

Unique reconstruction for discretized inverse problems: a random sketching approach via subsampling

Ruhui Jin^{1,*} , Qin Li¹ , Anjali Nair²
and Samuel N Stechmann^{1,3} 

¹ Department of Mathematics, University of Wisconsin-Madison, Madison, WI, United States of America

² Department of Statistics and CCAM, University of Chicago, Chicago, IL, United States of America

³ Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, United States of America

E-mail: rjin@math.wisc.edu, qinli@math.wisc.edu, anjalinair@uchicago.edu
and stechmann@wisc.edu

Received 8 March 2024; revised 4 February 2025

Accepted for publication 17 March 2025

Published 2 April 2025



CrossMark

Abstract

Computational inverse problems utilize a finite number of measurements to infer a discrete approximation of the unknown parameter function. With motivation from the setting of PDE-based optimization, we study the unique reconstruction of discretized inverse problems by examining the positivity of the Hessian matrix. What is the reconstruction power of a fixed number of data observations? How many parameters can one reconstruct? Here we describe a probabilistic approach, and spell out the interplay of the observation size (r) and the number of parameters to be uniquely identified (m). The technical pillar here is the random sketching strategy, in which the matrix concentration inequality and sampling theory are largely employed. By analyzing a randomly subsampled Hessian matrix, we attain a well-conditioned

* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

reconstruction problem with high probability. Our main theory is validated in numerical experiments, using an elliptic inverse problem as an example.

Supplementary material for this article is available [online](#)

Keywords: inverse problem, PDE-constrained optimization, well-posedness, condition number of Hessian matrix, matrix sketching, random subsampling

1. Introduction

In many PDE-based inverse problems, measurements of a PDE model are used to infer unknown parameters (often a function) of a PDE. A very common numerical strategy is to use the framework of the PDE-constrained optimization.

Typically we are to minimize a loss function, defined as the mismatch between PDE-simulated data and the measured data:

$$\min_{\sigma \in L^2(\Omega)} \text{Loss}[\sigma] = \frac{1}{2} \|\mathcal{M}[\sigma] - d\|_{L^2(\mathcal{D})}^2, \quad \text{where } d = \mathcal{M}[\sigma_*] \in L^2(\mathcal{D}), \quad (1)$$

where $\sigma(x) \in L^2(\Omega)$ is the parameter function to be inferred with x supported in Ω (assumed to be compactly supported and with C^1 boundary). The mismatch is quantified by the functional $\text{Loss} : L^2(\Omega) \rightarrow \mathbb{R}^+$. The parameter-to-observation map is $\mathcal{M} : L^2(\Omega) \rightarrow L^2(\mathcal{D})$, where the measured location is denoted by the set $\mathcal{D}$ and is not necessarily the same as Ω . We are to compare this measured PDE-simulated solution with the true measurement data $d(x)$ in $L^2(\mathcal{D})$ norm, assuming the data has no noise pollution, i.e. $d = \mathcal{M}[\sigma_*]$. Since both the to-be-reconstructed parameter $\sigma : \Omega \rightarrow \mathbb{R}$, and the available data $d : \mathcal{D} \rightarrow \mathbb{R}$ are functions, the problem is posed in the infinite-to-infinite dimensional setting.

In computations, this formulation is adjusted to fit a discrete setting. In practice, let the basis of interest be a subset of a complete basis: $\text{span}\{\phi_1, \phi_2, \dots, \phi_N\} \subseteq L^2(\Omega)$. Then the to-be-reconstructed function $\sigma \in L^2(\Omega)$ is approximated by $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_N] \in \mathbb{R}^N$, with $\sigma \approx \sum_{i=1}^N \sigma_i \phi_i$. The number of available data points is also finite. Letting $\{\psi_j\}_{j=1}^r$ be the collection of test functions used to detect the PDE solution, the discrete finite-to-finite dimensional problem is reformulated as:

$$\min_{\boldsymbol{\sigma} \in \mathbb{R}^N} \text{loss}(\boldsymbol{\sigma}) = \sum_{j=1}^r \frac{1}{2} \underbrace{(\langle \mathcal{M}(\boldsymbol{\sigma}), \psi_j \rangle - \mathbf{d}_j)^2}_{\text{loss}_j}, \quad \text{where } \mathbf{d}_j = \langle \mathcal{M}(\boldsymbol{\sigma}_*), \psi_j \rangle_{\mathcal{D}} \in \mathbb{R}. \quad (2)$$

Here we abuse the notation and set $\mathcal{M}(\boldsymbol{\sigma}) = \mathcal{M}[\sum_{i=1}^N \sigma_i \phi_i]$. The collected data points $\{\mathbf{d}_j\}$ are assumed to be noise-free, and thus have the form of $\mathbf{d}_j = \langle \mathcal{M}(\boldsymbol{\sigma}_*), \psi_j \rangle$ where $\boldsymbol{\sigma}_*$ is the ground-truth media. In this case, the ground-truth $\boldsymbol{\sigma}^*$ is a global optimum to (2), since it achieves zero-loss: $\text{loss}(\boldsymbol{\sigma}_*) = 0$. The notation of ‘loss : $\mathbb{R}^N \rightarrow \mathbb{R}^+$ ’ is no longer a functional, but a real-valued function.

1.1. Motivation and objective of this work

On the surface, the continuous (1) and the discrete (2) problems are rather similar. However, their theoretical behaviors are actually drastically different. In particular, close to the ground-truth $\boldsymbol{\sigma}_*$, by deploying the unique reconstruction theory proved in the continuous setting, we can deduce that the Hessian of the cost functional is not degenerate, ensuring that $\boldsymbol{\sigma}_*$ is the unique optimum point in the neighborhood. We denote the Hessian of the functional (1) at $\boldsymbol{\sigma}_*$

as $\mathcal{H}\text{Loss}[\sigma_*] : L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}$. With mild assumptions (see supplementary material), one deduces:

$$\mathcal{H}\text{Loss}[\sigma_*] \succ 0, \quad \text{or equivalently, for any } \tilde{\sigma} \neq 0 \in L^2(\Omega), \quad \mathcal{H}\text{Loss}[\sigma_*](\tilde{\sigma}, \tilde{\sigma}) > 0. \quad (3)$$

As a consequence, σ_* is a local optimum point

One might expect the same thing to be true in the discrete setting for (2), i.e. that the loss function also has the strict positivity of the Hessian that may hopefully be straightforwardly extended from (3). Unfortunately this is not true generally. Efforts have been made to connect the problems in continuous and discrete settings, notably in [1] and some references therein for generic problems, and in [5, 14, 22] for elliptic problems. All these works beautifully integrate the mathematical understanding in the continuous setting to address the setting of finite-dimensional problems. In so doing, assumptions are made. Particularly, in [1], it is assumed that the to-be-inferred function is finite-dimensional, and for elliptic operators, the structural off-diagonal low rank property is deployed.

Instead of conducting functional analysis in the spirit of the references above, in this paper, we directly tackle the finite dimensional object and examine the positivity of the discrete Hessian matrix. According to the definition of $\text{loss} = \sum_{j=1}^r \text{loss}_j$ in (2), with the parameter $\sigma \in \mathbb{R}^N$, the Hessian is

$$\mathcal{H}\text{loss}(\sigma) = \sum_{j=1}^r \mathcal{H}\text{loss}_j(\sigma) \in \mathbb{R}^{N \times N}. \quad (4)$$

Naturally the Hessian is of size $\mathbb{R}^{N \times N}$ and has a rank at most r . The positivity is not guaranteed. Especially in the case of $N > r$, when the number of parameters to be recovered is more than the number of data points given, the Hessian is guaranteed to be rank-deficient. The statement on the uniqueness of the local minimum of σ_* is no longer valid.

