Last updated: 2025-11-24

888 Algebraic Foundations of Data (Learning outcomes)

Instructor: Jose Israel Rodriguez, Course website

A preview of the learning outcomes

Many optimization problems in data and learning are built on algebraic ideas. For example, principal component analysis finds a low-rank approximation of a matrix, a problem central to linear algebra. This course builds out from this example to study the algebraic foundations of several optimization problems to find representations and understanding of data.

Topics[1]

There are many topics listed here, some of which entire courses can be devoted to. Our goal in this course is to identify important problems for motivation and to highlight algebraic foundations of data. Lectures will be important to achieving this goal. Textbook supplementary references are provided for those who want to build up their background, and research articles are listed to point to areas that could be exciting to dive into. There are six required homeworks to cement understanding by engaging with the material at the start of the semester. As the topics become more specialized optional challenge problems will be posted to inspire new ideas and approaches.

Part one of the course is a linear algebra review. The second part of course uses SVD as a spring board into several different topics including Procrustes problems. From Topic 13 and onward, a selection of these topics will be made based on time permitting. In a perfect world, we would cover all of these, but the more likely scenario is to cover half.

This list is continuously updated to match the pace of the class:

- Topic 1: Least squares
  Supplementary References: VMLS–Chapter 12
  Outcomes: Derive several expressions for the solution of the least squares problem; review QR factorization; Pseudo-inverses and how to compute them from a QR factorization
  Lecture: 1

[1] Topics are subject to change.

- Topic 2: Least squares data fitting
  Supplementary References: VMLS-Chapter 13.1-13.2
  Outcomes: Find a mathematical model, or an approximate model, of some relation, given some observed data; Fitting univariate functions; polynomial fit; regression; validation and overfitting; AR model example (page 259, 266-268); Vandermonde matrices.
  Lecture: 2, 3

- Topic 3: Eigenvalues of symmetric matrices and Rayleigh–Ritz Quotients
  Supplementary References: ALA–Chapter 8.5 Chebfun resource; Guangliang Chen's Notes
  Outcomes: Quadratic forms, optimizing a quadratic form over the unit ball, power iteration, useful properties of derivaties in linear algebra; power iteration (*ALA* Chapter 9.5), generalized Raleigh-Ritz Quotients
  Lecture: 3, 4

- Topic 4: Singular value decomposition for matrices
  Supplementary References: ALA–Chapter 8.7; NLA–Chapter 1 (Lecture 4); Stanford notes; Lim's article; additional article for $2 \times 2 \times 2$ setting; best rank one approximation and ill-posedness results for tensors; Tensors in computation.
  Outcomes: An optimization principle for SVD, connection to pseudo-inverse; geometric interpretation of SVD; singular values of tensors; issues in generalizing SVD to tensors.
  Lecture: 5, 6 (fire drill), 7

- Topic 5: Principle component analysis
  Supplementary References: ALA–Chapter 8.8; EDRML–Chapter 5.1-5.2, Tutorial
  Outcomes: Realizing how to compute principle components using SVD; framing PCA as an optimization problem; variance and covariance.
  Lecture: 8

- Topic 6: Procrustes Problems: Satellite rotations (Wahba's problem)
  Supplementary References: APAM-PCV– Chapter 2
  Outcomes: Putting Procrustes problems into practice ; Generalized version of Wahba's problem; a historical note by Wahba and more;
  Lecture: 9

- Topic 7: Matrix groups and manifolds (Part one)
  Supplementary References: Krishnamurthy notes; MGU–Chapter 3-7 (Orthogonal group; topology of matrix groups; importance of closed sets; tangent space and Lie Algebras; dimension of several matrix groups; matrix exponentiation; manifolds)
  Outcomes: Review matrix groups; background knowledge to dive deeper into manifold learning
  Lecture: 10, 11

- Topic 8: A snapshot of algebra for computer vision (Part one)
  Supplementary References: [18]; Stanford course
  Outcomes: Projective space; triangulation; rational maps; pinhole cameras; motivating implicitization.
  Lecture: 12

- Topic 9: First order optimization (Part 1)
  Supplementary References: Embedded Geometry [4, Chapter 3-4]; EDRML–Chapter 4.1-4.4, Chapter 5.4
  Outcomes: Definition of smooth embedded manifolds; calculus review of differentials and geometric intuition; definition of retraction.
  Lecture: 13

- Topic 10: Matrix groups and manifolds (Part two)
  Supplementary References: Krishnamurthy notes; MGU–Chapter 3-7 (Orthogonal group; topology of matrix groups; importance of closed sets; tangent space and Lie Algebras; dimension of several matrix groups; matrix exponentiation; manifolds)
  Outcomes: Matrix exponential, one parameter subgroups;
  Lecture: 14

- Topic 11: First order optimization (Part 2)
  Supplementary References: [4, Chapter 4].
  Outcomes: Riemmannian gradients, Riemannian gradient descent
  Lecture: 15, 16

