\big|\big|\, T[h] \,\big|\big|_W^2 + \big|\big|\, S^*[h] \,\big|\big|_U^2\Big]. \tag{2.1}$$

Then the main result is the following:

LEMMA 2.2. *The range of S equals the null space of T. In addition, there is a unique bounded linear operator $N : V \to H \subset V$ with the following properties:*

(1) If C is the constant from inequality (2.1), then $\big|\big|\, N[v] \,\big|\big|_V \leq C^2 \,\big|\big|\, v \,\big|\big|_V$ for all $v \in V$.

(2) For any $h \in H$ and $v \in V$ we have $(h, v) = Q\big(h, N[v]\big)$.

(3) $N : V \to Dom(L) \subset H$. Moreover, $N : N(T) \to H_1 = N(T) \cap Dom(S^*)$ and $N : N(S^*) \to H_2 = N(S^*) \cap Dom(T)$.

(4) N is one-to-one on V and L is one-to-one on $Dom(L)$.

(5) N is the inverse of L in the following sense:

$$\begin{aligned} L\big[N[v]\big] &= v \qquad \textit{for every } v \in V, \\ N\big[L[v]\big] &= v \qquad \textit{for every } v \in Dom(L), \end{aligned}$$

(6) The operator N is self adjoint, and the operator L is closed, densely defined, and self-adjoint.

(7) If $v \in V \cap Dom(T)$ with $T[v] = 0$, then $u = S^*N[v] \in U$ is the unique solution to $S[u] = v$ which is orthogonal to the null space of S.

The proof will requires several preliminary steps. We first establish the following facts. Part (4) is the first statement of Lemma 2.2.

PROPOSITION 2.3.

(1) $\overline{R(S)} = N(T)$.

(2) $R(S^*) = \Big\{u \in V \,\Big|\, \big(\exists v \in Dom(S^*) \cap N(T)\big)\,\big(u = S^*[v]\big)\Big\}$.

(3) $R(S^*)$ is closed.

(4) $R(S)$ is closed, and hence $R(S) = N(T)$.

(5) $R(T) = \Big\{w \in W \,\Big|\, \big(\exists v \in Dom(T) \cap N(S^*)\big)\,\big(w = S^*[v]\big)\Big\}$.

(6) $R(T)$ is closed, and hence $R(T^*)$ is closed.

PROOF OF (1): We have $\overline{R(S)} \subset N(T)$. If these closed subspaces are not equal, there is a non-zero vector $v \in N(T) \cap R(S)^{\perp}$. Then for every $u \in \text{Dom}(S)$ we have $|(S[u], v)| = 0 = 0 \cdot \big|\big| u \big|\big|_U$. Hence $v \in \text{Dom}(S^*)$ and $S^*[v] = 0$. But then $\big|\big| v \big|\big|_V^2 \leq C\big[\big|\big| T[v] \big|\big|_W^2 + \big|\big| S^*[v] \big|\big|_U^2\big] = 0$, so $v = 0$.

PROOF OF (2): We certainly have

$$R(S^*) \supset \Big\{u \in V \,\Big|\, \big(\exists v \in \text{Dom}(S^*) \cap N(T)\big)\,\big(u = S^*[v]\big)\Big\}.$$

On the other hand, since $N(S^*)^{\perp} = \overline{R(S)}$, we also have

$$V = N(S^*) \oplus \overline{R(S)} = N(S^*) \oplus N(T)$$

by (1). Thus if $v \in \text{Dom}(S^*)$, we can write $v = v_1 + v_2$ with $v_1 \in N(S^*)$ and $v_2 \in \text{Dom}(S^*) \cap N(T)$. But then $S^*[v] = S^*[v_1] + S^*[v_2] = S^*[v_2]$. Thus

$$R(S^*) \subset \Big\{u \in V \,\Big|\, \big(\exists v \in \text{Dom}(S^*) \cap N(T)\big)\,\big(u = S^*[v]\big)\Big\}.$$

PROOF OF (3): For every $v \in \text{Dom}(S^*) \cap N(T) \subset \text{Dom}(S^*) \cap \text{Dom}(T)$, it follows from the estimate in equation (2.1) that $\big|\big| v \big|\big|_V \leq C\big|\big| S^*[v] \big|\big|_U$. Let $\{u_n\}$ be a sequence in the range of S^* which converges to an element $u_0 \in U$. By (2), we can find $v_n \in \text{Dom}(S^*) \cap N(T)$ so that $u_n = S^*[v_n]$, and hence

$$\big|\big| v_m - v_n \big|\big|_V \leq C \,\big|\big| S^*[v_m - v_n] \big|\big|_U = C \,\big|\big| u_m - u_n \big|\big|_U.$$

Since $\{u_n\}$ is a Cauchy sequence, it follows that $\{v_n\}$ is also a Cauchy sequence, and hence converges to a vector $v_0 \in V$. But then since S^* is a closed operator, it follows that $v_0 \in S^*$ and $u_0 = S^*[v_0] \in R(S^*)$. Thus $R(S^*)$ is closed.

PROOF OF (4): We saw in Chapter 5, Lemma 2.6, that if $T : V \to W$ is any closed, densely defined linear operator, then the range of T is closed if and only if the range of T^* is closed.

PROOF OF (5): This follows in the same way as (2). If $v \in \text{Dom}(T)$ we can write $v = v_1 + v_2$ with $v_1 \in N(S^*)$ and $v_2 \in N(T)$. Then $T(v) = T(v_1)$.

PROOF OF (6): The proof that $R(T)$ is closed follows in the same way as (3). If $\{w_n\}$ is a sequence in $R(T)$ which converges to $w_0 \in W$, we can write $w_n = T(v_n)$ with $v_n \in \text{Dom}(T) \cap N(S^*)$. Then $\big|\big| v_m - v_n \big|\big|_V \leq C\big|\big| w_m - w_n \big|\big|_W$, and so $\{v_n\}$ converges to an element $v_0 \in V$. Since T is a closed operator, it follows that $w_0 = T[v_0] \in R(T)$, so $R(T)$ is closed. Then, as in (4), this implies that $R(T^*)$ is closed. This completes the proof. □

In preparation for the definition of the operator N, we observe the following.

PROPOSITION 2.4. *The quadratic form Q is an inner product on H which makes H into a Hilbert space with norm denoted by $\big|\big| \cdot \big|\big|_Q$, and*

$$\big|\big| v \big|\big|_V \leq C \,\big|\big| v \big|\big|_Q$$

for all $v \in H$. Moreover,

$$H = \big(N(T) \cap H\big) \oplus \big(N(S^*) \cap H\big) = H_1 \oplus H_2,$$

where $H_1 = \big(N(T) \cap H\big)$ and $H_2 = \big(N(S^*) \cap H\big)$ are closed subspaces of H with respect to the norm $\{\cdot\}_Q$

PROOF. The basic inequality (2.1) shows that $\big|\big|\,h\,\big|\big|_V \leq C\big|\big|\,h\,\big|\big|_Q$ for all $h \in H$. Let $\{h_n\}$ be a sequence in H which is Cauchy with respect to the norm $\big|\big|\;\cdot\;\big|\big|_Q$. Then

$$C^{-1}\big|\big|\,h_m - h_n\,\big|\big|_V^2 = \big|\big|\,T[h_m] - T[h_m]\,\big|\big|_W^2 + \big|\big|\,S^*[h_m] - S^*[h_n]\,\big|\big|_U^2 = \big|\big|\,h_m - h_n\,\big|\big|_Q^2$$

so $\{h_n\}$ is Cauchy in V, $\{T[h_n]\}$ is Cauchy in W, and $\{S^*[h_n]\}$ is Cauchy in U. Suppose $h_n \to h_0$ in V, $T[h_n] \to w_0$ in W, and $S^*[h_n] \to u_0$ in U. Since T and S^* are closed operators, it follows that $h_0 \in H$. Thus H is complete, and hence is a Hilbert space under the norm $\big|\big|\;\cdot\;\big|\big|_Q$.

Next, since $V = N(T) \oplus N(S^*)$, it follows that if $h \in H$, we can write $h = v_1 + v_2$ where $v_1 \in N(T)$ and $v_2 \in N(S^*)$. In particular, $v_1 \in \mathrm{Dom}(T)$ and $v_2 \in \mathrm{Dom}(S^*)$. It follows that $v_1 = h - v_2 \in \mathrm{Dom}(S^*)$ and $v_2 = h - v_1 \in \mathrm{Dom}(T)$. Thus $v_1, v_2 \in H$, and so we $H = \big(N(T) \cap H\big) + \big(N(S^*) \cap H\big)$. These two spaces are orthogonal with respect to the inner product Q, and hence they are both closed. This completes the proof. □

PROOF OF LEMMA 2.2. We begin with the construction of the operator N. This is done by defining N on the two complementary subspaces $N(T)$ and $N(S^*)$. For any $v_1 \in N(T)$, define a linear functional L_v on H_1 by setting $L_{v_1}[h] = (h, v_1)_V$. Then

$$\big|L_{v_1}[h]\big| = \big|(h, v_1)_V\big| \leq \big|\big|\,v_1\,\big|\big|_V\,\big|\big|\,h\,\big|\big|_V \leq C\,\big|\big|\,v_1\,\big|\big|_V\,\big|\big|\,h\,\big|\big|_Q,$$

and so L_{v_1} is bounded on H_1. By the Riesz representation theorem, there exists a unique $g_1 \in H_1 = \big(N(T) \cap H\big)$ with $\big|\big|\,g_1\,\big|\big|_Q \leq C\,\big|\big|\,v_1\,\big|\big|_V$, so that for all $h \in H_1$ we have

$$(h, v_1) = Q(h, g_1) = \big(S^*[h], S^*[g_1]\big)_U + \big(T[h], T[g_1]]\big)_W = \big(S^*[h], S^*[g_1]\big)_U$$

since $T[h] = T[g_1] = 0$. If we write $g_1 = N_1[v_1]$, it follows that $N_1 : N(T) \to H_1 \subset V$ is a linear transformation, and

$$\big|\big|\,N_1[v_1]\,\big|\big|_V \leq C\,\big|\big|\,N[v_1]\,\big|\big|_Q \leq C^2\big|\big|\,v_1\,\big|\big|_V$$

Thus $N_1 : N(T) \to H$ is bounded, with norm bounded by C, and $N_1 : N(T) \to V$ with norm bounded by C^2.

Similarly, for any $v_2 \in N(S^*)$, define a linear functional M_v on H_2 by setting $M_{v_2}[h] = (h, v_2)_V$. Then

$$\big|M_{v_2}[h]\big| = \big|(h, v_2)_V\big| \leq \big|\big|\,v_2\,\big|\big|_V\,\big|\big|\,h\,\big|\big|_V \leq C\,\big|\big|\,v_2\,\big|\big|_V\,\big|\big|\,h\,\big|\big|_Q,$$

and so M_{v_1} is bounded on H_2. By the Riesz representation theorem, there exists a unique $g_2 \in H_1 = \big(N(S^*) \cap H\big)$ with $\big|\big|\,g_2\,\big|\big|_Q \leq C\,\big|\big|\,v_2\,\big|\big|_V$ so that for all $h \in H_2$ we have

$$(h, v_2) = Q(h, g_2) = \big(S^*[h], S^*[g_2]\big)_U + \big(T[h], T[g_2]]\big)_W = \big(T[h], T[g_2]]\big)_W$$

since $S^*[h] = S^*[g_2] = 0$. If we write $g_2 = N_2[v_2]$, it follows that $N_2 : N(T) \to H_1 \subset V$ is a linear transformation, and

$$\big|\big|\,N_2[v_2]\,\big|\big|_V \leq C\,\big|\big|\,N[v_2]\,\big|\big|_Q \leq C^2\big|\big|\,v_2\,\big|\big|_V$$

Thus $N_2 : N(T) \to H$ is bounded, with norm bounded by C, and $N_2 : N(T) \to V$ with norm bounded by C^2.

We can now put the two operators together. For $v \in V$, write $v = v_1 + v_2$ with $v_1 \in N(T)$ and $v_2 \in N(S^*)$. Define

$$N(v) = N_1(v_1) + N_2(v_2) \in H_1 + H_2 = H.$$

Then $N : V \to H \subset V$ is a linear mapping. Since H_1 and H_2 are orthogonal, we have

$$\begin{aligned}||\,N(v)\,||_V^2 = ||\,N_1[v_1]\,||_V^2 + ||\,N_1[v_1]\,||_V^2 &\leq C^2 \left[||\,N_1[v_1]\,||_Q^2 + ||\,N_1[v_1]\,||_Q^2\right] \\ &\leq C^4 \left[||\,v_1\,||_V^2 + ||\,v_2\,||_V^2\right] = C^4 ||\,v\,||_V^2,\end{aligned}$$

and so

$$||\,N[v]\,||_V \leq C^2\,||\,v\,||_V,$$

which establishes statement (1) of Lemma 2.2.

Let $h = h_1 + h_2 \in H_1 + H_2 = H$ and $v = v_1 + v_2 \in N(T) + N(S^*) = V$. Then since $S^*[h_2] = 0 \in U$ and $T[h_1] = 0 \in W$, we have

$$\begin{aligned}(h, v)_V &= (h_1, v_1)_V + (h_2, v_2)_v \\ &= (S^*[h_1], S^*[N_1[v_1]])_U + (T[h_2], T[N_2[v_2]])_W \\ &= (S^*[h], S^*[N[v]])_U + (T[h], T[N[v]])_W \\ &= Q\big(h, N[h]\big).\end{aligned} \tag{2.2}$$

This establishes statement (2).

We next show that if $v \in V$, we have $S^*[N[v]] \in \operatorname{Dom}(S)$ and $T[N[v]] \in \operatorname{Dom}(T^*)$. Since $(S^*)^* = S$, it follows that in order to show $S^*[N[v]] \in \operatorname{Dom}(S)$, we need to show that for every $h \in \operatorname{Dom}(S^*)$ we have

$$\big|(S^*[h], S^*[N[v]])_U\big| \leq C\,||\,h\,||_V.$$

If $h \in \operatorname{Dom}(S^*) \subset V$, write $h = h_1 + h_2$ where $h_1 \in N(T)$ and $h_2 \in N(S^*)$. It follows that $h_1 \in N(T) \cap \operatorname{Dom}(S^*) \subset H$. Thus $S^*[h] = S^*[h_1] + S^*[h_2] = S^*[h_1]$. Next, write $N[v] = N_1[v_1] + N_2[v_2]$ where $N_1[v_1] \in H_1 = N(T) \cap H$ and $N_2[v_2] \in N(S^*)$. Thus $S^*[N[v]] = S^*[N_1[v_1]]$. But then we have

$$\begin{aligned}\big|(S^*[h], S^*[N[v]])_U &= \big|(S^*[h_1], S^*[N_1[v_1]])_U\big| = \big|(h_1, v_1)_V\big| \\ &\leq ||\,h_1\,||_V\,||\,v_1\,||_V \leq C\,||\,h_1\,||_V\,||\,v_1\,||_Q \leq C\,||\,h\,||_V\,||\,v_1\,||_Q.\end{aligned}$$

The proof that $T[N[v]] \in \operatorname{Dom}(T^*)$ is similar. Let $h \in \operatorname{Dom}(T)$, and write $h = h_1 + h_2$ as before. It follows that $h_2 \in N(S^*) \cap \operatorname{Dom}(T) \subset H$. Then $T[h] = T[h_2]$, and if we decompose $N[v] = N_1[v_1] + N_2[v_2]$ as before, we have $T[N[v]] = T[N_2[v_2]]$. Thus

$$\begin{aligned}\big|(T[h], T[N[v]])_U &= \big|(T[h_2], T[N_2[v_2]])_U\big| = \big|(h_2, v_2)_V\big| \\ &\leq ||\,h_2\,||_V\,||\,v_2\,||_V \leq C\,||\,h_2\,||_V\,||\,v_2\,||_Q \leq C\,||\,h\,||_V\,||\,v_2\,||_Q.\end{aligned}$$

Thus we have shown that $N[v] \in \operatorname{Dom}(L)$, and this is assertion (3).