This contrast between positivity of the Hessian in the continuous (3) and discrete form (4) prompts the key question of the current work:

Given fixed data, under what circumstance is the positivity of the Hessian (4) preserved?

Our toolbox here is random matrix theory. Knowing the matrix is at most rank r , we are to use matrix concentration inequalities to study the possibility of the sketched submatrix of $\mathcal{H}\text{loss}$ being positive.

It is worthwhile to note that our framework takes the perspective of PDE-constrained optimization, and examines the Hessian's local behavior around the global minimum. Preserving positivity of the Hessian is only one of the first quantities to be examined. A more substantial research effort is required to complete the full picture of well-posedness.

The pillars of our work are PDE-constrained optimization and random matrix theory, and we now briefly discuss the relevant results that we most heavily leverage.

PDE-constrained optimization: The problem that we study is formulated and originated from PDE-constrained optimization [15], a formulation extensively used for inverse problems [3, 23] and optimal control [19]. There are many challenges. Most problems suffer from notorious non-convexity, with landscapes sometimes very rugged. Often, regularization is introduced to convexify the loss function [7–9, 17], and in other approaches, specifics of the problems are built in algorithm design, as is done in full-wave inversion using multi-frequency data [6, 11, 32] that mitigate cycle skipping [10].

Randomized sketching and subsampling techniques: Random matrix sketching provides the technical foundation for our current work. Randomized numerical linear algebra has attracted a lot of attention in the past decade, mainly in the context of data analysis and machine learning tasks. With connections across the fields of numerical linear algebra [13, 20, 21], applied probability [29], and theoretical computer science [12, 31], randomized linear algebra techniques and theory have largely been developed and have ripened in recent years. Broadly speaking, the subject is to look for a low-dimensional representation of high-dimensional data. This is supported by the bulk of theoretical foundations built upon the celebrated Johnson–Lindenstrauss lemma [18].

1.2. Key strategy of this work

Throughout the paper, we assume $N > r$ and thus it is impossible to reconstruct the full parameter $\sigma \in \mathbb{R}^N$. However, it is possible to recover a sector of it. We are to understand how big this sector can be.

Suppose we reduce our problem from recovering $\sigma \in \mathbb{R}^N$ to a subset of $\sigma|_{\Lambda} \in \mathbb{R}^m$ with the cardinality of $|\Lambda| = m$, with Λ standing for a mask of the smaller set of parameters under reconstruction. The relation between the number of available data points (r), and the number of parameters needing to be reconstructed (m) is intriguing. In this paper, we present a probabilistic approach to partially answer this question. We examine the possibility of ‘sketching’ the $N \times N$ matrix (4) and projecting it down to a smaller matrix, and we look for conditions under which the projected matrix attains positivity, making the problem locally strictly convex when confined in Λ .

In the current context, it is reformulated as subsampling the Hessian matrix:

$$\mathcal{H}\text{loss}(\sigma) \in \mathbb{R}^{N \times N} \rightarrow \mathcal{H}\text{loss}(\sigma|_{\Lambda}) = \mathbf{S}(\mathcal{H}\text{loss}(\sigma))\mathbf{S}^{\top} \in \mathbb{R}^{m \times m}. \quad (5)$$

Here the ‘subsampling’ matrix is $\mathbf{S} \in \mathbb{R}^{m \times N}$ (see explicit definition in (8)). We are to investigate the conditions on $\mathbf{S}$ so that the sketched submatrix (5) has a high probability of being strictly positive definite. To ensure the sketched Hessian is of full rank, and noting that the base Hessian (4) is rank r , it is necessary to assume $m \leq r$.

1.3. Organization of the paper

The remainder of the paper is organized as follows. In section 2 we set notations and present our main theoretical results. In particular, we quantify the probability of the subsampled Hessian matrix being well-conditioned. The success rate is analyzed against various parameters (notably m and r). In section 3 we garner major technical components of the proof for the theorem. Finally in section 4 we present numerical examples over synthetic data and PDE simulation data from an elliptic inverse problem. The numerical results showcase the agreement with the theoretical guarantees.

2. Main results

We present the formulation of the problem and unify notation in section 2.1. The formulation leads to a problem of analyzing the conditioning of a specifically defined matrix. The main results are presented in subsection 2.2. We discuss the proof ingredients and the general strategy in subsection 2.3.

2.1. Problem formulation

We formulate the problem and unify notation in this section.

Recalling the definition in (2), with straightforward derivation, we compute each entry of the Hessian of the loss function. At the ground-truth $\sigma = \sigma_*$, for $i_1, i_2 = 1, \dots, N$, we have

$$\begin{aligned} (\mathcal{H}\text{loss})_{i_1, i_2}(\sigma_*) &= \sum_{j=1}^r \langle \partial_{i_1} \mathcal{M}(\sigma_*), \psi_j \rangle \langle \partial_{i_2} \mathcal{M}(\sigma_*), \psi_j \rangle + (\mathcal{H}\mathcal{M})_{i_1, i_2}(\sigma_*) (\langle \mathcal{M}(\sigma_*), \psi_j \rangle - \mathbf{d}_j), \\ &= \sum_{j=1}^r \langle \partial_{i_1} \mathcal{M}(\sigma_*), \psi_j \rangle \langle \partial_{i_2} \mathcal{M}(\sigma_*), \psi_j \rangle, \end{aligned} \quad (6)$$

where the second term drops out because $\mathbf{d}_j = \langle \mathcal{M}(\sigma_*), \psi_j \rangle$, recovering the form of the Gauss–Newton Hessian. The first-order derivative has the standard definition $\partial_i \mathcal{M}(\sigma) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{M}[\sigma + \epsilon \phi_i] - \mathcal{M}[\sigma]}{\epsilon}$ and when $\{\phi_i\}$ are orthonormal, $\partial_i \mathcal{M}(\sigma) = \langle \frac{\delta \mathcal{M}}{\delta \sigma}, \phi_i \rangle_\Omega$ where $\frac{\delta \mathcal{M}}{\delta \sigma}$ is the Fréchet derivative.

This symmetric structure of the Hessian (6) strongly suggests the use of the short-hand notation of

$$\mathbf{H}^N = \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N,1} & \dots & \mathbf{H}_{N,N} \end{bmatrix} = \Phi^N \Phi^{N\top} \in \mathbb{R}^{N \times N}, \quad (7)$$

with $\Phi^N \in \mathbb{R}^{N \times r}$ and $\Phi_{i,j} = \langle \partial_i \mathcal{M}(\sigma_*), \psi_j \rangle$.

We are interested in the regime where $N \gg r$, so many rows in $\Phi^N \in \mathbb{R}^{N \times r}$ are linearly dependent, and $\mathbf{H}$ is rank deficient and non-invertible.

We look for a subsampled Hessian matrix that can be inverted, and has moderate condition number. Since the Hessian matrix is always positive semidefinite, we are to look for the subsampling that guarantees *strict* positive definiteness. Mathematically, randomly selecting $m \leq r$ rows from Φ^N amounts to multiplying the sampling matrix $\mathbf{S}^N \in \mathbb{R}^{m \times N}$ on Φ^N :

$$\mathbf{S}^N \Phi^N = \begin{bmatrix} -\Phi_{\omega_1, :}^N & - \\ \vdots & \\ -\Phi_{\omega_m, :}^N & - \end{bmatrix} \in \mathbb{R}^{m \times r}, \quad \text{with } \mathbf{S}_{i,:}^N = e_{\omega_i}, \quad (8)$$

where e_i is the unit vector placing 1 on i th entry, and the random variables $\{\omega_i\}_{i=1}^m$ are generated uniformly from the integer set $\{1, \dots, N\}$.