- Topic 12: Critical points: A computational algebra perspective
  Supplementary References: [10, ED degree paper]
  Outcomes: Euclidean distance degrees; low rank matrices, a quartic surface https://www.desmos.com/3d/ovasz9zcau; translating smooth manifold optimization into computational algebra
  Lecture: 17, 18

- Topic 13: Eigenvalue methods polynomial system solving: multiplication maps
  Supplementary References: UAG–Division Algorithm, Chapter 2.2 and 2.4 (Theorem 2.4.5)
  Outcomes: Multivariate polynomial division; cosets; Finiteness Theorem (a,b); Solving polynomial systems via eigenvalues.
  Lecture: 18, 19

- Topic 14: A snapshot of algebra for computer vision (Part 2)
  Supplementary References: [18]; Stanford course
  Outcomes: Triangulation, relating ED degrees to sizes of multiplication maps
  Lecture: 20

- Topic 15: Feedforward neural networks viewed as tensors
  Supplementary References: [19]; [11] INA–Chapter 9; Online resource

Outcomes: Identifiability of polynomial neural networks; minimal problems; filling versus thick polynomial neural networks; [13, 28, 29, 11];
Lecture: 21, 22

- Topic 16: Projection and elimination theory: semi-algebraic versus algebraic
Supplementary References: UAG, INA–Chapter 4
Outcomes: Generalizing linear algebra concepts of REF to the nonlinear algebra setting
Lecture: 23

- Topic 17: Attention mechanisms and their neuromanifolds
Supplementary References: [23, 22, 14].
Outcomes: Saw when neuromanifolds are algebraic (lightning) and semialgebraic (Relu); stated dimension results for lightning attention;
Lecture: 24, 25

- Topic 18: From Transformers to splines and back again
Supplementary References: [23, 22, 14].
Outcomes: Built a framework of functions that underlie the mathematics of a transformer; saw how this new technology is related to splines
Lecture: 25, 26

- Topic 19: Elimination for phylogenetics — curves
Supplementary References: UAG and AS-Sullivant
Outcomes: We gave an example on how to calculate the maximum likelihood degree [**?**] of a parameterize curve. This was a special case of the maximum likelihood degree of a log-linear model.
Lecture: 27

- Topic 20: EM Algorithm and maximum likelihood estimation
Supplementary References: [20]; [1, Chapter 1.3]; AS-Sullivant; [26, 25, 17, 16, 5]
Outcomes: We say the difference between non-negative rank and how it impacts maximum likelihood estimation. We compared and contrasted the maximum likelihood degree (counting a finite set of points in the relative interior) and fixed points of the EM algorithm that live on the models boundary. We follows through carefully an example of the EM algorithm appearing in the *Annals of Statistics* [20] on the Naive Bayes model. We saw how algebra leads to understanding why local hill climbing algorithms can fail.
Lecture: 28, 29

- Topic 21: Elimination for phylogenetics — trees and networks
Supplementary References: UAG and AS-Sullivant
Outcomes: Division algorithm revisited; graphs and their projections
Lecture: 30, 31

- Topic 22: Grassmannians and likelihood geometry of determinantal point process
  Supplementary References: [8, 12] [15, 24, 2] For more background on Plucker coordinates see page 454 Chapter 8 in [7] (2025 edition). For more details on Grassmannians see Example 4.9 and Chapter 5 in INA.
  Outcomes: Projected DPP; apply likelihood geometry. In class we centered on these three references [12],[9], [21] Lecture: 32, 33

- Topic 23: Tropical maximum likelihood estimation
  Supplementary References: INA–Chapter 7, [3, 27]
  Outcomes: Extremal statistics through an algebra lens
  Lecture: 34, 35

- Topic 24: Continuous maximization for inverse spectral problems
  Supplementary References: [6]
  Outcomes: Bridge between convex and nonconvex optimization; Stiefel manifolds
  Lecture: 36

- Topic 25: Algebraic optimization degrees and data discriminants
  Supplementary References: TBD
  Outcomes: Computing data discriminants; connections to physics
  Lecture: 37

- Topic 26: Kernel Methods
  Supplementary References: EDRML– Chapter 3
  Outcomes: Kernel methods
  Lecture: 38, 39

- Topic 27: Kernel Principle component analysis
  Supplementary References: EDRML–Chapter 5.4
  Outcomes: Kernel methods
  Lecture: 40, 41

- Topic 28: Expressiveness of Relu neural networks
  Supplementary References: INA–Chapter 7; [30]
  Outcomes: Expressive power of Relu neural in terms of tropical geometry
  Lecture: 42, 43

- Topic 29: Orthogonally decomposable tensors
  Supplementary References: TBD
  Outcomes: Alternative ways to compress data
  Lecture: 44

- Topic 30: Sums of squares and SDP relaxations (Time permitting)
  Supplementary References: INA— Chapter 12
  Outcomes: Expanding the algebraic optimization toolkit
  Lecture: 45, 46

- Topic 31: Log-linear models and exact tests (Time permitting)
  Supplementary References: TBD
  Outcomes: Quantifying uncertainty using algebraic methods
  Lecture: 47, 48

References

[1] *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.