If $v \in V$ and $N[v] = 0$, it follows from assertion (2) that

$$(h, v)_V = Q\big(h, N[v]\big) = Q(h, 0) = 0$$

for all $h \in H$. Since H is dense in V, this implies that $v = 0$, so the operator N is one-to-one. Suppose that $v \in \text{Dom}(L)$ and $L[v] = 0$. Then

$$\begin{aligned} 0 &= (v, L[v])_V = (v, S\,S^*[v])_V + (v, T^*\,T[v])_V \\ &= (S^*[v], S^*[v])_U + (T[v], T[v])_W = \big|\big|\, S^*[v]\,\big|\big|_U^2 + \big|\big|\, T[v]\,\big|\big|_W^2. \end{aligned}$$

It follows follows from the basic inequality that $\big|\big|\, v\,\big|\big|_V = 0$, so the operator L is also one-to-one. This establishes assertion (4).

Now returning to equation (2.3), it now follows from (3) that if $v \in V$ then

$$\begin{aligned} (h, v)_V &= (S^*[h], S^*[N[v]])_U + (T[h], T[N[v]])_W \\ &= \big(h, L\big[N[v]\big]\big) \end{aligned}$$

for every $h \in H$. Since H is dense in V, it follows that $L\big[N[v]\big] = v$ for all $v \in V$. Finally, if $v \in \text{Dom}(L)$, then $L[v] \in V$, and so $N\big[L[v]\big] \in \text{dom}(L)$. But then

$$L\Big[N\big[L[v]\big]\Big] = L[v].$$

Since L is one-to-one, it follows that $v = N\big[L[v]\big]$, and this establishes assertion (5).

Now let $v_1, v_2 \in V$. Then $v_2 = L\big[N[v_2]\big]$, so we have

$$\begin{aligned} \big(N[v_1], v_2\big)_V &= \big(N[v_1], L\big[N[v_2]\big]\big)_V \\ &= \big(S^*\big[N[v_1]\big], S^*\big[N[v_2]\big]\big)_U + \big(T\big[N[v_1]\big], T\big[N[v_2]\big)_W \\ &= \big(L\big[N[v_1]\big], N[v_2]\big)_V \\ &= \big(v_1, N[v_2]\big)_V. \end{aligned}$$

Thus N is self-adjoint. Moreover, since N is one-to-one, N has dense range. Since L is the inverse to N, it follows that L is densely defined and is self-adjoint. This establishes (6).

If $v \in N(T)$, the same argument as in Proposition 2.1 shows that $u = S^*N[v]$ is the unique solution to $S[u] = v$. This completes the proof. □

2.3. Notation.

If we try to imitate the finite dimensional situation, we want to replace the operator S by the mapping $\overline{\partial}_{q-1}$ and the operator T by the mapping $\overline{\partial}_q$. We also want to replace U, V, and W by certain Hilbert spaces of differential forms. Thus let $\Omega \subset \mathbb{C}^n$ be a domain with $\mathcal{C}^2$-boundary. Thus we assume there is an open neighborhood U of the boundary $\partial\Omega$ and a function $\rho : U \to \mathbb{R}$ of class $\mathcal{C}^2$ such that

$$\Omega \cap U = \Big\{ z \in U \,\Big|\, \rho(z) < 0 \Big\}$$

and $|\nabla\rho(z)| \neq 0$ for all $z \in U$. In fact, we shall assume that

$$\big|\nabla\rho(z)\big| = 1$$

for all $z \in \partial\Omega$.

We let dm denote Lebesgue measure on Ω. It will be important to consider certain weighted spaces. Thus if $\varphi : \overline{\Omega} \to \mathbb{R}$ is continuous, we let L_φ denote the space of (equivalence classes) of complex-valued measurable functions f such that

$$\big|\big| f \big|\big|_\varphi^2 = \int_\Omega |f(z)|^2 \, e^{-\varphi(z)} \, dm(z) < \infty.$$

L_φ is a Hilbert space with inner product

$$(f, g)_\varphi = \int_\Omega f(z) \, \overline{g(z)} \, e^{-\varphi(z)} \, dm(z).$$

In order to put a Hilbert space norm on differential forms, we need to introduce a metric on the complexified tangent space. We shall use the Riemannian metric on $\Omega \subset \mathbb{C}^n \cong \mathbb{R}^{2n}$, in which

$$\Big\{ \frac{1}{\sqrt{2}} \partial_{x_1}, \ldots, \frac{1}{\sqrt{2}} \partial_{x_n}, \frac{1}{\sqrt{2}} \partial_{y_1}, \ldots, \frac{1}{\sqrt{2}} \partial_{y_n} \Big\}$$

is an orthonormal basis for T_z, and the dual basis

$$\big\{ \sqrt{2} dx_1, \ldots, \sqrt{2} dx_n, \sqrt{2} dy_1, \ldots, \sqrt{2} dy_n \big\}$$

is an orthonormal basis for $T_z^* = \Lambda_z^1$. If we extend this metric to an Hermitian metric $< \cdot, \cdot >$ on $\mathbb{C}\Lambda_z^1$, we have

$$\begin{aligned}
&< dz_j, dz_k > = < dx_j + i dy_j, dx_k + i dy_k > = < dx_j, dx_k > + < dy_j, dy_k > = \delta_{j,k} \\
&< d\bar{z}_j, d\bar{z}_k > = < dx_j - i dy_j, dx_k - i dy_k > = < dx_j, dx_k > + < dy_j, dy_k > = \delta_{j,k} \\
&< dz_j, d\bar{z}_k > = < dx_j + i dy_j, dx_k - i dy_k > = < dx_j, dx_k > - < dy_j, dy_k > = 0
\end{aligned}$$

Thus the elements

$$\{ dz_1, \ldots, dz_n, d\bar{z}_1, \ldots, d\bar{z}_n \}$$

form an orthonormal basis for $\mathbb{C}\Lambda_z^1$. This metric then naturally extends to the spaces $\Lambda_z^{p,q}$, so that the elements $\{dz_J \wedge d\bar{z}_K\}$ for $J \in \mathbb{I}_p^*$ and $K \in \mathbb{I}_q^*$ form an orthonormal basis. In particular, if $\omega = \sum \omega_{J,K} \, dz_J \wedge d\bar{z}_K \in \Lambda_z^{p,q}$, then

$$\big|\big| \omega \big|\big|_z^2 = \big\langle \omega, \omega \big\rangle_z = \sum_{J \in \mathbb{I}_p^*, K \in \mathbb{I}_q^*} \big| \omega_{J,K} \big|^2.$$

The inner product on Λ_z^m allows us to introduce the notion of contraction of two differential forms. This is the dual of the wedge product. Given $\alpha \in \mathbb{C}\Lambda_z^k$ and $\beta \in \mathbb{C}\Lambda_z^j$ with $j \leq k$, we define $\alpha \vee \beta \in \mathbb{C}\Lambda_z^{k-j}$ by requiring that

$$\big(\alpha \vee \beta, \gamma \big)_z = \big(\alpha, \beta \wedge \gamma \big)_z \tag{2.3}$$

for all $\gamma \in \mathbb{C}\Lambda_z^{k-j}$. Suppose that

$$\alpha = \sum_{K \in \mathbb{I}_k^*} \alpha_K \, d\bar{z}_k, \qquad \beta = \sum_{J \in \mathbb{I}_j^*} \beta_J \, d\bar{z}_J, \qquad \text{and} \qquad \gamma = \sum_{L \in \mathbb{I}_{k-j}^*} \gamma_L \, d\bar{z}_L.$$

Then

$$\big(\alpha, \beta \wedge \gamma \big)_z = \sum_{J \in \mathbb{I}_j^*} \sum_{K \in \mathbb{I}_k^*} \sum_{L \in \mathbb{I}_{k-j}^*} \epsilon_K^{(J,L)} \alpha_K \, \overline{\beta}_J \, \overline{\gamma}_L.$$

Thus

$$\beta \vee \alpha = \sum_{L \in \mathbb{I}_{k-j}^*} \Big[\sum_{J \in \mathbb{I}_j^*} \sum_{K \in \mathbb{I}_k^*} \epsilon_K^{(J,L)} \alpha_K \, \overline{\beta}_J \Big] \, d\bar{z}_L.$$

In particular, if ρ is a real valued function and if $\omega = \sum_{K\in\mathbb{I}_q^*} \omega_K \, d\bar{z}_K \in \Lambda_z^{0,q}$, then

$$\begin{aligned}\overline{\partial}\rho \vee \omega &= \sum_{J\in\mathbb{I}_{q-1}^*} \Big[\sum_{K\in\mathbb{I}_q^*}\sum_{j=1}^{n} \epsilon_K^{(j,J)} \frac{\partial\rho}{\partial z_j}\,\omega_K\Big]\, d\bar{z}_J \\ &= \sum_{J\in\mathbb{I}_{q-1}^*} \Big[\sum_{j=1}^{n} \epsilon_{[(j,J)]}^{(j,J)} \frac{\partial\rho}{\partial z_j}\,\omega_{[(j,J)]}\Big]\, d\bar{z}_J \\ &= \sum_{J\in\mathbb{I}_{q-1}^*} \Big[\sum_{j=1}^{n} \frac{\partial\rho}{\partial z_j}\,\omega_{(j,J)}\Big]\, d\bar{z}_J.\end{aligned} \tag{2.4}$$

We shall use the following general notation. If B is a space of functions or distributions on Ω or its closure $\overline{\Omega}$, we let $B^{p,q}$ denote the space of (p,q)-forms with coefficients in B. In particular:

(1) $\mathcal{D}(\Omega)^{p,q}$ denotes the space of (p,q)-forms with coefficients which are infinitely differentiable and compactly supported in Ω;

(2) $\mathcal{E}(\Omega)^{p,q}$ denotes the space of forms which are infinitely differentiable on Ω;

(3) $\mathcal{E}(\overline{\Omega})^{p,q}$ denotes the space of forms which are restrictions to $\overline{\Omega}$ of smooth forms on $\mathbb{C}^n$;

(4) $\mathcal{E}(\overline{\Omega})_0^{p,q}$ denotes the space of forms on $\overline{\Omega}$ which are restrictions of forms compactly supported in $\mathbb{C}^n$ (but not necessarily compactly supported in Ω. Note that if Ω is bounded, then $\mathcal{E}(\overline{\Omega})^{p,q} = \mathcal{E}(\overline{\Omega})_0^{p,q}$.

(5) $L_\varphi^{p,q}(\Omega)$ denotes the space of (p,q)-forms with coefficients in L_φ, and is itself a Hilbert space. If $\omega, \eta \in L_\varphi^{p,q}(\Omega)$ with

$$\omega = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \omega_{J,K}\, dz_J \wedge d\bar{z}_K, \qquad \text{and} \qquad \eta = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \eta_{J,K}\, dz_J \wedge d\bar{z}_K,$$

then

$$\big(\omega,\eta\big)_\varphi = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \int_\Omega \omega_{J,K}(z)\,\overline{\eta_{J,K}(z)}\, e^{-\varphi(z)}\, dm(z).$$

If $\varphi = 0$, we will simply write $L^{p,q}(\Omega)$ instead of $L_0^{p,q}(\Omega)$ for the space of (p,q)-forms with coefficients which are square integrable on Ω.

We will also need to consider inner products taken on the boundary $\partial\Omega$ of Ω. Let σ denote surface area measure. If $\omega, \eta \in \mathcal{E}(\overline{\Omega})_0^{p,q}$ with

$$\omega = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \omega_{J,K}\, dz_J \wedge d\bar{z}_K, \qquad \text{and} \qquad \eta = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \eta_{J,K}\, dz_J \wedge d\bar{z}_K,$$

we put

$$\big[\omega,\eta\big]_\varphi = \sum_{J\in\mathbb{I}_p^*, K\in\mathbb{I}_q^*} \int_{\partial\Omega} \omega_{J,K}(z)\,\overline{\eta_{J,K}(z)}\, e^{-\varphi(z)}\, d\sigma(z).$$

2.4. Computation of $\overline{\partial}$ and its formal adjoint.

From now on, we will take $p = 0$, since it plays no role for domains in $\mathbb{C}^n$. Recall that if $K \in \mathbb{I}_q^*$ and $1 \leq k \leq n$, we have $d\bar{z}_k \wedge d\bar{z}_K = \sum_{L\in\mathbb{I}_{q+1}^*} \epsilon_L^{(k,K)}\, d\bar{z}_L$. Thus we have the following formulae for $\overline{\partial}[\omega]$.

Proposition 2.5. *If* $\omega = \sum_{K\in\mathbb{I}_q^*} \omega_K\, d\bar{z}_K \in \mathcal{E}(\Omega)^{0,q}$ *then*

$$\overline{\partial}[\omega] = \sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^n \frac{\partial\omega_K}{\partial\bar{z}_k}\, d\bar{z}_k \wedge d\bar{z}_K = \sum_{L\in\mathbb{I}_{q+1}^*}\Big[\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^n \epsilon_L^{(k,K)}\,\frac{\partial\omega_K}{\partial\bar{z}_k}\Big]\, d\bar{z}_L.$$

The operator $\overline{\partial}$ has a formal adjoint, obtained by formal integration by parts. We compute this, using the following consequence of the divergence theorem.

Proposition 2.6. *Suppose* Ω *has a defining function* ρ *which is of class* $\mathcal{C}^2$ *such that* $|\nabla\rho(z)| = 1$ *for* $z \in \partial\Omega$*. If* $f \in \mathcal{E}(\overline{\Omega})_0$*, we have*

$$\int_\Omega \frac{\partial f}{\partial z_j}(z)\, dm(z) = \int_{\partial\Omega} f(z)\,\frac{\partial\rho}{\partial z_j}(z)\, d\sigma(z).$$

If $\eta \in \mathcal{E}(\overline{\Omega})_0^{0,q-1}$ and $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$, we shall want to integrate by parts in the inner product $(\overline{\partial}_{q-1}[\eta], \omega)_\psi$, and move the differentiation from η to ω. This will lead to a differential operator on Ω.

Definition 2.7. *The* formal adjoint *of* $\overline{\partial}_{q-1}$ *is the mapping* $\vartheta_q : \mathcal{E}(\Omega)^{0,q} \to \mathcal{E}(\Omega)^{0,q-1}$ *given by*

$$\begin{aligned}
\vartheta_q\Big[\sum_{K\in\mathbb{I}_q^*}\omega_K\, d\bar{z}_K\Big] &= -\sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{K\in\mathbb{I}_q^*}\sum_{j=1}^n \epsilon_K^{(j,J)}\, e^\lambda\,\frac{\partial}{\partial z_j}\big[e^{-\psi}\,\omega_K\big]\Big]\, d\bar{z}_J\\
&= -\sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{j=1}^n \epsilon_{[(j,J)]}^{(j,J)}\, e^\lambda\,\frac{\partial}{\partial z_j}\big[e^{-\psi}\,\omega_{[(j,J)]}\big]\Big]\, d\bar{z}_J\\
&= -\sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{j=1}^n e^\lambda\,\frac{\partial}{\partial z_j}\big[e^{-\psi}\,\omega_{(j,J)}\big]\Big]\, d\bar{z}_J \qquad .
\end{aligned}$$

The last equality follows from our notation so that $\epsilon_{[(j,J)]}^{(j,J)}\,\omega_{[(j,J)]} = \omega_{(j,J)}$.