Accordingly, the sampled Hessian submatrix becomes:

$$\mathbf{S}^N \mathbf{H}^N \mathbf{S}^{N\top} = \mathbf{S}^N \Phi^N \Phi^{N\top} \mathbf{S}^{N\top} := \mathbf{H}_s^N \in \mathbb{R}^{m \times m}. \quad (9)$$

The problem now translates to finding conditions for this subsampled matrix $\mathbf{H}_s^N$ to be well-conditioned.

2.2. Conditioning of the Hessian

One key quantity we examine is the conditioning of the sub-sampled matrix. This follows the conventional notation. Label the eigenvalues of the sub-sampled matrix (5) in descending order:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m,$$

the condition number is simply:

$$\text{cond}(\mathbf{S}(\mathcal{H}\text{loss}(\boldsymbol{\sigma}))\mathbf{S}^\top) = \frac{\lambda_1}{\lambda_m}. \quad (10)$$

We are to estimate the upper bound of the condition number of $\mathbf{H}_s^N$ when m is small. To start, we define a few key quantities for properties of $\mathbf{H}^N$.

Definition 1. (i) Diagonal variation parameters: We denote by ℓ and L the diagonal variation parameters. In particular, supposing all the diagonals are strictly positive, we define

$$\ell := \frac{N}{\text{Tr}(\mathbf{H}^N)} \min_{i=1,\dots,N} \mathbf{H}_{i,i}^N, \quad L := \frac{N}{\text{Tr}(\mathbf{H}^N)} \max_{i=1,\dots,N} \mathbf{H}_{i,i}^N. \quad (11)$$

(i) The coherence parameter: We denote by $\mu > 0$ the coherence:

$$\mu := \frac{N}{\|\mathbf{H}^N\|_F} \max_{i_1 \neq i_2=1,\dots,N} |\mathbf{H}_{i_1,i_2}^N|. \quad (12)$$

We note that all these quantities are inherent attributes of $\mathbf{H}^N$ and are ‘relative’ in nature. To be specific, the quantities are preserved when $\mathbf{H}^N$ is scaled by a constant. For example, $L(\mathbf{H}^N) = L(c\mathbf{H}^N)$ for any $c > 0$. Clearly, $\frac{\text{Tr}(\mathbf{H}^N)}{N}$ is the averaged value along the diagonal for $\mathbf{H}^N$, and $\frac{\|\mathbf{H}^N\|_F}{N}$ is the root-mean-square of the entries in $\mathbf{H}^N$.

We next provide the main probabilistic statement of the Hessian submatrix $\mathbf{H}_s^N$ being well-conditioned. This is the rigorous version of the informal statement shown in section 1.2.

Theorem 1. Consider the discrete Hessian matrix $\mathbf{H}^N \in \mathbb{R}^{N \times N}$ of rank r from (7). Fix the sampling dimension $m \in \mathbb{N}^+$, where $m \leq r \leq N$, and moreover let m satisfy

$$m \leq \min \left\{ \frac{\ell}{146e^{\frac{1}{4}}} \left(\frac{\text{Tr}(\mathbf{H}^N)}{\|\mathbf{H}^N\|_2} - 1 \right), \quad \frac{\ell^2}{149e^{\frac{1}{2}}\mu^2 \log r} \frac{\text{Tr}(\mathbf{H}^N)^2}{\|\mathbf{H}^N\|_F^2} \right\}. \quad (13)$$

The subsampled Hessian $\mathbf{H}_s^N \in \mathbb{R}^{m \times m}$ (8) and (9), generated from uniform sampling, is well-conditioned with high probability:

$$\mathbb{P} \left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L + \tau(m)}{\ell - \tau(m)} \right) \geq 1 - \frac{1}{r}, \quad (14)$$

where the condition number is defined in (10), and the distortion quantity $\tau(m)$ is

$$\tau(m) = e^{\frac{1}{4}} \left(2m \frac{\|\mathbf{H}^N\|_2}{\text{Tr}(\mathbf{H}^N)} + 12\mu \sqrt{m \log r} \frac{\|\mathbf{H}^N\|_F}{\text{Tr}(\mathbf{H}^N)} \right). \quad (15)$$

We leave the proof to the later part of this section. Besides the general interpretations previously stated in (i)–(iii) section 1.2, we provide more thorough explanations below.

- The theorem uses a specific criterion to define the ‘success’: the conditioning of the sketched Hessian $\mathbf{H}_s^N$ is lower bounded by a certain threshold. This bound appearing in (14) is approximately $\frac{L}{\ell}$, the ratio of the maximal and the minimal diagonal entries in $\mathbf{H}^N$, with perturbations encoded in τ that we call the distortion term. Notably, $\tau(m)$ is asymptotically linear in the sample size m for large m . Intuitively, a smaller sample size m brings less distortion $\tau(m)$ and then the conditioning is closer to $\frac{L}{\ell}$. This matches the intuition that the system behaves better when the number of to-be-reconstructed parameter (m) is small.
- The success probability (RHS of (14)) is $1 - \frac{1}{r}$. This means bigger r gives higher probability for the subsampled Hessian to be well-conditioned. The reconstruction is better conditioned when one has a large number of measurements (r).

- To make sense of the theorem, we also need the term $\ell - \tau(m)$ to be strictly greater than 0. This assumption is satisfied by the requirement on m in (13). A more careful derivation also suggests

$$0 < \frac{L + \tau(m)}{\ell - \tau(m)} \leq \frac{73r(L + \ell)}{\ell}. \quad (16)$$

Please see section 3.2 for the complete proof of the claim above.

- Noticing that $\mathbf{H}^N$ has rank r , due to Cauchy–Schwarz, we obtain the basic estimates of

$$1 \leq \frac{\text{Tr}(\mathbf{H}^N)}{\|\mathbf{H}^N\|_2} = \frac{\sum_{i=1}^r \lambda_i}{\max_i \lambda_i} \leq r, \quad 1 \leq \frac{\text{Tr}(\mathbf{H}^N)^2}{\|\mathbf{H}^N\|_F^2} = \frac{(\sum_{i=1}^r \lambda_i)^2}{\sum_{i=1}^r \lambda_i^2} \leq r.$$

This suggests the upper bound for m in (13) is roughly $\mathcal{O}(r)$ up to a constant factor. When the matrix is incoherent, meaning with a large μ , such constant scaling becomes worse. Theorem 1 concludes that when the number m (of to-be-reconstructed parameters) scales linearly with the number of measurements r , then with high probability, the global minimizer is locally unique and the problem is well-conditioned in its neighborhood.

2.3. Proof ingredients

In this subsection, we illustrate the key technical results and prepare for the proof of theorem 1.

We use the following short-hand notation for the off-diagonal Gram matrix:

$$\mathbf{M}^N = \mathbf{H}^N - \text{diag}(\mathbf{H}^N) \in \mathbb{R}^{N \times N}.$$

Similarly, the sketched submatrix of $\mathbf{M}^N$ is denoted as

$$\mathbf{M}_s^N = \mathbf{H}_s^N - \text{diag}(\mathbf{H}_s^N) = \mathbf{S}^N \mathbf{H}^N \mathbf{S}^{N\top} - \text{diag}(\mathbf{S}^N \mathbf{H}^N \mathbf{S}^{N\top}) \in \mathbb{R}^{m \times m}. \quad (17)$$

Lemma 1. *Under the same assumption as in theorem 1, the probability of the well-conditioned Hessian has the following upper bound:*

$$\mathbb{P}\left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L + \tau(m)}{\ell - \tau(m)}\right) \geq \mathbb{P}\left(\|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} \tau(m)\right). \quad (18)$$

The proof of lemma 1 is in section 3.1. This lemma transfers our focus from analyzing the conditioning of $\mathbf{H}_s^N$ to controlling the tail bound of $\|\mathbf{M}_s^N\|_2$. As we shall see below, this spectral norm quantity $\|\mathbf{M}_s^N\|_2$ is indeed a sub-Gaussian random variable, hence a sub-Gaussian type concentration inequality can be applied. We now recall the specific definition.