[2] G. Baverez, A. I. Bufetov, and Y. Qiu, *A survey on determinantal point processes*, in Stochastic analysis, random fields and integrable probability—Fukuoka 2019, vol. 87 of Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, [2021] ©2021, pp. 59–88. DOI.

[3] E. Boniface, K. Devriendt, and S. Hoşten, *Tropical toric maximum likelihood estimation*, 2025. arXiv: 2404.10567.

[4] N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, Cambridge, 2023.

[5] F. Catanese, S. Ho¸sten, A. Khetan, and B. Sturmfels, *The maximum likelihood degree*, Amer. J. Math., 128 (2006), pp. 671–697. URL.

[6] V. Chandrasekaran, T. Duff, J. I. Rodriguez, and K. Shu, *Lagrangian dual sections: A topological perspective on hidden convexity*, (2025). arXiv: 2510.06112.

[7] D. A. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, vol. 185 of Graduate Texts in Mathematics, Springer, New York, second ed., 2005.

[8] K. Devriendt, H. Friedman, B. Reinke, and B. Sturmfels, *The two lives of the Grassmannian*, Acta Univ. Sapientiae Math., 17 (2025), p. 8. DOI.

[9] K. Devriendt, H. Friedman, B. Reinke, and B. Sturmfels, *The two lives of the Grassmannian*, Acta Univ. Sapientiae Math., 17 (2025), pp. Paper No. 8, 18. DOI.

[10] J. Draisma, E. Horobe¸t, G. Ottaviani, B. Sturmfels, and R. R. Thomas, *The Euclidean distance degree of an algebraic variety*, Found. Comput. Math., 16 (2016), pp. 99–149. DOI.

[11] B. Finkel, J. I. Rodriguez, C. Wu, and T. Yahl, *Activation degree thresholds and expressiveness of polynomial neural networks*, 2025. arXiv: 2408.04569.

[12] H. Friedman, B. Sturmfels, and M. Zubkov, *Likelihood geometry of determinantal point processes*, Algebr. Stat., 15 (2024), pp. 15–25. DOI.

[13] A. Grosdos, E. Robeva, and M. Zubkov, *Algebraic geometry of rational neural networks*, (2025). arXiv: 2509.11088.

[14] N. W. Henry, G. L. Marchetti, and K. Kohn, *Geometry of lightning self-attention: Identifiability and dimension*, (2025). arXiv: 2408.17221.

[15] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Determinantal processes and independence*, Probab. Surv., 3 (2006), pp. 206–229. DOI.

[16] J. Huh, *The maximum likelihood degree of a very affine variety*, Compos. Math., 149 (2013), pp. 1245–1266. DOI.

[17] J. Huh, *Varieties with maximum likelihood degree one*, J. Algebr. Stat., 5 (2014), pp. 1–17. DOI.

[18] J. Kileel and K. Kohn, *Snapshot of algebraic vision*, 2023. arXiv: 2210.11443.

[19] J. Kileel, M. Trager, and J. Bruna, *On the expressive power of deep polynomial neural networks*, in Advances in Neural Information Processing Systems, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds., vol. 32, Curran Associates, Inc., 2019. URL.

[20] K. Kubjas, E. Robeva, and B. Sturmfels, *Fixed points EM algorithm and nonnegative rank boundaries*, Ann. Statist., 43 (2015), pp. 422–461. DOI.

[21] A. Kulesza and B. Taskar, *Determinantal Point Processes for Machine Learning*, Foundations and trends in machine learning, Now Publishers, 2012. URL.

[22] Z. Lai and L.-H. Lim, *Pierce-birkhoff conjecture is true for splines*, (2025). arXiv: 2507.07976.

[23] Z. Lai, L.-H. Lim, and Y. Liu, *Attention is a smoothed cubic spline*, (2024). arXiv: 2408.09624.

[24] R. Lyons, *Determinantal probability: basic properties and conjectures*, in Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV, Kyung Moon Sa, Seoul, 2014, pp. 137–161.

[25] J. I. Rodriguez and X. Tang, *A probabilistic algorithm for computing data-discriminants of likelihood equations*, J. Symbolic Comput., 83 (2017), pp. 342–364. DOI.

[26] J. I. Rodriguez and B. Wang, *The maximum likelihood degree of mixtures of independence models*, SIAM J. Appl. Algebra Geom., 1 (2017), pp. 484–506. DOI.

[27] A.-L. Sattelberger and R. van der Veer, *Maximum likelihood estimation from a tropical and a Bernstein-Sato perspective*, Int. Math. Res. Not. IMRN, (2023), pp. 5263–5292. DOI.

[28] V. Shahverdi, G. L. Marchetti, and K. Kohn, *Learning on a razor's edge: the singularity bias of polynomial neural networks*, (2025). arXiv: 2505.11846.

[29] K. Usevich, C. Dérand, R. Borsoi, and M. Clausel, *Identifiability of deep polynomial neural networks*, (2025). arXiv: 2506.17093.

[30] L. Zhang, G. Naitzat, and L. Lim, *Tropical geometry of deep neural networks*, CoRR, abs/1805.07091 (2018). URL.