Integration by parts also leads to a boundary integral, which in our case will involve $\overline{\partial}\rho \vee \omega$. Recall from equation (2.3) that we can write

$$\overline{\partial}\rho \vee \omega = \sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{j=1}^n \frac{\partial\rho}{\partial z_j}(z)\,\omega_{(j,J)}(z)\Big]\, d\bar{z}_J.$$

Proposition 2.8. *Let* ρ *be the defining function of the domain* Ω*, and assume that* $|\nabla\rho(z)| = 1$ *for* $z \in \partial\Omega$*.*

(1) If $\eta \in \mathcal{E}(\overline{\Omega})_0^{0,q-1}$ *and* $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$*. Then*

$$\big(\overline{\partial}_{q-1}[\eta], \omega\big)_\psi = \big(\eta, \vartheta_q[\omega]\big)_\lambda + \big[\eta, \overline{\partial}\rho \vee \eta\big]_\psi.$$

(2) *If $\eta \in \mathcal{D}(\Omega)_0^{0,q-1}$ (so η has compact support in Ω), and $\omega \in \mathcal{E}(\Omega)_0^{0,q}$ (so there is no assumption about the behavior of ω at the boundary), then*
$$\big(\overline{\partial}_{q-1}[\eta],\omega\big)_{\psi} = \big(\eta,\vartheta_q[\omega]\big)_{\lambda}.$$

(3) *If $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$ then $\big(\overline{\partial}_{q-1}[\eta],\omega\big)_{\psi} = \big(\eta,\overline{\partial}\rho\vee\omega\big)_{\lambda}$ for every $\eta \in \mathcal{E}(\overline{\Omega})_0^{0,q-1}$ if and only if for all $z \in \partial\Omega$ we have $\overline{\partial}\rho\vee\omega(z) = 0$.*

PROOF. We have
$$(\overline{\partial}[\eta],\omega)_{\psi} = \sum_{K\in\mathbb{I}_q^*}\sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j=1}^{n}\epsilon_K^{(j,J)}\int_{\Omega}\frac{\partial\eta_J}{\partial\bar z_j}(z)\,\overline{\omega_K(z)}\,e^{-\psi(z)}\,dm(z).$$
Integration by parts shows that this equals
$$\begin{aligned}
&-\sum_{J\in\mathbb{I}_{q-1}^*}\sum_{K\in\mathbb{I}_q^*}\sum_{j=1}^{n}\epsilon_K^{(j,J)}\int_{\Omega}\eta_J(z)\,\overline{e^{\lambda}\frac{\partial}{\partial z_j}\big[e^{-\psi}\,\omega_K\big](z)}\,e^{-\lambda(z)}\,dm(z)\\
&\qquad\qquad+\sum_{K\in\mathbb{I}_q^*}\int_{\partial\Omega}\eta_J(z)\,\overline{\sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j=1}^{n}\epsilon_K^{(j,J)}\frac{\partial\rho}{\partial z_j}(z)\,\omega_K(z)\,e^{-\psi(z)}}\,dA(z)\\
&=\Big(\eta,\vartheta_q[\omega]\Big)_{\lambda}+\Big[\eta,\overline{\partial}\rho\vee\omega\Big]_{\psi}.
\end{aligned}$$
This gives assertion (1), and assertions (2) and (3) then follow easily. □

2.5. The basic identity.

If $\omega = \sum_{K\in\mathbb{I}_q^*}\omega_K\,d\bar z_K \in \mathcal{E}(\overline{\Omega})_0^{0,q}$, we want to compute the quadratic form
$$Q_q(\omega,\omega) = \big|\big|\,\overline{\partial}_q[\omega]\,\big|\big|_{\varphi}^2 + \big|\big|\,\vartheta_q[\omega]\,\big|\big|_{\lambda}^2.$$
We first consider the case when $\lambda = \phi = \varphi$.

THEOREM 2.9. *Suppose that $\lambda = \phi = \varphi$ and $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$. If $\overline{\partial}\rho\vee\omega = 0$ on $\partial\Omega$, then*
$$\begin{aligned}
&\big|\big|\,\overline{\partial}_q[\omega]\,\big|\big|_{\varphi}^2 + \big|\big|\,\vartheta_q[\omega]\,\big|\big|_{\varphi}^2\\
&\qquad=\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar z_k}\,\Big|\Big|_{\varphi}^2\\
&\qquad\qquad+\sum_{J\in\mathbb{I}_{q-1}^*}\int_{\Omega}\sum_{j,k=1}^{n}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar z_k}(z)\,dm(z)\\
&\qquad\qquad\qquad+\sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^{n}\int_{\partial\Omega}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\rho}{\partial z_j\partial\bar z_k}(z)\,d\sigma(z)
\end{aligned}$$

If the weights are not all the same, we do not have such a nice identity, but if $2\phi = \varphi + \lambda$ we do have an inequality. Define
$$B_q\Big[\sum_{K\in\mathbb{I}_q^*}\omega_K\,d\bar z_K\Big] = \sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{j=1}^{n}\frac{\partial(\varphi-\psi)}{\partial z_j}\,\omega_{(j,J)}\Big]\,d\bar z_J.$$
Thus $B_q : \Lambda^{0,q}(\Omega)\to\Lambda^{0,q-1}(\Omega)$ is a multiplication operator.

THEOREM 2.10. *Suppose that* $2\phi = \varphi + \lambda$ *and* $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$. *If* $\overline{\partial}\rho \vee \omega = 0$ *on* $\partial\Omega$, *then*

$$\begin{aligned}\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2 &+ \sum_{J\in\mathbb{I}_{q-1}^*}\int_{\Omega}\sum_{j,k=1}^{n}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar{z}_k}(z)\,e^{-\varphi(z)}\,dm(z)\\ &\qquad + \sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^{n}\int_{\partial\Omega}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\rho}{\partial z_j\partial\bar{z}_k}(z)\,e^{-\varphi(z)}\,d\sigma(z).\\ &\leq \Big|\Big|\,\overline{\partial}_q[\omega]\,\Big|\Big|_{\varphi}^2 + 2\Big|\Big|\,\vartheta_q[\omega]\,\Big|\Big|_{\lambda}^2\\ &\qquad + \sum_{J\in\mathbb{I}_{q-1}^*}\int_{\Omega}\Big|\sum_{j=1}^{n}\frac{\partial(\varphi-\psi)}{\partial z_j}(z)\,\omega_{(j,J)}(z)\Big|^2\,e^{-\varphi(z)}\,dm(z)\end{aligned}$$

The proofs of these results is a long and ingenious calculation, which probably should not be attempted in public. We begin with an identity for the term $\big|\big|\,\overline{\partial}[\varphi]\,\big|\big|_{\varphi}^2$.

PROPOSITION 2.11. *Let* $\omega = \sum_{K\in\mathbb{I}_q^*}\omega_K\,d\bar{z}_K$. *Then*

$$\Big|\Big|\,\overline{\partial}[\omega]\,\Big|\Big|_{\varphi}^2 = \sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2 - \sum_{L\in\mathbb{I}_{q-1}^*}\sum_{k_1,k_2=1}^{n}\Big(\frac{\partial\omega_{(k_2,L)}}{\partial\bar{z}_{k_1}},\frac{\partial\omega_{(k_1,L)}}{\partial\bar{z}_{k_2}}\Big)_{\varphi}.$$

PROOF. We have $\overline{\partial}[\omega] = \displaystyle\sum_{L\in\mathbb{I}_{q+1}^*}\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\epsilon_L^{(k,K)}\frac{\partial\omega_K}{\partial\bar{z}_k}d\bar{z}_L$. Thus

$$\begin{aligned}\Big|\Big|\,\overline{\partial}[\omega]\,\Big|\Big|_{\varphi}^2 &= \sum_{L\in\mathbb{I}_{q+1}^*}\Big|\Big|\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\epsilon_L^{(k,K)}\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2\\ &= \sum_{L\in\mathbb{I}_{q+1}^*}\sum_{K_1,K_2\in\mathbb{I}_q^*}\sum_{k_1,k_2=1}^{n}\epsilon_L^{(k_1,K_1)}\,\epsilon_L^{(k_2,K_2)}\Big(\frac{\partial\omega_{K_1}}{\partial\bar{z}_{k_1}},\frac{\partial\omega_{K_2}}{\partial\bar{z}_{k_2}}\Big)_{\varphi}.\end{aligned}$$

If the product $\epsilon_L^{(k_1,K_1)}\,\epsilon_L^{(k_2,K_2)} \neq 0$, there are two possibilities.

(a) We can have $k_1 = k_2 = k$, $K_1 = K_2 = K$, and $k \notin K$. In this case $\epsilon_L^{k_1,K_1} = \epsilon_L^{k_2,K_2}$ so $\epsilon_L^{k_1,K_1}\,\epsilon_L^{k_2,K_2} = 1$.

(b) We can have $k_1 \neq k_2$ and $K_1 = [(k_2,J)]$, $K_2 = [(k_1,J)]$ for some $J \in \mathbb{I}_{q-1}^*$. In this case $\epsilon_L^{(k_1,K_1)} = \epsilon_L^{(k_1,[k_2,J])}$ and $\epsilon_L^{(k_2,K_2)} = \epsilon_L^{(k_2,[k_1,J])}$, so

$$\epsilon_L^{(k_1,[k_2,J])}\,\epsilon_L^{(k_2,[k_1,J])} = -\epsilon_{[(k_1,J)]}^{(k_1,J)}\,\epsilon_{[(k_2,J)]}^{(k_2,J)}.$$

The terms for which possibility (a) holds gives rise to the sum

$$\sum_{K\in\mathbb{I}_q^*}\sum_{k\notin K}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2.$$

The terms for which possibility (b) hold give the sum

$$-\sum_{J\in\mathbb{I}^*_{q-1}}\sum_{\substack{j_1\neq j_2\\ j_1,j_2\notin J}}\epsilon^{(j_1,J)}_{[(j_1,J)]}\,\epsilon^{(j_2,J)}_{[(j_2,J)]}\Big(\frac{\partial\omega_{[(j_2,J)]}}{\partial\bar z_{j_1}},\frac{\partial\omega_{[(j_1,J)]}}{\partial\bar z_{j_2}}\Big)_\varphi$$
$$=-\sum_{J\in\mathbb{I}^*_{q-1}}\sum_{\substack{j_1\neq j_2\\ j_1,j_2\notin J}}\Big(\frac{\partial\omega_{(j_2,J)}}{\partial\bar z_{j_1}},\frac{\partial\omega_{(j_1,J)}}{\partial\bar z_{j_2}}\Big)_\varphi.$$

We rearrange the inner sum as follows:

$$-\sum_{\substack{j_1\neq j_2\\ j_1,j_2\notin J}}\Big(\frac{\partial\omega_{(j_2,J)}}{\partial\bar z_{j_1}},\frac{\partial\omega_{(j_1,J)}}{\partial\bar z_{j_2}}\Big)_\varphi$$
$$=\Bigg(-\sum_{j_1,j_2=1}^{n}+\sum_{j_1=j_2\notin J}+\sum_{j_1,j_2\in J}+\sum_{\substack{j_1\in J\\ j_2\notin J}}+\sum_{\substack{j_2\in J\\ j_1\notin J}}\Bigg)\Big(\frac{\partial\omega_{(j_2,J)}}{\partial\bar z_{j_1}},\frac{\partial\omega_{(j_1,J)}}{\partial\bar z_{j_2}}\Big)_\varphi.$$

Now if $j_1\in J$ then $\omega_{(j_1,J)}=0$, and if $j_2\in J$ then $\omega_{(j_2,J)}=0$, so the last three summations are zero. Thus we have

$$-\sum_{J\in\mathbb{I}^*_{q-1}}\sum_{\substack{j_1\neq j_2\\ j_1,j_2\notin J}}\Big(\frac{\partial\omega_{(j_2,J)}}{\partial\bar z_{j_1}},\frac{\partial\omega_{(j_1,J)}}{\partial\bar z_{j_2}}\Big)_\varphi$$
$$=-\sum_{J\in\mathbb{I}^*_{q-1}}\sum_{j_1,j_2=1}^{n}\Big(\frac{\partial\omega_{(j_2,J)}}{\partial\bar z_{j_1}},\frac{\partial\omega_{(j_1,J)}}{\partial\bar z_{j_2}}\Big)_\varphi+\sum_{k\in K}\Big|\Big|\frac{\partial\omega_K}{\partial\bar z_k}\Big|\Big|_\varphi.$$

Putting cases (a) and (b) together, we obtain the stated equality. □

We next turn to the term $\big|\big|\,\vartheta_q[\omega]\,\big|\big|_\lambda$. Define an operator D_k by setting

$$D_k[f]=e^{\varphi}\,\frac{\partial}{\partial z_k}\big[e^{-\varphi}\,f\big]=\frac{\partial f}{\partial z_k}-\frac{\partial\varphi}{\partial z_k}\,f.$$

PROPOSITION 2.12.

$$\big(D_j[f],D_k[g]\big)_\varphi=\Big(\frac{\partial f}{\partial\bar z_k},\frac{\partial g}{\partial\bar z_j}\Big)_\varphi+\Big(f\,\frac{\partial^2\varphi}{\partial z_j\partial\bar z_k},g\Big)_\varphi$$
$$+\Big[f\,\frac{\partial\rho}{\partial z_j},D_k[g]\Big]_\varphi-\Big[f\,\frac{\partial\rho}{\partial\bar z_k},\frac{\partial g}{\partial\bar z_j}\Big]_\varphi.$$

PROOF. Integration by parts shows that

$$\big(D_j[f],g\big)_\varphi=\Big(f,\frac{\partial g}{\partial\bar z_j}\Big)_\varphi+\Big[f\,\frac{\partial\rho}{\partial z_j},g\Big]_\varphi\qquad\text{and}$$
$$\big(f,D_k[g]\big)_\varphi=\Big(\frac{\partial g}{\partial\bar z_k},g\Big)_\varphi+\Big[f\,\frac{\partial\rho}{\partial\bar z_k},g\Big]_{\varphi 0.}$$

Thus

$$\begin{aligned}\big(D_j[f], D_k[g]\big)_\varphi &= -\Big(f, \frac{\partial}{\partial \bar z_j}\big[D_k[g]\big]\Big)_\varphi + \Big[f\frac{\partial \rho}{\partial z_j}, D_k[g]\Big]_\varphi \\ &= -\Big(f, D_k\big[\frac{\partial g}{\partial \bar z_k}\big]\Big)_\varphi + \Big(f, \Big[\frac{\partial}{\partial \bar z_j}, D_k\Big][g]\Big)_\varphi + \big[f\frac{\partial \rho}{\partial z_j}, D_k[g]\big]_\varphi \\ &= \Big(\frac{\partial f}{\partial \bar z_k}, \frac{\partial g}{\partial \bar z_j}\Big)_\varphi + \Big(f, \Big[\frac{\partial}{\partial \bar z_j}, D_k\Big][g]\Big)_\varphi \\ &\qquad + \Big[f\frac{\partial \rho}{\partial z_j}, D_k[g]\Big]_\varphi - \Big[f\,\frac{\partial \rho}{\partial \bar z_k}, \frac{\partial g}{\partial \bar z_j}\Big]_\varphi\end{aligned}$$