Proposition 1 (Sub-Gaussian distribution (proposition 10 of [27])). *Let $\alpha, \beta, D > 0$ be finite numbers such that the non-negative random variable X satisfies*

$$(\mathbb{E}X^p)^{\frac{1}{p}} \leq \alpha\sqrt{p} + \beta, \quad \text{for all } p \geq D. \quad (19)$$

Then for all $u \geq \sqrt{D}$, one has the tail probability

$$\mathbb{P}\left\{X \geq e^{\frac{1}{4}}(\alpha u + \beta)\right\} \leq e^{-\frac{u^2}{4}}. \quad (20)$$

There are a number of equivalent definitions for sub-Gaussian and the characterization of the tail bound. We refer the readers to L 2.5.2 in [29].

In the following lemma, we provide the moment estimation of the random variable $\|\mathbf{M}_s^N\|_2$ and show that it exhibits a sub-Gaussian decay by its higher moment bound.

Lemma 2. *Under the same assumption as in theorem 1, for $p \geq 2$, the p th moment is bounded by*

$$\left(\mathbb{E} \|\mathbf{M}_s^N\|_2^p\right)^{\frac{1}{p}} \leq \frac{2m}{N} \|\mathbf{H}^N\|_2 + 12 \sqrt{\max\left(\log m, \frac{p}{2}\right)} \mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F. \quad (21)$$

The proof of lemma 2 is in section 3.2.

Equipped with these two lemmas, we proceed with the proof for theorem 1.

Proof of theorem 1. Based on the moment estimation (21) in lemma 2, the random variable $\|\mathbf{S}^N \mathbf{M}^N \mathbf{S}^{N\top}\|_2$ is of sub-Gaussian type. In particular, according to proposition 1, we set

$$\beta = \frac{2m}{N} \|\mathbf{H}^N\|_2, \quad \alpha = 6\mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F, \quad D = 4 \log m.$$

Therefore, for $u = 2\sqrt{\log r} > \sqrt{D} = 2\sqrt{\log m}$, ($r \geq m$) by the distortion term τ (15), we achieve the probability tail bound

$$\begin{aligned} & \mathbb{P} \left(\|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} \tau(m) \right) \\ &= \mathbb{P} \left(\|\mathbf{M}_s^N\|_2 \leq e^{\frac{1}{4}} \left(\frac{m}{N} \|\mathbf{H}^N\|_2 + 6\mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F 2\sqrt{\log r} \right) \right) \\ &\geq 1 - e^{-\frac{4 \log r}{4}} = 1 - \frac{1}{r}. \end{aligned} \quad (22)$$

Applying the events relation in lemma 1, we prove the general well-conditioning Hessian probability (14),

$$\mathbb{P} \left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L + \tau(m)}{\ell - \tau(m)} \right) \geq \mathbb{P} \left(\|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} \tau(m) \right) \geq 1 - e^{-\frac{4 \log r}{4}} = 1 - \frac{1}{r}.$$

This completes the proof. $\square$

We make a final remark on potential improvement for the main theorem 1.

Remark 1. In our proof, the way that the parameters L, ℓ (11) are employed is pessimistic. Indeed, their definitions are to look at the extreme large and the extreme small ratio compared to the average diagonal entries, so they represent the worst case scenario. If we are to examine the distribution of diagonal entries, most entries are centered around the average. Hence the worst case bound is not tight, i.e. the extreme condition imposed is more stringent than necessary for most other entries. It is clear that the entry magnitude is well-concentrated although there is a big disparity between ℓ and L .

To obtain a better bound, a quick relaxation to this requirement is to examine a high-probability cut-off: Fix a small failure probability $0 < \eta < \frac{1}{2}$, and define the modified version of $\ell_0, L_0 > 0$ as

$$\mathbb{P} \left\{ \min_{i=1, \dots, m} (\mathbf{H}_s^N)_{i,i} < \frac{\ell_0}{\frac{\text{Tr}(\mathbf{H}^N)}{N}} \right\} \leq \frac{\eta}{2}, \quad \mathbb{P} \left\{ \max_{i=1, \dots, m} (\mathbf{H}_s^N)_{i,i} > \frac{L_0}{\frac{\text{Tr}(\mathbf{H}^N)}{N}} \right\} \leq \frac{\eta}{2}. \quad (23)$$

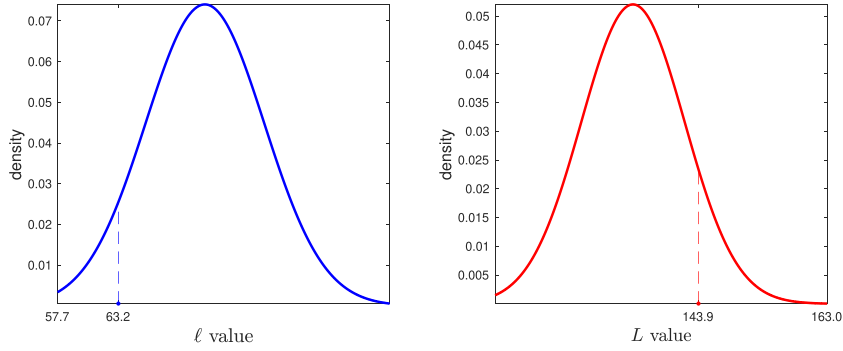


Figure 1. Distributions of min and max diagonal entries in $\mathbf{H}_s^N$ out of 10 000 simulations. We sample $m = 30$ from the Gauss–Newton Hessian $\mathbf{H}^N = \Phi^N \Phi^{N\top}$ where $\Phi^N \in \mathbb{R}^{N \times r}$ is generated by a random Gaussian matrix in the size of $N = 5000, r = 100$.

Clearly, we have $\ell_0 \geq \ell$, $L_0 \leq L$. With these re-defined quantities, the success probability (14) in the main theorem is improved to:

$$\mathbb{P}\left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L_0 + \tau(m)}{\ell_0 - \tau(m)}\right) \geq 1 - \frac{1}{r} - \frac{\eta}{2} - \frac{\eta}{2} = 1 - \frac{1}{r} - \eta$$

where the condition number threshold has a more moderate upper bound. In figure 1 we show ℓ_0 and L_0 with cut-off probability $\frac{\eta}{2} = 0.1$. In this case, the bound $\frac{L_0}{\ell_0} = \frac{143.9}{63.2} \approx 2.28$ is an improvement over $\frac{L}{\ell} = \frac{163}{57.7} \approx 2.83$, and the sacrifice on the probability is $\eta = 0.2$.

3. Details of the proofs

The main theorem is built upon two technical lemmas: lemma 1 on estimating the probability of well-conditioning, and lemma 2 on estimating high moments of an intermediate variable $\|\mathbf{M}_s^N\|_2$ (17). We present their proof in the following two subsections respectively.

3.1. Probability bound, lemma 1

In order to certify a good conditioning of the subsampled Hessian $\mathbf{H}_s^N$, lemma 1 informs us that the key quantity is the spectral norm of the hollow Gram matrix $\mathbf{M}_s^N$ in (17). We establish this connection and prove lemma 1 as follows.