But it is easy to check that

$$\Big[\frac{\partial}{\partial \bar z_j}, D_k\Big][g] = \frac{\partial^2 \varphi}{\partial \bar z_j \partial z_k}\, g,$$

and this completes the proof. □

Now

$$e^\lambda\,\frac{\partial}{\partial z_k}\big[e^{-\psi}\, f\big] = e^{(\lambda-\psi)}\,\Big[D_k[f] + \frac{\partial(\varphi-\psi)}{\partial z_k}\, f\Big],$$

and so we can write

$$\begin{aligned}\vartheta_q[\omega] &= \sum_{J\in\mathbb{I}^*_{q-1}}\Big[\sum_{K\in\mathbb{I}^*_q}\sum_{k=1}^n \epsilon_K^{k,J}\, e^\lambda\frac{\partial}{\partial z_k}\big[e^{-\psi}\,\omega_K\big]\Big]\,d\bar z_J \\ &= e^{(\lambda-\psi)}\sum_{J\in\mathbb{I}^*_{q-1}}\Big[\sum_{K\in\mathbb{I}^*_q}\sum_{k=1}^n \epsilon_K^{k,J}\, D_k[\omega_K]\Big]\,d\bar z_J \\ &\qquad + e^{(\lambda-\psi)}\sum_{J\in\mathbb{I}^*_{q-1}}\Big[\sum_{K\in\mathbb{I}^*_q}\sum_{k=1}^n \epsilon_K^{k,J}\,\frac{\partial(\varphi-\psi)}{\partial z_k}\,\omega_K\Big]\,d\bar z_J \\ &= e^{(\lambda-\psi)}\,A_q[\omega] + e^{(\lambda-\psi)}\,B_q[\omega].\end{aligned}$$

There are two cases of particular interest. The first is when $\lambda = \phi = \varphi$, in which case $B_q[\omega] = 0$. The second is if $2\psi = \lambda + \varphi$, in which case we have

$$\begin{aligned}\big|\big|\,e^{(\lambda-\psi)}\,A_q[\omega]\,\big|\big|_\lambda &= \big|\big|\,A_q[\omega]\,\big|\big|_\varphi \\ \big|\big|\,e^{(\lambda-\psi)}\,B_q[\omega]\,\big|\big|_\lambda &= \big|\big|\,B_q[\omega]\,\big|\big|_\varphi.\end{aligned}$$

Thus we have

$$\begin{aligned}&\big|\big|\,A_q[\omega]\,\big|\big|^2_\varphi = \big|\big|\,\vartheta_q[\omega]\,\big|\big|^2_\lambda \qquad \text{in the first case, and} \\ &\big|\big|\,A_q[\omega]\,\big|\big|^2_\varphi \leq 2\big|\big|\,\vartheta_q[\omega]\,\big|\big|^2_\lambda + 2\big|\big|\,B_q[\omega]\,\big|\big|^2_\varphi \qquad \text{in the second case.}\end{aligned}$$

Now we have

$$\begin{aligned}\big|\big|\,A_q[\omega]\,\big|\big|^2_\varphi &= \sum_{J\in\mathbb{I}^*_{q-1}}\sum_{K_1,K_2\in\mathbb{I}^*_k}\sum_{j_1,j_2=1}^n \epsilon_{K_1}^{(j_1,J)}\,\epsilon_{K_2}^{(j_2,J)}\big(D_{j_1}[\omega_{K_1}], D_{j_2}[\omega_{K_2}]\big)_\varphi \\ &= \sum_{J\in\mathbb{I}^*_{q-1}}\sum_{j_1,j_2=1}^n \epsilon_{[(j_1,J)]}^{(j_1,J)}\,\epsilon_{[(j_2,J)]}^{(j_2,J)}\big(D_{j_1}[\omega_{[(j_1,J)]}], D_{j_2}[\omega_{[(j_2,J)]}]\big)_\varphi.\end{aligned}$$

But $\epsilon^{j_1,J}_{[j_1,J]}\,\epsilon^{j_2,J}_{[j_2,J]} = 0$ unless $j_1, j_2 \notin J$ and $[j_1, J] = [j_1, J]$ and $[j_2, J] = [j_2, J]$. Thus this sum is

$$\begin{aligned}
\big|\big|\, A_q[\omega]\,\big|\big|^2_\varphi &= \sum_{J\in\mathbb{I}^*_{q-1}} \sum_{j,k=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\epsilon^{(k,J)}_{[(k,J)]}\big(D_j[\omega_{[(j,J)]}], D_k[\omega_{[(k,J)]}]\big)_\varphi \\
&= \sum_{J\in\mathbb{I}^*_{q-1}} \sum_{j,k=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\epsilon^{(k,J)}_{[(k,J)]}\Big(\frac{\partial\omega_{[(j,J)]}}{\partial\bar z_k}, \frac{\partial\omega_{[(k,J)]}}{\partial\bar z_j}\Big)_\varphi \\
&\quad + \sum_{J\in\mathbb{I}^*_{q-1}} \sum_{j,k=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\epsilon^{(k,J)}_{[(k,J)]}\Big(\omega_{[(j,J)]}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar z_k}, \omega_{[(k,J)]}\Big)_\varphi \\
&\quad\quad + \sum_{J\in\mathbb{I}^*_{q-1}} \Big[\sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\frac{\partial\rho}{\partial z_j}\omega_{[(j,J)]}\,,\, \sum_{k=1}^n \epsilon^{(k,J)}_{[(k,J)]} D_k[\omega_{[(k,J)]}]\Big]_\varphi \\
&\quad\quad\quad - \sum_{J\in\mathbb{I}^*_{q-1}} \Big[\sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\omega_{[(j,J)]}, \sum_{k=1}^n \epsilon^{(k,J)}_{[(k,J)]}\,\frac{\partial\rho}{\partial z_k}\,\frac{\partial\omega_{[(k,J)]}}{\partial\bar z_j}\Big]_\varphi.
\end{aligned}$$

Suppose now that $\overline{\partial}\rho \vee \omega = 0$. Then for every $J \in \mathbb{I}^*_{q-1}$ and all $z \in \partial\Omega$ we have

$$\sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\frac{\partial\rho}{\partial z_j}(z)\,\omega_{[(j,J)]}(z) = \sum_{j=1}^n \omega_{(j,J)}\,\frac{\partial\rho}{\partial z_j}(z) = 0, \tag{2.5}$$

and this means that the third sum equals zero. Also, this means that the operator

$$L_J = \sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\omega_{[(j,J)]}\,\frac{\partial}{\partial z_j}$$

annihilates ρ along $\partial\Omega$, and hence is tangential. But

$$\begin{aligned}
&\Bigg\langle \sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\omega_{[(j,J)]}, \sum_{k=1}^n \epsilon^{(k,J)}_{[(k,J)]}\,\frac{\partial\rho}{\partial z_k}\,\frac{\partial\omega_{[(k,J)]}}{\partial\bar z_j}\Bigg\rangle_z \\
&\qquad = \sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\omega_{[(j,J)]}\,\overline{\sum_{k=1}^n \epsilon^{(k,J)}_{[(k,J)]}\,\frac{\partial\rho}{\partial z_k}\,\frac{\partial\omega_{[(k,J)]}}{\partial\bar z_j}} \\
&\qquad = \sum_{j=1}^n \epsilon^{(j,J)}_{[(j,J)]}\,\omega_{[(j,J)]}\,\frac{\partial}{\partial z_j}\overline{\Big[\sum_{k=1}^n \epsilon^{(k,K)}_{[(k,K)]}\,\omega_{[(k,J)]}\,\frac{\partial\rho}{\partial z_k}\Big]} \\
&\qquad\quad - \sum_{j,k=1}^m \epsilon^{(j,J)}_{[(j,J)]}\,\epsilon^{(k,K)}_{[(k,J)]}\,\omega_{[(j,J)]}\,\overline{\omega_{[(k,K)]}}\,\frac{\partial^2\rho}{\partial z_j\partial\bar z_k} \\
&\qquad = L_J\overline{\Big[\sum_{k=1}^n \epsilon^{(k,K)}_{[(k,K)]}\,\omega_{[(k,J)]}\,\frac{\partial\rho}{\partial z_k}\Big]} \\
&\qquad\quad - \sum_{j,k=1}^m \epsilon^{(j,J)}_{[(j,J)]}\,\epsilon^{(k,J)}_{[(k,J)]}\,\omega_{[(j,J)]}\,\overline{\omega_{[(k,J)]}}\,\frac{\partial^2\rho}{\partial z_j\partial\bar z_k}
\end{aligned}$$

Thus we have shown

$$\begin{aligned}
&\big|\big|\,\overline{\partial}_q[\omega]\,\big|\big|_\varphi^2 + \big|\big|\,A_q[\omega]\,\big|\big|_\varphi^2 \\
&= \sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^n \Big|\Big|\,\frac{\partial\omega_K}{\partial\bar z_k}\,\Big|\Big|_\varphi^2 \\
&\quad + \sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^n \int_\Omega \epsilon_{[(j,J)]}^{(j,J)}\,\epsilon_{[(k,J)]}^{(k,J)}\omega_{[(j,J)]}(z)\,\overline{\omega_{[(k,J)]}(z)}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar z_k}(z)\,dm(z) \\
&\qquad + \sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^n \int_{\partial\Omega} \epsilon_{[(j,J)]}^{(j,J)}\,\epsilon_{[(k,J)]}^{(k,J)}\omega_{[(j,J)]}(z)\,\overline{\omega_{[(k,J)]}(z)}\,\frac{\partial^2\rho}{\partial z_j\partial\bar z_k}(z)\,d\sigma(z).
\end{aligned}$$

2.6. Further estimates.

We now want to derive some consequences of Theorem 2.10. We shall assume that the domain Ω is pseudo-convex, so that

$$\sum_{j,k=1}^n \frac{\partial^2\rho}{\partial z_j\partial\bar z_k}(z)\,\xi_j\,\overline{\xi_k} \ge 0$$

for every $z\in\partial\Omega$ and every $\xi=(\xi_1,\ldots,\xi_n)\in\mathbb{C}^n$ such that

$$\sum_{j=1}^n \xi_j\,\frac{\partial\rho}{\partial z_j}(z) = 0.$$

According to equation (2.4), for each $J\in\mathbb{I}_{q-1}^*$, the vector $(\omega_{(1,J)},\ldots,\omega_{(n,J)}$ satisfies this condition. It follows that

$$\sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^n \int_{\partial\Omega}\omega_{(j,J)}\,\overline{\omega_{(k,J)}}\,e^{-\varphi(z)}\,d\sigma(z)\ge 0.$$

We also set $\psi=0$ so that $\lambda=-\varphi$. This means that

$$\vartheta_q\Big[\sum_{K\in\mathbb{I}_q^*}\omega_K\,d\bar z_K\Big] = e^\varphi\sum_{J\in\mathbb{I}_{q-1}^*}\Big[\sum_{j=1}^n\frac{\partial\omega_{(j,J)}}{\partial z_j}\Big]\,d\bar z_K$$

is the formal adjoint of $\overline{\partial}$ in the unweighted case, multiplied by e^φ. In this situation, Theroem 2.10 gives

$$\begin{aligned}
&\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^n \Big|\Big|\,\frac{\partial\omega_K}{\partial\bar z_k}\,\Big|\Big|_\varphi^2 + \sum_{J\in\mathbb{I}_{q-1}^*}\int_\Omega\sum_{j,k=1}^n \omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar z_k}(z)\,e^{-\varphi(z)}\,dm(z) \\
&\quad\le \big|\big|\,\overline{\partial}_q[\omega]\,\big|\big|_\varphi^2 + 2\big|\big|\,\vartheta_q[\omega]\,\big|\big|_{-\varphi}^2 + 2\sum_{J\in\mathbb{I}_{q-1}^*}\int_\Omega\Big|\sum_{j=1}^n\frac{\partial\varphi}{\partial z_j}(z)\,\omega_{(j,J)}(z)\Big|^2 e^{-\varphi(z)}\,dm(z)
\end{aligned}$$

Following Catlin [**Cat84**], we let $\varphi=\frac{1}{6}\,e^\lambda$. Then

$$\Big|\sum_{j=1}^n\frac{\partial\varphi}{\partial z_j}\,\omega_{(j,J)}\Big|^2 = \frac{1}{36}\,e^{2\lambda}\,\Big|\sum_{j=1}^n\frac{\partial\lambda}{\partial z_j}\omega_{(j,J)}\Big|^2$$

and

$$\sum_{j,k=1}^{n} \omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\varphi}{\partial z_j\partial\bar{z}_k}(z) = \frac{1}{6}\,e^{\lambda}\sum_{j,k=1}^{n}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\lambda}{\partial z_j\partial\bar{z}_k} + \frac{1}{6}\,e^{\lambda}\Big|\sum_{j=1}^{n}\frac{\partial\lambda}{\partial z_j}\omega_{(j,J)}\Big|^2.$$

Thus it follows that

$$\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2 + \frac{1}{6}\sum_{J\in\mathbb{I}_{q-1}^*}\int_\Omega\sum_{j,k=1}^{n}\omega_{(j,J)}(z)\,\overline{\omega_{(k,J)}(z)}\,\frac{\partial^2\lambda}{\partial z_j\partial\bar{z}_k}(z)\,e^{(\lambda-\varphi)(z)}\,dm(z)$$
$$+\frac{1}{6}\sum_{J\in\mathbb{I}_{q-1}^*}\int_\Omega\Big|\sum_{j=1}^{n}\frac{\partial\lambda}{\partial z_j}\omega_{(j,J)}\Big|^2\Big(e^{\lambda}-\frac{1}{3}\,e^{2\lambda}\Big)\,e^{-\varphi}\,dm(z)$$
$$\leq \Big|\Big|\,\overline{\partial}_q[\omega]\,\Big|\Big|_{\varphi}^2 + 2\Big|\Big|\,\vartheta_q[\omega]\,\Big|\Big|_{-\varphi}^2$$

Now if $0 \leq \lambda \leq 1$, we have

$$e^{\lambda} - \frac{1}{3}\,e^{2\lambda} \geq 0,$$
$$e^{\lambda-\varphi} \geq \frac{3}{4}$$
$$e^{-\varphi} \leq 1$$
$$e^{\varphi} \leq e^{\frac{1}{6}} \leq 2.$$

Thus we have established

THEOREM 2.13. *Suppose that $\lambda \in \mathcal{C}^2(\overline{\Omega})$, with $0 \leq \lambda \leq 1$. If $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}$ and if $\overline{\partial}\rho \vee \omega = 0$ on $\partial\Omega$, then*

$$\sum_{K\in\mathbb{I}_q^*}\sum_{k=1}^{n}\Big|\Big|\,\frac{\partial\omega_K}{\partial\bar{z}_k}\,\Big|\Big|_{\varphi}^2 + \sum_{J\in\mathbb{I}_{q-1}^*}\sum_{j,k=1}^{n}\int_\Omega\omega_{(j,J)}(z)\,\omega_{(k,J)}(z)\,\frac{\partial^2\lambda}{\partial z_j\partial\bar{z}_k}(z)\,dm(z)$$
$$\leq C\left[\Big|\Big|\,\overline{\partial}[\omega]\,\Big|\Big|_{L^2(\Omega}^2 + \Big|\Big|\,\vartheta[\omega]\,\Big|\Big|_{L^2(\Omega)}^2\right],$$

where C is an absolute constant.

2.7. Hilbert spaces.

Let $\{\lambda,\, \psi,\, \varphi\}$ be three real valued functions defined on $\overline{\Omega}$. When we try to imitate the finite dimensional model, we replace the sequence

$$U \xrightarrow{S} V \xrightarrow{T} W.$$

by the sequence

$$L_{\lambda}^{0,q-1}(\Omega) \xrightarrow{\overline{\partial}_{q-1}} L_{\psi}^{0,q}(\Omega) \xrightarrow{\overline{\partial}_q} L_{\varphi}^{0,q+1}(\Omega).$$

These differential operators are not defined on the whole Hilbert space. Instead, we turn again to the theory of unbounded operators and be satisfied with linear transformations which are closed and densely defined. We focus on the operator $\overline{\partial}_q$; the operator $\overline{\partial}_{q-1}$ is handled similarly.

We need to specify a domain for $\overline{\partial}_q$. Let us temporarily distinguish between the operator $\overline{\partial}_q$ defined on $\mathcal{E}^{0,q}(\Omega)$ and the unbounded Hilbert space operator D_q we wish to define. Let

$$\mathrm{Dom}(\overline{D}_q) = \Big\{\omega \in L^{0,q}_{\psi}(\Omega) \,\Big|\, \overline{\partial}[\omega] \in L^{0,q+1}_{\varphi} \text{ in the sense of distributions}\Big\}.$$

Precisely, this means that $\omega \in L^{0,q}_{\psi}(\Omega)$ belongs to the domain of $\overline{D}_q$ if and only if there exists a form $\eta \in L^{0,q+1}_{\varphi}(\Omega)$ such that for every form $\alpha \in \mathcal{D}^{0,q+1}(\Omega)$ (smooth forms with compact support in Ω), we have

$$\big(\eta, \alpha\big)_{\varphi} = \big(\omega, \vartheta_{q+1}[\alpha]\big)_{\psi}. \tag{2.6}$$

If it exists, the form η is uniquely determined since the space $\mathcal{D}^{0,q+1}(\Omega)$ is dense in $L^{0,q+1}_{\varphi}(\Omega)$, and we write $\overline{D}_q[\omega] = \eta$.

The operator $\big(\overline{D}_q, \mathrm{Dom}(\overline{D}_q)\big)$ is closed, densely defined, and is a natural extension of the differential operator $\overline{\partial}_q$. Thus if $\omega \in \mathcal{E}(\Omega)^{0,q} \cap L^{0,q}_{\psi}(\Omega)$ and if $\overline{\partial}_q[\omega] \in L^{0,q+1}_{\varphi}(\Omega)$, then $\eta = \overline{\partial}_q[\omega]$ satisfies equation (2.6), and so $\omega \in \mathrm{Dom}(\overline{D}_q)$ with $\overline{D}_q[\omega] = \overline{\partial}_q[\omega]$. The operator $\overline{D}_q$ is densely defined since the space $\mathcal{D}^{0,q}(\Omega)$ is certainly dense in $L^{0,q}_{\psi}(\Omega)$ and $\mathcal{D}^{0,q}(\Omega) \subset \mathrm{Dom}(\overline{D}_q)$. Finally, $\big(\overline{D}_q, \mathrm{Dom}(\overline{D}_q)\big)$ is closed since if $\omega_n \in \mathrm{Dom}(\overline{D}_q)$ and if $\omega_n \to \omega_0$ in $L^{0,q}_{\psi}$ and $\overline{D}_q[\omega_n] \to \eta_0 \in L^{0,q+1}_{\varphi}(\Omega)$, then for any $\alpha \in \mathcal{D}^{0,q+1}(\Omega)$ we have

$$\big(\eta_0, \alpha\big)_{\varphi} = \lim_{n\to\infty} \big(\overline{D}_q[\omega_n], \alpha\big)_{\varphi} = \lim_{n\to\infty} \big(\omega_n, \vartheta_{q+1}[\alpha]\big)_{\psi} = \big(\omega_0, \vartheta_{q+1}[\alpha]\big)_{\psi}.$$

Thus $\omega_0 \in \mathrm{Dom}(\overline{D}_q)$ and $\overline{D}_q[\omega_0] = \eta_0$.

In fact, we can say more about the domain of $\overline{D}_q$.

Lemma 2.14. *The space $\mathcal{E}^{0,q}_{\psi}(\Omega)_0$ is dense in Dom$(\overline{D}_q)$ in the graph norm $\big|\big|\,\omega\,\big|\big|_{\psi} + \big|\big|\,\overline{D}_q[\omega]\,\big|\big|_{\varphi}$. Precisely, given $\omega \in$ Dom$(\overline{D}_q)$, there is a sequence $\{\omega_n\}$ in $\mathcal{E}^{0,q}(\overline{\Omega})_0$ so that*

$$\lim_{n\to\infty} \big|\big|\,\omega - \omega_n\,\big|\big|_{\psi} = 0, \qquad \textit{and}$$
$$\lim_{n\to\infty} \big|\big|\,\overline{D}_q[\omega] - \overline{\partial}_q[\omega_n]\,\big|\big|_{\varphi} = 0.$$

GIVE PROOF

Since the Hilbert space operator $\big(\overline{D}_q, \mathrm{Dom}(\overline{D}_q)\big)$ is closed and densely defined, it follows that it has a closed, densely defined adjoint $\overline{D}^*_q$. Recall that

$$\mathrm{Dom}(\overline{D}^*_q) = \Big\{\eta \in L^{0,q+1}_{\varphi}(\Omega) \,\Big|\, \big|\big(\overline{D}_q[\omega], \eta\big)_{\varphi}\big| \leq C_\eta \,\big|\big|\,\omega\,\big|\big|_{\psi} \text{ for all } \omega \in \mathrm{Dom}(\overline{D}_q)\Big\},$$

and that for $\omega \in \mathrm{Dom}(\overline{D}_q)$ and $\eta \in \mathrm{Dom}(\overline{D}^*_q)$ we have

$$\big(\omega, \overline{D}^*_q[\eta]\big)_{\psi} = \big(\overline{D}_q[\omega], \eta\big)_{\varphi}. \tag{2.7}$$

It is easy to calculate the action of $\overline{D}^*_q$ on forms which are smooth on Ω, and to characterize the elements of the domain of $\overline{D}^*_q$ which are smooth up to the boundary.

Lemma 2.15.

(1) If $\eta \in \mathcal{E}^{0,q+1}(\Omega) \cap$ Dom$(\overline{D}^*_q)$, then $\overline{D}^*_q[\eta] = \vartheta_{q+1}[\eta]$.

(2) If $\eta \in \mathcal{E}(\overline{\Omega})^{0,q+1}_0(\Omega)$, then $\eta \in$ Dom$(\overline{D}^*_q)$ if and only if $\overline{\partial}\rho \vee \eta(z) = 0$ for all $z \in \partial\Omega$.

PROOF. Let $\eta \in \mathcal{E}^{0,q+1}(\Omega) \cap \mathrm{Dom}(\overline{D}_q^*)$. Since $\mathcal{D}^{0,q}(\Omega) \subset \mathrm{Dom}(\overline{D}_q)$, for every $\omega \in \mathcal{D}^{0,q}(\Omega)$ we have

$$\big(\overline{\partial}_q[\omega], \eta\big)_\varphi = \big(\overline{D}_q[\omega], \eta\big)_\varphi = \big(\omega, \overline{D}_q^*[\eta]\big)_\psi.$$

But according to (2) in Proposition 2.8, we also have

$$\big(\overline{\partial}_q[\omega], \eta\big)_\varphi = \big(\omega, \vartheta_{q+1}[\eta]\big)_\psi.$$

Since $\mathcal{D}^{0,q}(\Omega)$ is dense, it follows that $\overline{D}_q^*[\eta] = \vartheta_{q+1}[\eta]$.

Next, suppose $\eta \in \mathcal{E}(\overline{\Omega})_0^{0,q+1}(\Omega) \cap \mathrm{Dom}(\overline{D}_q^*)$. Then for every $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}(\Omega) \subset \mathrm{Dom}(\overline{D}_q)$, we have

$$\big(\overline{\partial}_q[\omega], \eta\big)_\varphi = \big(\overline{D}_q[\omega], \eta\big)_\psi = \big(\omega, \overline{D}_q^*[\eta]\big)_\psi = \big(\omega, \vartheta_{q+1}[\eta]\big)_\psi.$$

Then according to (3) in Proposition 2.8 we have $\overline{\partial}\rho \vee \eta = 0$ on $\partial\Omega$.

Conversely, if $\eta \in \mathcal{E}(\overline{\Omega})_0^{0,q+1}(\Omega)$ and $\overline{\partial}\rho \vee \eta = 0$ on $\partial\Omega$, then according to (1) in Proposition 2.8, for every $\omega \in \mathcal{E}(\overline{\Omega})_0^{0,q}(\Omega)$ we have

$$\big(\overline{D}_q[\omega], \eta\big)_\varphi = \big(\overline{\partial}_q[\omega], \eta\big)_\varphi = \big(\omega, \vartheta_{q+1}[\eta]\big)_\psi,$$

Now let $\omega_0 \in \mathrm{Dom}(\overline{D}_q)$, and let $\{\omega_n\} \subset \mathcal{E}(\overline{\Omega})_0^{0,q}$ be an approximating sequence as in Lemma 2.14. Then

$$\begin{aligned}
\big|\big(\overline{D}_q[\omega], \eta\big)_\varphi\big| &= \lim_{n\to\infty} \big|\big(\overline{D}_q[\omega_n], \eta\big)_\varphi\big| \\
&= \lim_{n\to\infty} \big|\big(\omega_n, \vartheta_{q+1}[\eta]\big)_\varphi\big| \\
&= \big|\big(\omega_0, \vartheta_{q+1}[\eta]\big)_\varphi\big| \\
&\leq \big|\big|\, \vartheta_{q+1}[\eta] \,\big|\big|_\varphi \big|\big|\, \omega_0 \,\big|\big|_\varphi.
\end{aligned}$$

It follows that $\eta \in \mathrm{Dom}(\overline{D}_q^*)$, with $\overline{D}_q[\eta] = \vartheta_{q+1}[\eta]$, and this completes the proof. □

We give a special name to the smooth functions which satisfy this boundary condition.

DEFINITION 2.16. $\mathcal{N}^{0,q}(\Omega) = \Big\{\omega \in \mathcal{E}_0^{0,q}(\overline{\Omega}) \,\Big|\, \overline{\partial}\rho \wedge \omega(z) = 0 \textit{ for all } z \in \partial\Omega\Big\}.$

THEOREM 2.17. *The space $\mathcal{N}^{0,q}$ is dense in $Dom(\overline{D}_q) \cap Dom(\overline{D}_{q-1}^*)$ in the graph norm $\big|\big|\,\omega\,\big|\big|_\psi + \big|\big|\,\overline{D}_q[\omega]\,\big|\big|_\varphi + \big|\big|\,\overline{D}_{q-1}^*[\omega]\,\big|\big|_\lambda$. Precisely, if $\omega \in Dom(\overline{D}_q) \cap Dom(\overline{D}_{q-1}^*)$, there exists a sequence $\{\omega_n\}$ in $\mathcal{N}^{0,q}(\Omega)$ such that*

$$\begin{aligned}
\lim_{n\to\infty} \big|\big|\,\omega - \omega_n\,\big|\big|_\psi &= 0, \qquad \textit{and} \\
\lim_{n\to\infty} \big|\big|\,\overline{D}_q[\omega] - \overline{\partial}_q[\omega_n]\,\big|\big|_\varphi &= 0 \\
\lim_{n\to\infty} \big|\big|\,\overline{D}_{q-1}^*[\omega] - \vartheta_q[\omega_n]\,\big|\big|_\lambda &= 0.
\end{aligned}$$

2.8. The Neumann operator.

As a corollary of the density established in Theorem 2.17, we have

THEOREM 2.18. *Suppose that Ω is a bounded pseudo-convex with $\mathcal{C}^2$-boundary. Suppose the diameter of Ω is D. Then if $\omega \in Dom(\overline{D}_q) \cap Dom(\overline{D}^*_{q-1})$, we have*

$$\big|\big|\,\omega\,\big|\big|^2_{L^2(\Omega)} \leq 4D^2q^{-1}\Big[\big|\big|\,\overline{\partial}_q[\omega]\,\big|\big|^2_{L^2(\Omega)} + \big|\big|\,\vartheta_q[\omega]\,\big|\big|^2_{L^2(\Omega)}\Big].$$

PROOF. We can assume that Ω lies in the Euclidean ball in $\mathbb{C}^n$ centered at 0 with radius $D/2$. Then take $\lambda(z) = 4\,D^{-2}|z|^2$. In this case $A = 4D^{-2}$, and this completes the proof. □

Now it follows from our discussion in Section 2.2 that we have the following solution to the $\overline{\partial}$-Neumann problem. Let $\overline{D}_q$ be the Hilbert space extension of the operator $\overline{\partial}_q$. Put

$$\begin{aligned}
\mathcal{D}^{0,q}(\Omega) &= \Big\{\omega \in L^{0,q}(\Omega)\,\Big|\,\omega \in \mathrm{Dom}(\overline{D}_q) \cap \mathrm{Dom}(\overline{D}^*_{q-1})\Big\}\\
\mathrm{Dom}(\Box_q) &= \Big\{\omega \in \mathcal{D}^{0,q}(\Omega)\,\Big|\,\overline{D}_q[\omega] \in \mathrm{Dom}(\overline{D}^*_q) \text{ and } \overline{D}^*_{q-1}[\omega] \in \mathrm{Dom}(D_{q-1})\Big\}\\
\Box_q[\omega] &= \overline{D}_{q-1}\,\overline{D}^*_{q-1}[\omega] + \overline{D}^*_q\,\overline{D}_q[\omega] \quad \text{for } \omega \in \mathrm{Dom}(\Box_q).
\end{aligned}$$

THEOREM 2.19. *Let $\Omega \subset \mathbb{C}^n$ be a bounded pseudo-convex domain with $\mathcal{C}^2$-boundary. Let $1 \leq q \leq n$. Then the ranges of the operators $\overline{D}_{q-1}$ and $\overline{D}^*_q$ are closed subspaces of $L^{0,q}$, and*

$$\begin{aligned}
R(\overline{D}_{q-1}) &= N(\overline{D}_q) = N(\overline{D}^*_{q-1})^\perp\\
R(\overline{D}^*_q) &= N(\overline{D}^*_{q-1}) = N(D_q)^\perp\\
L^{0,q}(\Omega) &= N(\overline{D}_q) \oplus N(\overline{D}^*_{q-1}).
\end{aligned}$$

There is a bounded, self-adjoint operator $N_q : L^{0,q}(\Omega) \to L^{0,q}(\Omega)$ which is one-to-one, and which has the following properties:

$$N_q : L^{0,q} \to Dom(\Box_q) \tag{a}$$

$$N_q : N(\overline{D}_q) \to N(\overline{D}_q) \cap Dom(\overline{D}^*_{q-1}), \tag{2.8}$$

$$N_q : N(\overline{D}^*_{q-1}) \to N(\overline{D}^*_{q-1}) \cap Dom(\overline{D}_q); \tag{2.9}$$

$$\Box_q\big[N_q[\omega]\big] = \omega \qquad \text{for all } \omega \in L^{0,q}; \tag{2.10}$$

$$N_q\big[\Box[\omega]\big] = \omega \qquad \text{for all } \omega \in Dom(\Box); \tag{2.11}$$

$$\big|\big|\,N_q[\omega]\,\big|\big|_{L^2(\Omega)} \leq 4D^2q^{-1}\,\big|\big|\,\omega\,\big|\big|_{L^2(\Omega)}. \tag{2.12}$$

CHAPTER 9

Appendix on background material

In this chapter we gather together a summary of material that we use in the rest of the book.

1. Distributions, Fourier transforms, and Fundamental Solutions

The purpose of this section is to establish some standard terminology that will be used in the rest of the book, and to recall the definitions and basic properties of the spaces $\mathcal{D}(U)$ and $\mathcal{E}(U)$ of infinitely differentiable functions defined on open subsets $U \subset \mathbb{R}^n$. We also briefly discuss the properties of the dual spaces $\mathcal{D}'(U)$ and $\mathcal{E}'(U)$, which are spaces of distributions. We give the definition of the Fourier transform, and then use this to define the scale of Sobolev spaces $H^s(\mathbb{R}^n)$. Finally, we discuss the concepts of *fundamental solution* and *parametrix* for a partial differential operator.