Proof of lemma 1. Recall

$$\mathbf{M}_s^N = \mathbf{H}_s^N - \text{diag}(\mathbf{H}_s^N) \in \mathbb{R}^{m \times m}.$$

By Weyl's inequality (originally in [30], or corollary 6.3.8 of [16]), for $k = 1, \dots, m$,

$$|\lambda_k(\mathbf{H}_s^N) - \lambda_k(\text{diag}(\mathbf{H}_s^N))| \leq \|\mathbf{M}_s^N\|_2. \quad (24)$$

Take $k = 1$, and using the reverse triangle inequality, we know that

$$\begin{aligned}\lambda_1(\mathbf{H}_s^N) &\leq \lambda_1(\text{diag}(\mathbf{H}_s^N)) + \|\mathbf{M}_s^N\|_2 \\ &= \max_{i=1,\dots,m} (\mathbf{H}_s^N)_{i,i} + \|\mathbf{M}_s^N\|_2 \\ &\leq L \frac{\text{Tr}(\mathbf{H}^N)}{N} + \|\mathbf{M}_s^N\|_2.\end{aligned}$$

The last inequality is obtained due to (11). Analogously as above derivations, let $k = m$, we have

$$\lambda_m(\mathbf{H}_s^N) \geq \ell \frac{\text{Tr}(\mathbf{H}^N)}{N} - \|\mathbf{M}_s^N\|_2.$$

The event

$$\left\{ \|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} \tau(m) \right\} \quad (25)$$

implies the outcomes:

$$\begin{cases} \lambda_1(\mathbf{H}_s^N) \leq L \frac{\text{Tr}(\mathbf{H}^N)}{N} + \|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} (L + \tau(m)), \\ \lambda_m(\mathbf{H}_s^N) \geq \ell \frac{\text{Tr}(\mathbf{H}^N)}{N} - \|\mathbf{M}_s^N\|_2 \geq \frac{\text{Tr}(\mathbf{H}^N)}{N} (\ell - \tau(m)). \end{cases}$$

And consequently, assuming $\ell - \tau(m) > 0$, the event (25) leads to another event:

$$\left\{ \text{cond}(\mathbf{H}_s^N) = \frac{\lambda_1(\mathbf{H}_s^N)}{\lambda_m(\mathbf{H}_s^N)} \leq \frac{L + \tau(m)}{\ell - \tau(m)} \right\}.$$

Thus we have the probability relation of the two events:

$$\mathbb{P} \left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L + \tau(m)}{\ell - \tau(m)} \right) \geq \mathbb{P} \left(\|\mathbf{M}_s^N\|_2 \leq \frac{\text{Tr}(\mathbf{H}^N)}{N} \tau(m) \right).$$

To complete the proof, we still need to show that (13) leads to $\ell - \tau(m) > 0$. To do so, we recall from (15):

$$\tau(m) = \underbrace{2e^{\frac{1}{4}} m \frac{\|\mathbf{H}^N\|_2}{\text{Tr}(\mathbf{H}^N)}}_{\text{Term I}} + \underbrace{12\mu e^{\frac{1}{4}} \sqrt{m \log r} \frac{\|\mathbf{H}^N\|_F}{\text{Tr}(\mathbf{H}^N)}}_{\text{Term II}}. \quad (26)$$

According to the condition on the sampling dimension m (13),

$$\text{Term I} \leq \frac{\ell}{73} \left(1 - \frac{\|\mathbf{H}^N\|_2}{\text{Tr}(\mathbf{H}^N)} \right), \quad \text{Term II} \leq \frac{12\ell}{\sqrt{149}} \leq \frac{72\ell}{73}.$$

Consequently:

$$\tau(m) \leq \ell - \frac{\ell \|\mathbf{H}^N\|_2}{73 \text{Tr}(\mathbf{H}^N)}, \quad \text{and} \quad \ell - \tau(m) \geq \frac{\ell \|\mathbf{H}^N\|_2}{73 \text{Tr}(\mathbf{H}^N)} > 0.$$

This completes the proof. Moreover, a more careful derivation derives:

$$\frac{L + \tau(m)}{\ell - \tau(m)} \leq \frac{L + \ell}{\frac{\ell \|\mathbf{H}^N\|_2}{73 \text{Tr}(\mathbf{H}^N)}} = \frac{73(L + \ell)}{\ell} \frac{\|\mathbf{H}^N\|_2}{\text{Tr}(\mathbf{H}^N)} \leq \frac{73r(L + \ell)}{\ell},$$

proving (16). $\square$

3.2. Moment bound lemma 2

Lemma 2 is to quantify the probability of $\|\mathbf{M}_s^N\|_2$ being controlled within a certain range. To get this concentration tail, it is helpful to understand the moments of $\|\mathbf{M}_s^N\|_2$. Since higher moment bounds are crucial for characterizing this random variable's distribution, we next pin down its concentration property.

We remark that lemma 2 is a generalization of equation (6.1) of Tropp [27]. The key difference is that we remove the assumption in [27] that Φ^N has equal-norm rows. Nevertheless, we present the essential analysis for the moment bound (21). The proof framework uses the decoupling technique below.

Lemma 3 (Decoupling—theorem 9 of [27]). Fix an even integer $N \in \mathbb{N}$. There exists a sub-matrix $\widehat{\mathbf{M}}^N \in \mathbb{R}^{\frac{N}{2} \times \frac{N}{2}}$ of $\mathbf{M}^N \in \mathbb{R}^{N \times N}$ so that

$$\left(\mathbb{E} \left\| \mathbf{S}^N \mathbf{M}^N \mathbf{S}^{N\top} \right\|_2^p \right)^{\frac{1}{p}} \leq 2 \max_{\substack{m_1 + m_2 = m, \\ m_1, m_2 \in \mathbb{N}^+}} \left(\mathbb{E} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \right)^{\frac{1}{p}}. \quad (27)$$

Here, $\mathbf{S}_1^N \in \mathbb{R}^{m_1 \times \frac{N}{2}}$ and $\mathbf{S}_2^N \in \mathbb{R}^{m_2 \times \frac{N}{2}}$ are independent random sampling matrices with sample size respectively $m_1, m_2 \in \mathbb{N}^+$. The maximization in RHS of (27) is taken over all possible combinations of m_1, m_2 satisfying $m_1 + m_2 = m$.

To quantify the moment of interest $\left\| \mathbf{S}^N \mathbf{H}^N \mathbf{S}^{N\top} \right\|_2$, we instead focus on analyzing

$$\left(\mathbb{E} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \right)^{\frac{1}{p}} = \left(\mathbb{E}_1 \left[\mathbb{E}_2 \left[\left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \mid \mathbf{S}_1 \right] \right] \right)^{\frac{1}{p}}. \quad (28)$$

Regarding that, we present the useful information of both column and row samplings. The result relates to one specific type of matrix norms defined below.

Definition 2. For a matrix $\mathbf{A}$ with N columns, we define the induced 1, 2-norm, or equivalently the largest column Euclidean norm:

$$\|\mathbf{A}\|_{1,2} := \sup_{\substack{\|\mathbf{x}\|_2=1, \\ \mathbf{x} \in \mathbb{R}^N}} \|\mathbf{A}\mathbf{x}\|_1 = \max_{j=1, \dots, N} \|\mathbf{A}_{:,j}\|_2.$$

Next we present the key technical results for uniform sampling.