1.1. Notation.

If X is a topological space and if $E \subset X$ is a subset, we denote by $\overline{E}$, E^o, and ∂E the closure, the interior, and the boundary of E. If $U \subset X$ is open, a subset $E \subset U$ is *relatively compact* if $\overline{E} \subset U$ and if $\overline{E}$ is compact. A subset $\Omega \subset \mathbb{R}^n$ is a *domain* if it is open and connected. We say that a domain Ω has $\mathcal{C}^m$-smooth boundary if there is an open neighborhood U of $\partial\Omega$ and an m-times continuously differentiable function $\rho : U \to \mathbb{R}$ so that

(a) $\Omega \cap U = \Big\{x \in U \,\Big|\, \rho(x) < 0\Big\}$;
(b) The gradient $\nabla\rho(x) \neq 0$ for all $x \in U$.

The function ρ is called a *defining function.* If ρ is infinitely differentiable, we sometimes say that Ω has smooth boundary.

We shall use standard multi-index notation. Let $\mathbb{Z}_+ = \{0, 1, \ldots\}$ be the set of non-negaitve integers. A multi-index α on $\mathbb{R}^n$ is an n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$. If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and if α is an n-tuple,

$$\begin{aligned} |\alpha| &= \alpha_1 + \cdots + \alpha_n \\ \alpha! &= \alpha_1! \cdots \alpha_n! \\ x^\alpha &= x_1^{\alpha_1} \cdots x_n^{\alpha_n} \\ \partial_x^\alpha \varphi &= \frac{\partial^{|\alpha|}\varphi}{\partial x^\alpha} = \frac{\partial^{\alpha_1 + \cdots + \alpha_n}\varphi}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \end{aligned} \tag{1.1}$$

1.2. Spaces of smooth functions.

Let $U \subset \mathbb{R}^n$ be open. The space $\mathcal{C}^m(U)$ is the space of m-times continuously differentiable functions on U and $\mathcal{C}^\infty(U)$ is the space of infinitely differentiable functions on U. When we need to distinguish between real- and (possibly) complex-valued functions on U, we write $\mathcal{C}^\infty_{\mathbb{R}}(U)$ or $\mathcal{C}^\infty_{\mathbb{C}}(U)$. For every $m \in \mathbb{Z}_+$, $\mathcal{C}^m(U) \subset \mathcal{C}^{m+1}(U) \supset \mathcal{C}^\infty(U)$. We focus primarily on the space $\mathcal{C}^\infty(U)$.

If $K \subset U$ is a compact set, $m \in \mathbb{Z}_+$, and $f \in \mathcal{C}^\infty(U)$, let

$$\big|\big|\, f \,\big|\big|_{K,m} = \sup_{|\alpha|\leq m} \sup_{x\in K} \big|\partial_x^\alpha \varphi(x)\big| \tag{1.2}$$

Then $\big|\big| \cdot \big|\big|_{K,m}$ is a semi-norm on the space $\mathcal{C}^\infty(U)$. We can define a topology on this space so that a sequence $\{f_n\} \subset \mathcal{C}^\infty(U)$ converges to a limit function $f_0 \in \mathcal{C}^\infty(U)$ if and only if $\lim_{n\to\infty} \big|\big|\, f_0 - f_n \,\big|\big|_{K,m} = 0$ for every compact subset $K \subset \Omega$ and for every non-negative integer m. The space $\mathcal{C}^\infty(U)$ equipped with this topology is denoted by $\mathcal{E}(U)$. It is easy to check that if $\{f_n\} \subset \mathcal{E}(U)$ is a sequence which is Cauchy in every semi-norm $\big|\big| \cdot \big|\big|_{K,m}$, then the sequence converges to a limit $f_0 \in \mathcal{E}(U)$.

If $f \in \mathcal{C}^\infty(U)$, the *support* of f is the closure of the set of points $x \in U$ such that $f(x) \neq 0$. The support of f is denoted by $\text{suppt}(f)$. Then $\mathcal{C}^\infty_0(U)$ is the subspace of $\mathcal{C}^\infty(U)$ consisting of functions with compact support. There is a topology on $\mathcal{C}^\infty_0(U)$ such that a sequence $\{f_1, f_n, \ldots, f_n, \ldots\} \subset \mathcal{C}^\infty_0(U)$ converges to a limit function $f_0 \in \mathcal{C}^\infty_0(U)$ if and only there is a compact set $K \subset U$ so that $\text{suppt}(f_n) \subset K$ for all $n \geq 0$ and $\lim_{n\to\infty} \big|\big|\, f_0 - f_n \,\big|\big|_{K,m} = 0$ for every non-negative integer m. The space $\mathcal{C}^\infty_0(U)$ with this topology is denoted by $\mathcal{D}(U)$. It should be noted that this topology is *not* the same as the topology that $\mathcal{C}^\infty_0(U)$ naturally inherits as a subspace of $\mathcal{E}(U)$. However, with the given topologies on the two spaces, the inclusion map $\mathcal{D}(U) \hookrightarrow \mathcal{E}(U)$ is continuous. Note that if $\chi \in \mathcal{C}^\infty_0(U)$, then the mapping $R_\chi : \mathcal{E}(U) \to \mathcal{D}(U)$ given by $R_\chi[f] = \chi\, f$ is continuous.

1.3. Spaces of distributions.

The space of *distributions* on U is the dual space $\mathcal{D}'(U)$ of continuous real- or complex-valued continuous linear functionals on $\mathcal{D}(U)$. The pairing between a distribution $T \in \mathcal{D}'(U)$ and a smooth, compactly supported function $\varphi \in \mathcal{D}(U)$ is denoted by $\big\langle T, \varphi \big\rangle$. Of course, if $f \in L^1_{loc}(U)$, then f induces a distribution T_f whose action is given by $\big\langle T_f, \varphi \big\rangle = \int_U f(x)\, \varphi(x)\, dx$. Abusing notation, we shall frequently denote the distribution T_f simply by f. We shall also engage in another, perhaps more serious abuse by sometimes writing the action of a distribution $u \in \mathcal{D}'(U)$ on a function $\varphi \in \mathcal{D}(U)$ by $\int_U u(x)\, \varphi(x)\, dx$.

The space $\mathcal{D}(U)$ is closed under the operation of multiplication by a smooth function in $\mathcal{E}(U)$, and also under differentiation. By duality, this allows us to define the product of a distribution and a smooth function, and to define the derivative of a distribution. Thus let $T \in \mathcal{D}'(U)$ be a distribution. If $a \in \mathcal{E}(U)$ is a smooth function, the product aT is the distribution defined by the requirement that

$$\big\langle aT, \varphi \big\rangle = \big\langle T, a\varphi \big\rangle. \tag{1.3}$$

For every multi-index $\alpha \in \mathbb{Z}^n_+$, the distribution $\partial^\alpha T$ is defined by the requirement that

$$\big\langle \partial^\alpha T, \varphi \big\rangle = (-1)^{|\alpha|} \big\langle T, \partial^\alpha \varphi \big\rangle. \tag{1.4}$$

In particular, if

$$P[\varphi](x) = \sum_{|\alpha|\leq M} a_\alpha(x) \partial_x^\alpha[\varphi](x) \tag{1.5}$$

is a linear partial differential operator of order M with coefficients $\{a_\alpha\} \subset \mathcal{E}(U)$ and if $u \in \mathcal{D}'(U)$ is a distribution, then $P[u]$ is also a distribution on U whose action on a test function $\varphi \in \mathcal{E}(U)$ is given by

$$\big\langle P[u], \varphi \big\rangle = \big\langle u, \sum_{|\alpha| \leq M} \partial_x^\alpha [a_\alpha \varphi] \big\rangle. \tag{1.6}$$

Note that the operator

$$P^*[\varphi](x) = \sum_{|\alpha| \leq M} \partial_x^\alpha \big[a_\alpha \, \varphi\big](x) \tag{1.7}$$

is also a linear partial differential operator of order M. It is called the formal adjoint of P. Thus equation (1.6) is equivalent to the statement

$$\big\langle P[u], \varphi \big\rangle = \big\langle u, P^*[\varphi] \big\rangle. \tag{1.6$'$}$$

If $T \in \mathcal{D}'(U)$, the support of T is the smallest closed set $E \subset U$ such that if $\varphi \in \mathcal{C}_0^\infty(U - E)$, then $\big\langle T, \varphi \big\rangle = 0$. We again denote the support of a distribution T by suppt(T). The dual space $\mathcal{E}'(U)$ is the space of continuous real- or complex-valued linear functionals on $\mathcal{E}(U)$. Since the inclusion map $\mathcal{D}(U) \hookrightarrow \mathcal{E}(U)$ is continuous, every continuous linear functional on $\mathcal{E}(U)$ restricts to a continuous linear functional on $\mathcal{D}(U)$, so there is an induced mapping $\mathcal{E}'(U) \to \mathcal{D}'(U)$. It is not hard to see that this mapping is one-to-one, so $\mathcal{E}'(U)$ is a subspace of distributions on U. In fact, $\mathcal{E}'(U)$ is exactly the space of distributions on U with compact support.

The *Schwartz space* $\mathcal{S}(\mathbb{R}^n)$ is the space of complex-valued infinitely differentiable functions φ defined on $\mathbb{R}^n$ such that for every multi-index α and every positive integer N it follows that for all $x \in \mathbb{R}^n$

$$\big|\partial_x^\alpha \varphi(x)\big| \leq C_{\alpha,\beta}(1 + |x|)^{-N}. \tag{1.8}$$

The collection of semi-norms

$$\big|\big|\, \varphi \,\big|\big|_{\alpha,N} = \sup_{x \in \mathbb{R}^n} \big|(1 + |x|)^N \partial_x^\alpha f(x)\big| \tag{1.9}$$

allows us to put a topology on $\mathcal{S}(\mathbb{R}^n)$ so that a sequence $\{\varphi_n\} \subset \mathcal{S}(\mathbb{R}^n)$ converges to a limit φ_0 if and only if $\lim_{n \to \infty} \big|\big|\, \varphi_0 - \varphi_n \,\big|\big|_{\alpha,N} = 0$ for every α and N. It is easy to see that $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$.

1.4. The Fourier Transform.

We briefly recall some basic facts about the Fourier transform. This material can be found in many places, such as [**SW71**]. For $f \in L^1(\mathbb{R}^n)$, the *Fourier transform* $\mathcal{F}[f] = \widehat{f}$ is defined by the absolutely convergent integral

$$\mathcal{F}[f](\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i x \cdot \xi} \, f(x) \, dx, \tag{1.10}$$

where $x \cdot \xi = x_1\xi_1 + \cdots + x_n\xi_n$. The inverse Fourier transform $\mathcal{F}^{-1}[f] = f^\vee$ is

$$\mathcal{F}^{-1}[f](x) = f^\vee(x) = \int_{\mathbb{R}^n} e^{2\pi i x \cdot \xi} \, \varphi(\xi) \, d\xi. \tag{1.11}$$

If $\varphi \in \mathcal{S}(\mathbb{R}^n)$, one can show that $\mathcal{F}[\varphi] \in \mathcal{S}(\mathbb{R}^n)$, and the Fourier transform is a continuous and invertible linear mapping of $\mathcal{S}(\mathbb{R}^n)$ to itself. The inversion formula asserts that

$$\varphi(x) = \mathcal{F}^{-1}\big[\mathcal{F}[\varphi]\big](x) = \int_{\mathbb{R}^n} e^{2\pi i x \cdot \xi} \, \widehat{\varphi}(\xi) \, d\xi.. \tag{1.12}$$

If $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n)$, Fubini's theorem shows that

$$\int_{\mathbb{R}^n} \varphi(x)\, \widehat{\psi}(x)\, dx = \int_{\mathbb{R}^n} \widehat{\varphi}(y)\, \psi(y)\, dy. \tag{1.13}$$

Also, if $\varphi \in \mathcal{S}(\mathbb{R}^n)$, the Plancherel formula asserts that

$$\big|\big|\, \varphi \,\big|\big|^2_{L^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\varphi(x)|^2\, dx = \int_{\mathbb{R}^n} |\widehat{\varphi}(\xi)|^2\, d\xi = \big|\big|\, \widehat{\varphi} \,\big|\big|^2_{L^2(\mathbb{R}^n)}, \tag{1.14}$$

and it follows that the Fourier transform extends by continuity to an isometry of $L^2(\mathbb{R}^n)$.

For $\varphi \in \mathcal{S}(\mathbb{R}^n)$, define

$$M^s[\varphi](x) = (1 + 4\pi^2|x|^2)^{\frac{s}{2}}\, \varphi(x). \tag{1.15}$$

Then it is easy to check that M^s maps the space $\mathcal{S}(\mathbb{R}^n)$ to itself, and is continuous and invertible with inverse M^{-s}. Hence we can define a mapping $\Lambda^s : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ by setting $\Lambda^s = \mathcal{F}^{-1} M^s \mathcal{F}$, or explicitly

$$\widehat{\Lambda^s[\varphi]}(\xi) = (1 + 4\pi^2|\xi|^2)^{\frac{s}{2}}\, \widehat{\varphi}(\xi) \tag{1.16}$$

The space of *tempered distributions* $\mathcal{S}'(\mathbb{R}^n)$ is the space of continuous real- or complex-valued linear functionals on $\mathcal{S}(\mathbb{R}^n)$. It is clear that every $f \in L^2(\mathbb{R}^n)$ defines a tempered distribution by setting $\langle T_f, \varphi \rangle = \int_{\mathbb{R}^n} f(x)\, \varphi(x)\, dx$ for $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Since the Fourier transform is continuous and invertible on $\mathcal{S}(\mathbb{R}^n)$, it is possible to extend the defintion of the Fourier transform to the space of tempered distributions by setting

$$\langle \widehat{T}, \varphi \rangle = \langle T, \widehat{\varphi} \rangle \tag{1.17}$$

for $T \in \mathcal{S}'(\mathbb{R}^n)$ and $\varphi \in \mathcal{S}(\mathbb{R}^n)$. We can also extend the mapping Λ^s to the space $\mathcal{S}'(\mathbb{R}^n)$ by setting

$$\langle \Lambda^s T, \varphi \rangle = \langle T, \Lambda^s \varphi \rangle. \tag{1.18}$$

We then have $\widehat{\Lambda^s T} = M^s T$ since

$$\langle \widehat{\Lambda^s T}, \varphi \rangle = \langle \Lambda^s T, \widehat{\varphi} \rangle = \langle T, \Lambda^s \widehat{\varphi} \rangle = \langle \widehat{T}, (\Lambda^s \widehat{\varphi})^{\vee} \rangle = \langle \widehat{T}, M^s \varphi \rangle = \langle M^s \widehat{T}, \varphi \rangle. \tag{1.19}$$

For every $s \in \mathbb{R}^n$, the *Sobolev space* $H^s(\mathbb{R}^n)$ is the subspace of tempered distributions $T \in \mathcal{S}'(\mathbb{R}^n)$ such that $\Lambda^s T \in L^2(\mathbb{R}^n)$. This means that $\widehat{T}$ is given by a locally integrable function and

$$\big|\big|\, T \,\big|\big|^2_s = \int_{\mathbb{R}^n} |\widehat{T}(\xi)|^2 (1 + 4\pi^2|\xi|^2)^s\, d\xi < \infty. \tag{1.20}$$

In particular. $H^0(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, and we often write the norm of $f \in L^2(\mathbb{R}^n)$ as $\big|\big|\, f \,\big|\big|_{L^2(\mathbb{R}^n)} = \big|\big|\, f \,\big|\big|_0$.

1.5. Fundamental solutions and parametrices.

We can now define the concept of a fundamental solution or a parametrix for a partial differential operator P given in equation (1.5).

DEFINITION 1.1.

(1) *A distribution* $K \in \mathcal{D}'(\Omega \times \Omega)$ *is a* fundamental solution *for* P *if for all* $\varphi, \psi \in \mathcal{D}(\Omega)$ *we have*

$$\big\langle K, \psi \otimes P[\varphi]\big\rangle = \big\langle K, P^*[\psi] \otimes \varphi\big\rangle = \int_\Omega \varphi(x)\,\psi(x)\,dx.$$

Equivalently, a continuous linear mappping $\mathcal{K} : \mathcal{D}(\Omega) \to \mathcal{D}'(\Omega)$ *is a* fundamental solution *for the differential operator* P *if*

$$P\big[\mathcal{K}[\varphi]\big] = \mathcal{K}\big[P[\varphi]\big] = \varphi$$

for all $\varphi \in \mathcal{D}(\Omega)$.

(2) *A continuous linear mapping* $\mathcal{S} : \mathcal{D}(\Omega) \to \mathcal{D}'(\Omega)$ *is* infinitely smoothing *if the Schwartz kernel* $S \in \mathcal{C}^\infty(\Omega \times \Omega)$.

(3) *A distribution* $K \in \mathcal{D}'(\Omega \times \Omega)$ *is a* parametrix *for* P *if for all* $\varphi, \psi \in \mathcal{D}(\Omega)$ *we have*

$$\big\langle K, \psi \otimes P[\varphi]\big\rangle = \int_\Omega \varphi(x)\,\psi(x)\,dx + \int_{\Omega\times\Omega} \psi(x)\,\varphi(y)\,S_1(x,y)\,dx\,dy, \quad \textit{and}$$

$$\big\langle K, P^*[\psi] \otimes \varphi\big\rangle = \int_\Omega \varphi(x)\,\psi(x)\,dx + \int_{\Omega\times\Omega} \psi(x)\,\varphi(y)\,S_2(x,y)\,dx\,dy$$

where $S_1, S_2 \in \mathcal{C}^\infty(\Omega \times \Omega)$. *Equivalenty, a continuous linear mappping* $\mathcal{K} : \mathcal{D}(\Omega) \to \mathcal{D}'(\Omega)$ *is a* parametrix *for the differential operator* P *if*

$$P\big[\mathcal{K}[\varphi]\big] = \varphi + \mathcal{S}_1[\varphi] \quad \textit{and} \quad \mathcal{K}\big[P[\varphi]\big] = \varphi + \mathcal{S}_2[\varphi]$$

for all $\varphi \in \mathcal{D}(\Omega)$, *where* $\mathcal{S}_1, \mathcal{S}_2$ *are infinitely smoothing operators. Equivalently,*

Note that $K \in \mathcal{D}'(\Omega \times \Omega)$ is a fundamental solution for P if and only if

$$P_x \otimes I[K] = I \otimes P^*_y[K] = \delta$$

where δ is the distribution on $\Omega \times \Omega$ given by

$$\big\langle \delta, \varphi \otimes \psi\big\rangle = \int_\Omega \varphi(x)\,\psi(x)\,dx.$$

Here $P_x \otimes I$ or $I \otimes P^*_y$ indicates that the operator P or P^* is applied in the x-variables with y fixed, or in the y-variables with x fixed. In particular, a fundamental solution is not unique. For example, one can add to a given fundamental solution K once can add any distribution L such that $P_x \otimes I[L] = I \otimes P^*_y[L] = 0$.

As we will see, properties of fundamental solutions or parametrices for P, if they exist, can provide information about the existence and the regularity of solutions u to the equation $P[u] = g$. In particular, an explicit knowledge of the distribution K, or even estimates on the size of K and its derivatives away from its singularities, can be used to prove estimates for solution in a variety of function spaces.

CHAPTER 10

Pseudodifferential operators

We recall the definitions and basic properties of standard pseudodifferential operators on $\mathbb{R}^n$. Most of this material can be found in [**Ste93**], Chapter VI.

Throughout this chapter we will use the following notation. If $E \subset \mathbb{R}^n$, the closure of E is denoted by $\overline{E}$, the interior of E is denoted by E^o or $int(E)$, and the boundary of E is denoted by ∂E. A subset $\Omega \subset \mathbb{R}^n$ is a *domain* if it is open and connected. If $U \subset \mathbb{R}^n$ is open, a subset $E \subset U$ is *relatively compact* if $\overline{E} \subset U$ and if $\overline{E}$ is compact. This is written $E \Subset U$.

We shall also use standard multi-index notation. $\mathbb{Z}_+ = \{0, 1, \dots\}$ is the set of non-negaitve integers. A multi-index α on $\mathbb{R}^n$ is an n-tuple $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n$. We write

$$\begin{aligned} |\alpha| &= \alpha_1 + \cdots + \alpha_n \\ \alpha! &= \alpha_1! \cdots \alpha_n! \\ x^\alpha &= x_1^{\alpha_1} \cdots x_n^{\alpha_n} \\ \partial_x^\alpha \varphi &= \frac{\partial^{\alpha_1 + \cdots + \alpha_n} \varphi}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \end{aligned} \tag{0.21}$$

1. Functions, distributions, and Fourier transforms

In this section we recall the definitions and basic properties of the spaces $\mathcal{D}(U)$ and $\mathcal{E}(U)$ of infinitely differentiable functions defined on open subsets $U \subset \mathbb{R}^n$. We also discuss the properties of the dual spaces $\mathcal{D}'(U)$ and $\mathcal{E}'(U)$, which are spaces of distributions. We give the definition of the Fourier transform, and then use this to define the scale of Sobolev spaces $H^s(\mathbb{R}^n)$.

1.1. Spaces of smooth functions. Let $U \subset \mathbb{R}^n$ be open. The space $\mathcal{C}^\infty(U)$ is the space of infinitely differentiable functions on U. When we need to distinguish between real- and (possibly) complex-valued functions on U, we write $\mathcal{C}^\infty_{\mathbb{R}}(U)$ or $\mathcal{C}^\infty_{\mathbb{C}}(U)$. If $K \subset U$ is a compact set and $m \in \mathbb{Z}_+$, let

$$\big|\big|\, f \,\big|\big|_{K,m} = \sup_{|\alpha| \leq m} \sup_{x \in K} \big|\partial_x^\alpha \varphi(x)\big| \tag{1.1}$$

Then $\big|\big| \cdot \big|\big|_{K,m}$ is a semi-norm on the space $\mathcal{C}^\infty(U)$. We can define a topology on this space so that a sequence $\{f_n\} \subset \mathcal{C}^\infty(U)$ converges to a limit function $f_0 \in \mathcal{C}^\infty(U)$ if and only if $\lim_{n\to\infty} \big|\big|\, f_0 - f_n \,\big|\big|_{K,m} = 0$ for every compact subset $K \subset \Omega$ and for every non-negative integer m. The space $\mathcal{C}^\infty(U)$ equipped with this topology is denoted by $\mathcal{E}(U)$. It is easy to check that if $\{f_n\} \subset \mathcal{E}(U)$ is a sequence which is Cauchy in every semi-norm $\big|\big| \cdot \big|\big|_{K,m}$, then the sequence converges to a limit $f_0 \in \mathcal{E}(U)$.

If $f \in \mathcal{E}(U)$, the *support* of f is the closure of the set of points $x \in U$ such that $f(x) \neq 0$. The support of f is denoted by suppt(f). Then $\mathcal{C}_0^\infty(U)$ is the subspace of $\mathcal{E}(U)$ consisting of functions with compact support. There is a topology on $\mathcal{C}_0^\infty(U)$ such that a sequence $\{f_1, f_n, \ldots, f_n, \ldots\} \subset \mathcal{C}_0^\infty(U)$ converges to a limit function $f_0 \in \mathcal{C}_0^\infty(U)$ if and only there is a compact set $K \subset U$ so that suppt$(f_n) \subset K$ for all $n \geq 0$ and $\lim_{n\to\infty} \big|\big| f_0 - f_n \big|\big|_{K,m} = 0$ for every non-negative integer m. The space $\mathcal{C}_0^\infty(U)$ with this topology is denoted by $\mathcal{D}(U)$. It should be noted that this topology is not the same as the topology that $\mathcal{C}_0^\infty(U)$ naturally inherits as a subspace of $\mathcal{E}(U)$. However, with the given topologies on the two spaces, the inclusion map $\mathcal{D}(U) \hookrightarrow \mathcal{E}(U)$ is continuous. Note that if $\chi \in \mathcal{C}_0^\infty(U)$, then the mapping $R_\chi : \mathcal{E}(U) \to \mathcal{D}(U)$ given by $R_\chi[f] = \chi\, f$ is continuous.

1.2. Spaces of distributions. The space of *distributions* on U is the dual space $\mathcal{D}'(U)$ of continuous real- or complex-valued continuous linear functionals on $\mathcal{D}(U)$. The pairing between a distribution $T \in \mathcal{D}'(U)$ and a smooth, compactly supported function $\varphi \in \mathcal{D}(U)$ is denoted by $\big\langle T, \varphi \big\rangle$. Of course, if $f \in L^1_{loc}(U)$, then f induces a distribution T_f whose action is given by $\big\langle T_f, \varphi \big\rangle = \int_U f(x)\, \varphi(x)\, dx$. Abusing notation, we shall frequently denote the distribution T_f simply by f. We shall also engage in another, perhaps more serious abuse by sometimes writing the action of a distribution $u \in \mathcal{D}'(U)$ on a function $\varphi \in \mathcal{D}(U)$ by $\int_U u(x)\, \varphi(x)\, dx$.

The space $\mathcal{D}(U)$ is closed under the operation of multiplication by a smooth function in $\mathcal{E}(U)$, and also under differentiation. By duality, this allows us to define the product of a distribution and a smooth function, and to define the derivative of a distribution. Thus let $T \in \mathcal{D}'(U)$ be a distribution. If $a \in \mathcal{E}(U)$ is a smooth function, the product aT is the distribution defined by the requirement that

$$\big\langle aT, \varphi \big\rangle = \big\langle T, a\varphi \big\rangle. \tag{1.2}$$

For every multi-index $\alpha \in \mathbb{Z}^n_+$, the distribution $\partial^\alpha T$ is defined by the requirement that

$$\big\langle \partial^\alpha T, \varphi \big\rangle = (-1)^{|\alpha|} \big\langle T, \partial^\alpha \varphi \big\rangle. \tag{1.3}$$

In particular, if $P = \sum_{|\alpha| \leq M} a_\alpha \partial^\alpha$ is a linear partial differential operator of order M with coefficients $\{a_\alpha\} \subset \mathcal{E}(U)$ and if $u \in \mathcal{D}'(U)$ is a distribution, then $P[u]$ is also a distribution on U.

If $T \in \mathcal{D}'(U)$, the support of T is the smallest closed set $E \subset U$ such that if $\varphi \in \mathcal{C}_0^\infty(U - E)$, then $\big\langle T, \varphi \big\rangle = 0$. We again denote the support of a distribution T by suppt(T).

The dual space $\mathcal{E}'(U)$ is the space of continuous real- or complex-valued linear functionals on $\mathcal{E}(U)$. Since the inclusion map $\mathcal{D}(U) \hookrightarrow \mathcal{E}(U)$ is continuous, every continuous linear functional on $\mathcal{E}(U)$ restricts to a continuous linear functional on $\mathcal{D}(U)$, so there is an induced mapping $\mathcal{E}'(U) \to \mathcal{D}'(U)$. It is not hard to see that this mapping is one-to-one, so $\mathcal{E}'(U)$ is a subspace of distributions on U. In fact, $\mathcal{E}'(U)$ is exactly the space of distributions on U with compact support.

The *Schwartz space* $\mathcal{S}(\mathbb{R}^n)$ is the space of complex-valued infinitely differentiable functions φ defined on $\mathbb{R}^n$ such that for every multi-index α and every positive integer N it follows that for all $x \in \mathbb{R}^n$

$$\big|\partial^\alpha_x \varphi(x)\big| \leq C_{\alpha,\beta}(1 + |x|)^{-N}. \tag{1.4}$$

The collection of semi-norms

$$\big|\big|\,\varphi\,\big|\big|_{\alpha,N} = \sup_{x\in\mathbb{R}^n} \big|(1+|x|)^N \partial_x^\alpha f(x)\big| \tag{1.5}$$

allows us to put a topology on $\mathcal{S}(\mathbb{R}^n)$ so that a sequence $\{\varphi_n\} \subset \mathcal{S}(\mathbb{R}^n)$ converges to a limit φ_0 if and only if $\lim_{n\to\infty} \big|\big|\,\varphi_0 - \varphi_n\,\big|\big|_{\alpha,N} = 0$ for every α and N. It is easy to see that $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$.

For $f \in L^1(\mathbb{R}^n)$, the *Fourier transform* $\mathcal{F}[f]$ is defined by the absolutely convergent integral

$$\mathcal{F}[f](\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i<x,\xi>}\, f(x)\, dx, \tag{1.6}$$

where $< x, \xi >= x_1\xi_1 + \cdots + x_n\xi_n$. If $\varphi \in \mathcal{S}(\mathbb{R}^n)$, one can show that $\mathcal{F}[\varphi] \in \mathcal{S}(\mathbb{R}^n)$, and the Fourier transform is a continuous and invertible linear mapping of $\mathcal{S}(\mathbb{R}^n)$ to itself. The inverse to the Fourier transform is given for $\varphi \in \mathcal{S}(\mathbb{R}^n)$ by the inversion formula

$$\varphi(x) = \mathcal{F}^{-1}[\widehat{\varphi}](x) = \int_{\mathbb{R}^n} e^{2\pi i<x,\xi>}\, \widehat{\varphi}(\xi)\, d\xi. \tag{1.7}$$

If $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n)$, Fubini's theorem shows that

$$\int_{\mathbb{R}^n} \varphi(x)\, \widehat{\psi}(x)\, dx = \int_{\mathbb{R}^n} \widehat{\varphi}(y)\, \psi(y)\, dy. \tag{1.8}$$

Also, if $\varphi \in \mathcal{S}(\mathbb{R}^n)$, the Plancherel formula asserts that

$$\big|\big|\,\varphi\,\big|\big|^2_{L^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\varphi(x)|^2\, dx = \int_{\mathbb{R}^n} |\widetilde{\varphi}(\xi)|^2\, d\xi = \big|\big|\,\widehat{\varphi}\,\big|\big|^2_{L^2(\mathbb{R}^n)}, \tag{1.9}$$

and it follows that the Fourier transform extends by continuity to an isometry of $L^2(\mathbb{R}^n)$.

Define $M^s[\varphi](x) = (1 + 4\pi^2|x|^2)^{\frac{s}{2}}$ if $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Then it is easy to check that M^s maps the space $\mathcal{S}(\mathbb{R}^n)$ to itself, and is continuous and invertible with inverse M^{-s}. Hence we can define a mapping $\Lambda^s : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ by setting $\Lambda^s = \mathcal{F}^{-1}M^s\mathcal{F}$, or explicitly

$$\widehat{\Lambda^s[\varphi]}(\xi) = (1 + 4\pi^2|\xi|^2)^{\frac{s}{2}}\, \widehat{\varphi}(\xi) \tag{1.10}$$

The space of *tempered distributions* $\mathcal{S}'(\mathbb{R}^n)$ is the space of continuous real- or complex-valued linear functionals on $\mathcal{S}(\mathbb{R}^n)$. It is clear that every $f \in L^2(\mathbb{R}^n)$ defines a tempered distribution by setting $\big\langle T_f, \varphi\big\rangle = \int_{\mathbb{R}^n} f(x)\, \varphi(x)\, dx$ for $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Since the Fourier transform is continuous and invertible on $\mathcal{S}(\mathbb{R}^n)$, it is possible to extend the defintion of the Fourier transform to the space of tempered distributions by setting

$$\big\langle \widehat{T}, \varphi\big\rangle = \big\langle T, \widehat{\varphi}\big\rangle \tag{1.11}$$

for $T \in \mathcal{S}'(\mathbb{R}^n)$ and $\varphi \in \mathcal{S}(\mathbb{R}^n)$. We can also extend the mapping Λ^s to the space $\mathcal{S}'(\mathbb{R}^n)$ by setting

$$\big\langle \Lambda^s T, \varphi\big\rangle = \big\langle T, \Lambda^s\varphi\big\rangle. \tag{1.12}$$

Note that we then have

$$\big\langle \widehat{\Lambda^s T}, \varphi\big\rangle = \big\langle \Lambda^s T, \widehat{\varphi}\big\rangle = \big\langle T, \Lambda^s\widehat{\varphi}\big\rangle = \tag{1.13}$$

For every $s \in \mathbb{R}^n$, the *Sobolev space* $H^s(\mathbb{R}^n)$ is the subspace of tempered distributions $T \in \mathcal{S}'(\mathbb{R}^n)$ such that $\Lambda^s T \in L^2(\mathbb{R}^n)$. This means that $\widehat{T}$ is given by a locally integrable function and

$$\big|\big|\, T \,\big|\big|_s^2 = \int_{\mathbb{R}^n} |\widehat{T}(\xi)|^2 (1 + 4\pi^2|\xi|^2)^s \, d\xi < \infty. \tag{1.14}$$

In particular. $H^0(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, and we often write the norm of $f \in L^2(\mathbb{R}^n)$ as $\big|\big|\, f \,\big|\big|_{L^2(\mathbb{R}^n)} = \big|\big|\, f \,\big|\big|_0$.