Lemma 4 (Theorem 8 of [27]). For a matrix $\mathbf{A}$ with N columns. Draw a random coordinate selector $\mathbf{S} \in \mathbb{R}^{m \times N}$

$$\left(\mathbb{E} \left\| \mathbf{A} \mathbf{S}^\top \right\|_2^p \right)^{\frac{1}{p}} \leq \sqrt{\frac{m}{N}} \|\mathbf{A}\|_2 + 3 \sqrt{\max \left(2 \log m, \frac{p}{2} \right)} \|\mathbf{A}\|_{1,2}. \quad (29)$$

Similarly, for row sampling, given a matrix $\mathbf{B}$ with N rows. Draw a random coordinate selector $\mathbf{S} \in \mathbb{R}^{m \times N}$

$$\left(\mathbb{E} \|\mathbf{S} \mathbf{B}\|_2^p \right)^{\frac{1}{p}} \leq \sqrt{\frac{m}{N}} \|\mathbf{B}\|_2 + 3 \sqrt{\max \left(2 \log m, \frac{p}{2} \right)} \|\mathbf{B}^\top\|_{1,2}. \quad (30)$$

Lemma 4 (from [27]) estimates the spectral norm of a submatrix whose columns are in a uniformly random collection of column set in the original matrix. This lemma generalizes the previous results in [24] and [25] to moments of all orders regimes. The row sampling can be proved by simply taking the transpose of $\mathbf{B}$ and applying the column sampling result previously stated.

Proof of lemma 2. The proof adopts the techniques from section 6 in [27].

- (i) Guided by the quantity of interest (28), we first condition on $\mathbf{S}_1$ and consider the random column sampling of $\mathbf{S}_2$. Applying the spectral norm estimation of column sampling (29),

$$\left(\mathbb{E}_2 \left[\left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \mid \mathbf{S}_1 \right] \right)^{\frac{1}{p}} \leq \sqrt{\frac{2m_2}{N}} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_2 + 3 \sqrt{\max \left(2 \log m_2, \frac{p}{2} \right)} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_{1,2}.$$

It remains to study:

$$\begin{aligned} \left(\mathbb{E} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \right)^{\frac{1}{p}} &= \left(\mathbb{E}_1 \left[\mathbb{E}_2 \left[\left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \right\|_2^p \mid \mathbf{S}_1 \right] \right] \right)^{\frac{1}{p}} \\ &\leq \left(\mathbb{E}_1 \left(\sqrt{\frac{m_2}{N}} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_2 + 3 \sqrt{\max \left(2 \log m_2, \frac{p}{2} \right)} \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_{1,2}^p \right)^p \right)^{\frac{1}{p}}. \quad (31) \\ &\leq \underbrace{\sqrt{\frac{m_2}{N}} \left(\mathbb{E}_1 \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_2^p \right)^{\frac{1}{p}}}_{\text{term (I)}} + 3 \underbrace{\sqrt{\max \left(2 \log m_2, \frac{p}{2} \right)} \left(\mathbb{E}_1 \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_{1,2}^p \right)^{\frac{1}{p}}}_{\text{term (II)}} \end{aligned}$$

The last line is by triangle inequality.

- (ii) Next, we respectively tackle term (I) and (II) estimations.

Term (I): Applying the spectral norm estimation of row sampling (30),

$$\left(\mathbb{E}_1 \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_2^p \right)^{\frac{1}{p}} \leq \sqrt{\frac{m_1}{N}} \left\| \widehat{\mathbf{M}}^N \right\|_2 + 3 \sqrt{\max \left(2 \log m_1, \frac{p}{2} \right)} \left\| \widehat{\mathbf{M}}^{N\top} \right\|_{1,2}. \quad (32)$$

Given the definition of μ (12), the column norm of $\widehat{\mathbf{M}}^N \in \mathbb{R}^{\frac{N}{2} \times \frac{N}{2}}$ has a uniform bound, that is

$$\left\| \widehat{\mathbf{M}}^{N\top} \right\|_{1,2} \leq \sqrt{\frac{N}{2}} \mu \frac{\left\| \mathbf{H}^N \right\|_F}{N} = \mu \frac{\left\| \mathbf{H}^N \right\|_F}{\sqrt{2N}}.$$

Then continue from (32), the estimation of term (I) in (31) becomes

$$\left(\mathbb{E}_1 \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_2^p \right)^{\frac{1}{p}} \leq \sqrt{\frac{m_1}{N}} \left\| \widehat{\mathbf{M}}^N \right\|_2 + 3 \sqrt{\max \left(2 \log m_1, \frac{p}{2} \right)} \mu \frac{\left\| \mathbf{H}^N \right\|_F}{\sqrt{2N}}. \quad (33)$$

Term (II): For any random realization of $\mathbf{S}_1^N$, we have a uniform bound

$$\left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_{1,2} \leq \sqrt{m_1} \mu \frac{\left\| \mathbf{H}^N \right\|_F}{N} = \mu \frac{\sqrt{m_1}}{N} \left\| \mathbf{H}^N \right\|_F.$$

Consequently, the p th moment is within this uniform bound as well, i.e.

$$\left(\mathbb{E}_1 \left\| \mathbf{S}_1^N \widehat{\mathbf{M}}^N \right\|_{1,2}^p \right)^{\frac{1}{p}} \leq \left(\left(\mu \frac{\sqrt{m_1}}{N} \left\| \mathbf{H}^N \right\|_F \right)^p \right)^{\frac{1}{p}} = \mu \frac{\sqrt{m_1}}{N} \left\| \mathbf{H}^N \right\|_F. \quad (34)$$

Now we plug in the analysis of term (I) (33) and (II) (34) back into the overall estimation (31),

$$\begin{aligned} \left(\mathbb{E} \|\mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \|_2^p \right)^{\frac{1}{p}} &\leq \frac{\sqrt{m_1 m_2}}{\frac{N}{2}} \|\widehat{\mathbf{M}}^N\|_2 + 3 \sqrt{\max \left(2 \log m_1, 2 \log m_2, \frac{p}{2} \right)} \mu \frac{\sqrt{m_1} + \sqrt{m_2}}{N} \|\mathbf{H}^N\|_F \\ &\leq \frac{m}{N} \|\mathbf{M}^N\|_2 + 6 \sqrt{\max \left(\log m, \frac{p}{2} \right)} \mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F. \end{aligned} \quad (35)$$

To derive the last inequality, note that $\|\widehat{\mathbf{M}}^N\|_2 \leq \|\mathbf{M}^N\|_2$, since $\widehat{\mathbf{M}}^N$ is the submatrix of $\mathbf{M}^N$. We further replace m_1, m_2 in the last inequality using the relation that $m_1 + m_2 = m, m_1, m_2 \in \mathbb{N}^+$ which implies

$$\max(\log m_1, \log m_2) < \log m, \quad 2\sqrt{m_1 m_2} \leq m, \quad \sqrt{m_1} + \sqrt{m_2} \leq \sqrt{2m}.$$

Moreover, due to the relation of $\mathbf{M}^N$ and $\mathbf{H}^N$:

$$\mathbf{H}^N - \mathbf{M}^N = \text{diag}(\mathbf{H}^N),$$

and the fact that the difference $\text{diag}(\mathbf{H}^N)$ is positive semi-definite, by Weyl's monotonicity principle [4, 30],

$$\|\mathbf{M}^N\|_2 = \sigma_1(\mathbf{M}^N) \leq \sigma_1(\mathbf{H}^N) = \|\mathbf{H}^N\|_2.$$

We replace the norms of $\mathbf{M}^N$ in (35) with norm of $\mathbf{H}^N$ and obtain the main result (21):

$$\left(\mathbb{E} \|\mathbf{S}_1^N \widehat{\mathbf{M}}^N \mathbf{S}_2^{N\top} \|_2^p \right)^{\frac{1}{p}} \leq \frac{m}{N} \|\mathbf{H}^N\|_2 + 6 \sqrt{\max \left(\log m, \frac{p}{2} \right)} \mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F.$$

To obtain the ultimate moment bound, we trace back to the decoupling (27). Multiplying the above estimation with 2, we conclude the calculation (21)

$$\left(\mathbb{E} \|\mathbf{S}^N \mathbf{M}^N \mathbf{S}^{N\top} \|_2^p \right)^{\frac{1}{p}} \leq \frac{2m}{N} \|\mathbf{H}^N\|_2 + 12 \sqrt{\max \left(\log m, \frac{p}{2} \right)} \mu \frac{\sqrt{m}}{N} \|\mathbf{H}^N\|_F.$$

□

4. Numerical experiments

In this section, we present numerical evidence that corroborate our theory. According to theorem 1, we have stated that, from (14):

$$\mathbb{P} \left(\text{cond}(\mathbf{H}_s^N) \leq \frac{L + \tau(m)}{\ell - \tau(m)} \right) \geq 1 - \frac{1}{r}.$$

The theory implies that with high probability, the condition number of the sketched Hessian is well within control. The success probability is better if r is big (large number of measurements), and the condition number upper bound is smaller if m is small (small number of to-be-reconstructed parameters). We validate these observations in our numerical simulation, and focus our attention on the interplay between the data rank r and the sampling size m .

The following two subsections present numerical results run on two test cases: synthetic tests, and PDE-based simulation for an elliptic inverse problem. To obtain probabilistic results, 10 000 independent runs are applied in each test.

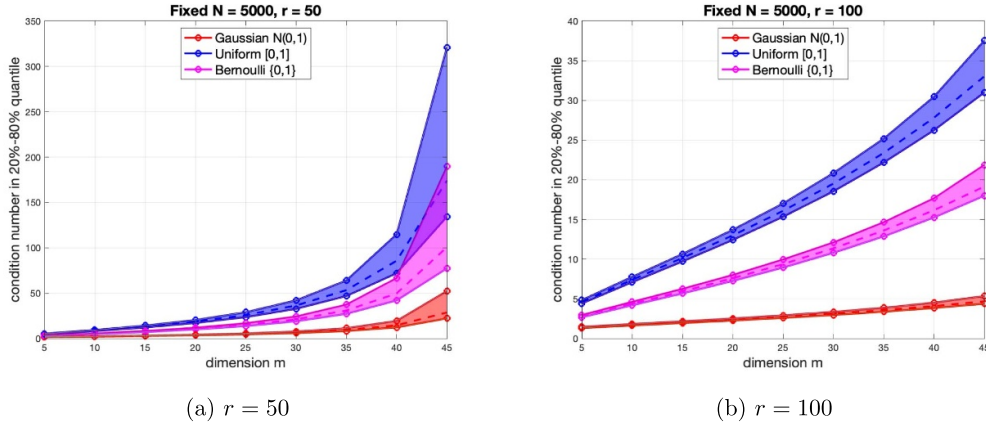


Figure 2. The concentration of condition number $\text{cond}(\mathbf{H}_s^N)$.

Table 1. Key parameter values in (14).

	Gaussian	Uniform	Bernoulli		Gaussian	Uniform	Bernoulli
$\frac{L}{\ell}$	4.1076	2.4819	3.0833	$\frac{L}{\ell}$	2.8263	2.0104	2.3000
$\frac{\ \mathbf{H}^N\ _2}{\text{Tr}(\mathbf{H}^N)}$	0.0241	0.7547	0.5095	$\frac{\ \mathbf{H}^N\ _2}{\text{Tr}(\mathbf{H}^N)}$	0.0130	0.7521	0.5043
$\frac{\ \mathbf{H}^N\ _F}{\text{Tr}(\mathbf{H}^N)}$	0.1422	0.7556	0.5143	$\frac{\ \mathbf{H}^N\ _F}{\text{Tr}(\mathbf{H}^N)}$	0.1010	0.7525	0.5068

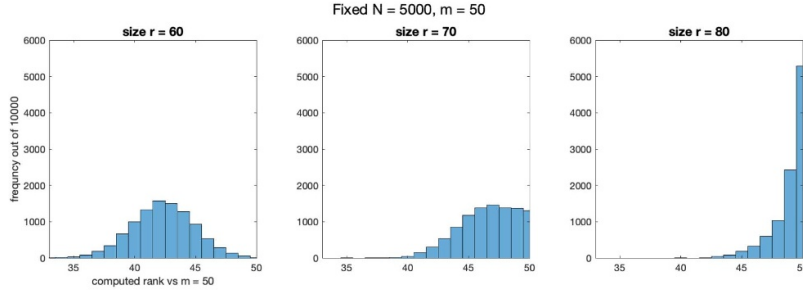
4.1. Synthetic tests

We start with the synthetic experiment. The full Hessian matrix $\mathbf{H}^N = \Phi^N \Phi^{N\top} \in \mathbb{R}^{N \times N}$ (7) is built via the rank- r matrix $\Phi^N \in \mathbb{R}^{N \times r}$. The entries of Φ^N are generated by synthetic data, from various choices of probability distributions shown below. We study the conditioning of the Hessian submatrix $\mathbf{H}_s^N = \mathbf{S}^N \Phi^N \Phi^{N\top} \mathbf{S}^{N\top} \in \mathbb{R}^{m \times m}$ (9) after the uniform sampling applied on $\mathbf{S}^N \Phi^N$ (8).

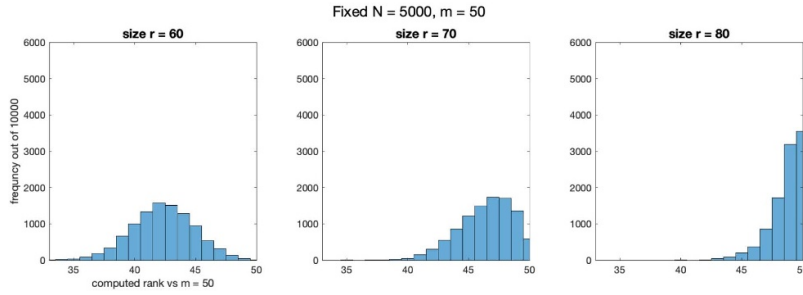
Fixed r with varying m . In the first set of experiments, we work with a fixed value of the full rank r (data points), and vary the sampling dimension m . This testing scenario allows us to fix the probability bound on the right hand side of (14) and check the cut-off value of condition number. The synthetic data (7) is built in three types. Each entry of the factor matrix $\Phi^N \in \mathbb{R}^{N \times r}$ is drawn from, respectively, standard Gaussian $\mathcal{N}(0, 1)$, uniform $[0, 1]$ and Bernoulli $\{0, 1\}$ distributions. 10 000 independent subsamplings have been run for various values of m , and in figure 2, we plot the percentage range $[20\%, 80\%]$ and the median value of the condition number of the sketched Hessian sub-matrices, for all three types of synthetic data matrix. The two subplots show results for $r = 50$ and $r = 100$ respectively.

From figure 2, we observe that for all three types of data matrices, the condition number of the sketched sub-matrices grows, statistically, as the sampling size m increases. The concentration variance also expands as m gets bigger. Moreover, sampling the standard Gaussian data produces the best conditioning in comparison to the other two data types.