2. Pseudodifferential Operators

The space S^m of standard symbols of order $m \in \mathbb{R}$ on $\mathbb{R}^n$ is the space of functions $a \in \mathcal{C}^\infty(\mathbb{R}^n)$ with the property that for every pair of multi-indices α and β there is a constant $C_{\alpha,\beta}$ so that for all $(x,\xi) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\big|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)\big| \leq C_{\alpha,\beta}(1 + |\xi|)^{m-|\alpha|}. \tag{2.1}$$

The collection of norms

$$p_{\alpha,\beta}(a) = \sup_{(x,\xi)\in\mathbb{R}^n\times\mathbb{R}^n} (1 + |\xi|)^{-m+|\alpha|} \big|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)\big| \tag{2.2}$$

make S^m into a Fréchet space.

If $a \in S^m$, the pseudodifferential operator $a(x,D)$ with symbol a defined on the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is given by

$$a(x,D)[f](x) = \int_{\mathbb{R}^n} e^{-2\pi i <x,\xi>} \, a(x,\xi) \, \widehat{f}(\xi) \, d\xi. \tag{2.3}$$

This integral converges absolutely since $f \in \mathcal{S}$ implies that the Fourier transform $\widehat{f} \in \mathcal{S}(\mathbb{R}^n)$ and thus has rapid decay. It is easy to check that $a(x,D)[f]$ is infinitely differentiable and all derivatives have rapid decay. Thus $a(x,D) : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$. By duality, $a(x,D)$ defines a mapping $\mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$, and so $a(x,D)$ is defined on the space of tempered distributions.

We denote the space of pseudodifferential operators of order m by $OP^m(\mathbb{R}^n)$, and we set

$$OP(\mathbb{R}^n) = \bigcup_{m\in\mathbb{R}} OP^m(\mathbb{R}^n), \tag{2.4}$$

and

$$OP^{-\infty}(\mathbb{R}^n) = \bigcap_{m\in\mathbb{R}} OP^m(\mathbb{R}^n). \tag{2.5}$$

If $A \in OP(\mathbb{R}^n)$, we write $o(A)$ for the order of the operator A.

Let $a \in S^m$. For each $\epsilon > 0$, set

$$K_a^\epsilon(x,y) = \int_{\mathbb{R}^n} e^{-2\pi i <x-y,\xi>} \, a(x,\xi) \, d\xi. \tag{2.6}$$

The integral is absolutely convergent. If $\chi \in \mathcal{C}(\mathbb{R}^n \times \mathbb{R}^n)$, then the expression

$$\big\langle K, \chi \big\rangle = \lim_{\epsilon\to 0} \int_{\mathbb{R}^n\times\mathbb{R}^n} K_a^\epsilon(x,y) \, \chi(x,y) \, dx \, dy \tag{2.7}$$

exists, and defines a distribution K_a on $\mathbb{R}^n \times \mathbb{R}^n$, which is the Schwartz kernel for the operator $a(x, D)$. That is, if $\varphi, \psi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} a(x,D)[\varphi](x)\,\psi(x)\,dx = \lim_{\epsilon \to 0} \iint_{\mathbb{R}^n \times \mathbb{R}^n} K_a^\epsilon(x,y)\varphi(y)\,\psi(x)\,dx\,dy = \big\langle K_a, \varphi \otimes \psi \big\rangle.$$

We define a family of symbols $\lambda^s \in S^s$ by

$$\lambda^m(x,\xi) = (1 + 4\pi^2|\xi|^2)^{\frac{s}{2}}. \tag{2.8}$$

The corresponding pseudodifferential operator is denoted by Λ^s, and we have

$$\widehat{\Lambda^s[f]}(\xi) = (1 + 4\pi^2|\xi|^2)^{\frac{s}{2}}\,\widehat{f}(\xi) \tag{2.9}$$

for each $f \in \mathcal{S}(\mathbb{R}^n)$. When s is a positive even integer $s = 2m$, the Fourier inversion formula shows that

$$\Lambda^{2m}[f] = (1 - \triangle_x)^m[f]. \tag{2.10}$$

For any $s \in \mathbb{R}$ we define the Sobolev space $H^s(\mathbb{R}^n)$ to be the subspace of tempered distributions $u \in \mathcal{S}'(\mathbb{R}^n)$ such that $\Lambda^s[u] \in L^2(\mathbb{R}^n)$. For $u \in H^s(\mathbb{R}^n)$, set

$$||u||_s = \big|\big|\,\Lambda^s[u]\,\big|\big|_{L^2(\mathbb{R}^n)} = \big|\big|\,\Lambda^s[u]\,\big|\big|_0. \tag{2.11}$$

2.1. Results.

2.1.1. *The symbolic calculus.*

The space $OP(\mathbb{R}^n)$ is an algebra under composition, and is closed under taking adjoints.

THEOREM 2.1. *Let $a_j \in S^{m_j}$ for $j = 1, 2$. Then there is a symbol $b \in S^{m_1+m_2}$ so that*

$$a_1(x,D) \circ a_2(x,D) = b(x,D). \tag{2.12}$$

Moreover, the symbol b has an asymptotic expansion so that

$$b - \sum_{|\alpha|<N} \frac{1}{(2\pi i)^{|\alpha|}\,\alpha!} \frac{\partial^\alpha a_1}{\partial \xi^\alpha} \frac{\partial^\alpha a_2}{\partial x^\alpha} \in S^{m_1+m_2-N}. \tag{2.13}$$

COROLLARY 2.2. *If $a_j \in S^{m_j}$ for $j = 1, 2$, the commutator*

$$[a_1(x,D),\, a_2(x,D)] = a_1(x,D) \circ a_2(x,D) - a_2(x,D) \circ a_1(x,D) \tag{2.14}$$

is a pseudodifferential operator of order $m_1 + m_2 - 1$.

THEOREM 2.3. *Let $a \in S^m$. Then there is a symbol $b \in S^m$ so that*

$$a(x,D)^* = b(x,D). \tag{2.15}$$

More precisely, if $f, g \in \mathcal{S}(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} a(x,D)[f](x)\,\overline{g(x)}\,dx = \int_{\mathbb{R}^n} f(x)\,\overline{b(x,D)[g](x)}\,dx. \tag{2.16}$$

Moreover, the symbol b has an asymptotic expansion so that

$$b - \sum_{|\alpha|<N} \frac{1}{(2\pi i)^{|\alpha|}\,\alpha!} \frac{\partial^{2\alpha}\overline{a}}{\partial \xi^\alpha \partial x^\alpha} \in S^{m-N}. \tag{2.17}$$

2.1.2. *Continuity on Sobolev spaces.*

Pseudodifferential operators induce bounded operators between appropriate Sobolev spaces.

THEOREM 2.4. *Let $a \in S^0$. Then $a(x,D) : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n$ is a bounded operator, with norm $\big|\big| a(x,D) \big|\big|_0$ depending on finitely many of the semi-norms $\{p_{\alpha,\beta}(a)\}$.*

It follows from Theorem 2.1 that if $a \in S^m$, then the pseudodifferential operators $\Lambda^s\, a(x,D)$ and $a(x,D)\,\Lambda^s$ are pseudodifferential operators of order $m+s$. Combining this with the definition of the Sobolev spaces $H^s(\mathbb{R}^n)$ and Theorem 2.3, we have

COROLLARY 2.5. *Let $b \in S^m$. Then for every $s \in \mathbb{R}$ the operator $b(x,D)$ is a bounded operator from $H^s(\mathbb{R}^n)$ to $H^{s-m}(\mathbb{R}^n)$. The norm of this operator depends on s, m, and on finitely many of the semi-norms $p_{\alpha,\beta}(b)$.*

2.1.3. *Pseudolocality.*

THEOREM 2.6. *If $a \in S^m$, the Schwartz kernel K_a for the operator $a(x,D)$ is of class $\mathcal{C}^\infty$ away from the diagonal of $\mathbb{R}^n \times \mathbb{R}^n$. Moreover, for all multi-indices α and β and every $N \geq 0$ with $m+n+|\alpha|+|\beta|+N > 0$, if $x \neq y$,*

$$\big|\partial_x^\alpha \partial_y^\beta K_a(x,y)\big| \leq C_{\alpha,\beta,N}\, |x-y|^{-n-m-|\alpha|-|\beta|-N}. \tag{2.18}$$

DEFINITION 2.7. *Let $a \in S^m$. The operator $a(x,D)$ is properly supported if for every $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, the functions $K_a(x,y)\,\varphi(y)$ and $K_a(x,y)\,\varphi(x)$ have compact support in $\mathbb{R}^n \times \mathbb{R}^n$. The space of properly supported pseudodifferential operators is denoted by $OP_p(\mathbb{R}^n)$.*

PROPOSITION 2.8. *If $a_1(x,D)$ and $a_2(x,D)$ are properly supported pseudodifferential operators, then so are $a_1(x,D) \circ a_2(x,D)$ and $a_j(x,D)^*$. If $a \in S^m$, there exists $\widetilde{a} \in S^m$ so that $\widetilde{a}(x,D)$ is properly supported and $a - \widetilde{a} \in OP^{-\infty}(\mathbb{R}^n)$.*

Bibliography

[BJS64] Lipman Bers, Fritz John, and Martin Schechter. *Partial Differential Equations*, volume III of *Lectures in Applied Mathematics*. Interscience Publishers, JohnWiley & Sons, 1964.

[Bou89] N. Bourbaki. *Lie Groups and Lie Algebras: chapters 1-3*. Elements of Mathematics. Springer-Verlag, Berlin, 1989.

[BR96] A. Bellaïche and J.-J. Risler. *Sub-Riemannian Geometry*, volume 144 of *Progress in Mathematics*. Birkhauser, Boston, 1996.

[Cat84] David W. Catlin. Global regularity of the $\overline{\partial}$-Neumann problem. In Yum-Tong Siu, editor, *Complex Analysis of Several Variables*, volume 41 of *Proceeding of Symposia in Pure Mathematics*, pages 39–49, Providence, Rhode Island, 1984. American Mathematical Society.

[Cho40] W.L. Chow. Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung. *Math. Ann.*, 117:85–115, 1940.

[CS01] So-Chin Chen and Mei-Chi Shaw. *Partial Differential Equations in Several Complex Variables*, volume 19 of *Studies in Advanced Mathematics*. American Mathematical Society, 2001.

[CW77] R. R. Coifman and G. Weiss. *Transference methods in Analysis*, volume 31 of *CBMS Reg. Conf. Series in Math.* American Mathematical Society, 1977.

[dG75] Miguel de Guzmán. *Differentiation of Integrals in* $\mathbb{R}^n$. Number 481 in Lecture Notes in Mathematics. Springer-Verlag, 1975.

[FK72] G.B. Folland and J.J. Kohn. *The Neumann problem for the Cauchy-Riemann Complex*, volume 75 of *Ann. of Math. Studies*. Princeton University Press, Princeton, New Jersey, 1972.

[Fol84] Gerald B. Folland. *Real Analysis*. John Wiley & Sons, 1984.

[Fri34] K. Friedrichs. Spektraltheorie halbbeschränkter Operatoren. *Math. Annalen*, 109:465–487, 685–713, 1934.

[FS74] G.B. Folland and E.M. Stein. Estimates for the $\overline{\partial}_b$-complex and analysis on the Heisenberg group. *Comm. Pure and Appl. Math.*, 27:429–522, 1974.

[GR65] Robert C. Gunning and Hugo Rossi. *Analytic Functions of Several Complex Variables*. Prentice Hall, Englewood Cliffs, NJ, 1965.

[Gru71] V.V. Grushin. A certain class of elliptic pseudodifferential operators that are degenerate on a submanifold. *Mat. Sb. (N.S.)*, 126:163–195, 1971.

[GT83] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 224 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1983.

[H6̈6] L. Hörmander. *An Introduction to Complex Analysis in Several Variables*. Van Nostrand, 1966.

[H6̈7] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[H8̈5] L. Hörmander. *The Analysis of Linear Partial Differential Operators I - IV*, volume 256, 257, 274, 275 of *Grund. der math. Wiss.* Springer-Verlag, 1983, 1985.

[H9̈4] L. Hörmander. *Notions of Convexity*, volume 127 of *Progress in Mathematics*. Birkhäuser, Bosten, 1994.

[Hal50] M. Hall. A basis for free Lie rings and higher commutators in free groups. *Proceedings of the AMS*, 1(5):575–581, 1950.

[Hum72] J.E. Humphreys. *Introduction to Lie Algebras and Representation Theory*. Springer-Verlag, New York, 1972.

[Jac62] N. Jacobson. *Lie Algebras*. Dover Publications, New York, 1962.

[Koe00] K. Koenig. *On maximal Sobolev and Hölder estimates for the tangential Cauchy-Riemann operator and boundary Laplacian.* PhD thesis, Princeton University, Department of Mathematics, 2000.

[Koe02] K. Koenig. On maximal Sobolev and Hölder estimates for the tangential Cauchy-Riemann operator and boundary Laplacian. *Am. J. Math*, 124:129–197, 2002.

[Nar85] R. Narasimhan. *Analysis on Real and Complex Manifolds.* North-Holland Mathematical Library. Elseevier Science Publishers, Amsterdam, New York, Oxford, 1985.

[NSW85] A. Nagel, E.M. Stein, and S. Wainger. Balls and metrics defined by vector fields I: Basic properties. *Acta Math.*, 155:103–147, 1985.

[Sta93] J. Stalker. *Hölder and L^p estimates for $\overline{\partial}$, $\overline{\partial}_b$ on domains of finite type.* PhD thesis, Princeton University, Department of Mathematics, 1993.

[Ste70] Elias M. Stein. *Singular Integrals and Differentiability Properties of Functions.* Princeton Mathematical Series; 30. Princeton University Press, Princeton, New Jersey, 1970.

[Ste72] E.M. Stein. *Boundary Behavior of Holomorphic Functions of Several Complex Variables.* Princeton University Press, Princeton, New Jersey, 1972.

[Ste93] Elias M. Stein. *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals.* Princeton Mathematical Series; 43. Princeton University Press, Princeton, New Jersey, 1993.

[SW71] Elias M. Stein and Guido Weiss. *Fourier Analysis on Euclidean Spaces.* Princeton University Press, Princeton, New Jersey, 1971.

[Tre80] François Treves. *Introduction to Pseudodifferential and Fourier Integral Operators.* Plenum Press, New York, 1980.