The test suggests the uniform $[0, 1]$ gives the least satisfying condition number, and we now proceed to examine the reason. To do so, we calculate the main parameters in (14). See the list in table 1. While the ratio $\frac{L}{\ell}$ for the three cases are all similar, though the distortion term τ



(a) Rank threshold 10^{-6} . The computed rank is the number of eigenvalues that are above this threshold factored by the top eigenvalue λ_1 .



(b) Rank threshold 10^{-2} . The rank counting way is the same as in (a).

Figure 3. The rank distribution of Hessian submatrix $\mathbf{H}_s^N$.

Table 2. Key parameters in (14) (averaged over 10 000 simulations).

	$r = 60$	$r = 70$	$r = 80$	
$\frac{L}{\ell}$	6.0995	5.6329	5.2802	↓
$\frac{\ \mathbf{H}^N\ _2}{\text{Tr}(\mathbf{H}^N)}$	0.0597	0.0555	0.0523	↓
$\frac{\ \mathbf{H}^N\ _F}{\text{Tr}(\mathbf{H}^N)}$	0.1707	0.1636	0.1582	↓

varies greatly. In particular, according to (15), τ depends on $\frac{\|\mathbf{H}^N\|_2}{\text{Tr}(\mathbf{H}^N)}$, $\frac{\|\mathbf{H}^N\|_F}{\text{Tr}(\mathbf{H}^N)}$, and it is clear that the uniform $[0, 1]$ distribution produces much larger values for these factors than the Gaussian case.

Fixed m with varying r . We then carry out another set of experiments, with the sample size m fixed and we are free to adjust r (conditioned that $r > m$). By (14), the probability of the well-conditioned Hessian is expected to decrease as we increase the rank r . To demonstrate this, we display the frequency histogram for the rank of $\mathbf{H}_s^N \in \mathbb{R}^{m \times m}$ in figure 3. As the full discrete Hessian $\mathbf{H}^N \in \mathbb{R}^{N \times N}$ (4) rank r rises, we expect the sampled Hessian $\mathbf{H}_s^N$ (9) rank to concentrate towards the size of m , its largest possible rank. For this test, we only generate synthetic data using the Gaussian distribution. In figure 3, we preset a threshold to compute the rank of the subsampled matrix $\mathbf{H}_s^N$, and the histogram demonstrates the distribution of the rank generated from 10 000 runs using different r values.

Figure 3 shows that when m is fixed at 50, as the original rank r increases, the distribution of the rank of the subsampled Hessian matrix also concentrates to m , the largest possible rank.

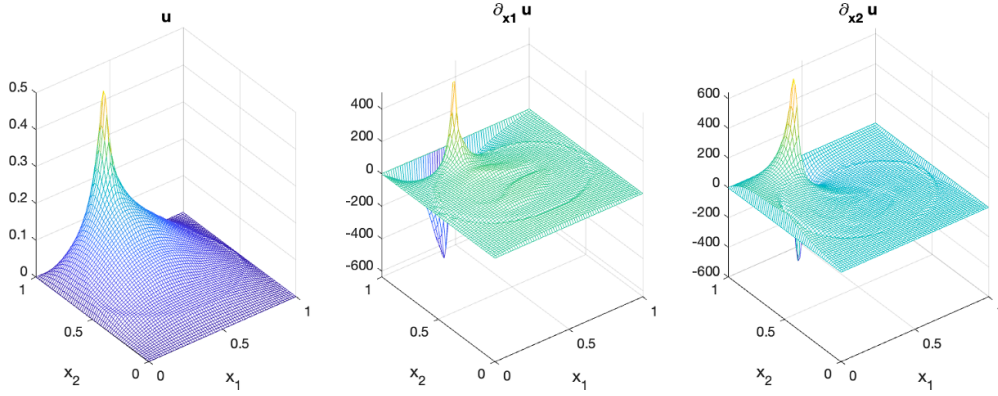


Figure 4. One solution to the forward model (36) at certain source location.

This validates the statement in theorem 1 that as r increases, one has higher probability of obtaining well-conditioned sub-matrices. The two rows in figure 3 showcases the cut-offs with different threshold. The top panel sets a more lenient rank condition and hence provides greater rank probability for the Hessian. Other key parameters in (14) are also provided in table 2. Their values drop according to the increasing r , which may lower the condition number threshold $\frac{L+\tau}{\ell-\tau}$ (14) and (15) as well.

4.2. Elliptic inverse problem

We now consider an inverse problem associated with elliptic type equations. This is similar to the classical electrical impedance tomography (EIT) problem from medical imaging but enjoys a better stability. This is because we use interior sensors to avoid the ill-conditioning issue commonly seen in the EIT problem [2]. Consider the elliptic equation

$$\begin{cases} \nabla_x \cdot (\sigma \nabla_x u) = S, & x \in \Omega \\ u = 0, & x \in \partial\Omega. \end{cases}$$

The task is to provide sources to the elliptic equation and take measurement of the solution inside the domain, to reconstruct the unknown media function $\sigma^* \in L^2(\Omega)$ ($\Omega \subset \mathbb{R}^2$), the elliptic coefficient, using finite measurements. We choose the ground-truth σ^* to be the Shepp–Logan phantom [26]. We define the loss function (2) to be the mismatch between the PDE-generated data and the true data. Suppose we fix the measurement index $j = (j_1, j_2)$. In particular, given the source S_{j_1} to the PDE, we compute the PDE solution, and test the solution by the testing function S_{j_2} . In this setting, the following u_{j_1} and h_{j_2} solve the forward and adjoint models respectively, and are used for the construction of the loss function derivative:

$$\begin{cases} \nabla_x \cdot (\sigma \nabla_x u_{j_1}) = S_{j_1}, & x \in \Omega \\ u_{j_1} = 0, & x \in \partial\Omega, \end{cases} \quad \begin{cases} \nabla_x \cdot (\sigma \nabla_x h_{j_2}) = S_{j_2}, & x \in \Omega \\ h_{j_2} = 0, & x \in \partial\Omega. \end{cases} \quad (36)$$

The forward map $\mathcal{M}_j : L^2(\Omega) \rightarrow L^2(\Omega)$ is given by $\mathcal{M}_j[\sigma] \equiv d_{j_1, j_2}$. See figure 4 below for a solution instance of equation (36) and its partial derivatives.

For the implementation, the window functions are chosen to be Gaussian functions as mollified versions of Dirac deltas:

$$S_j(x) = e^{-0.1(x-y_j)^2}, \text{ for } j = 1, \dots, N.$$

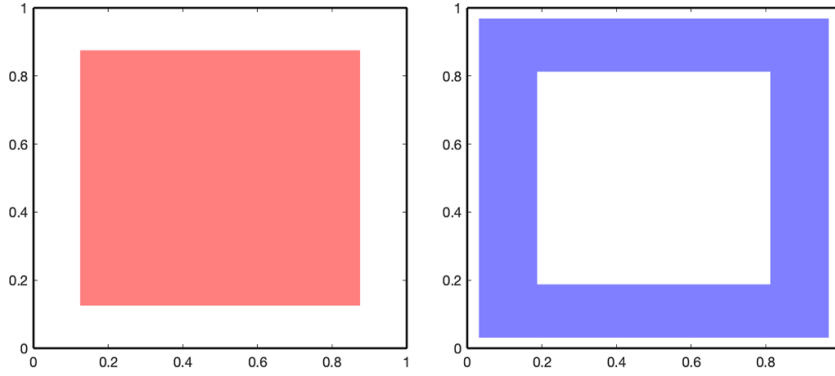


Figure 5. Measurement layouts. The two layouts present sub-domains $\mathcal{D}_1$ and $\mathcal{D}_2$.

Moreover, the orthonormal basis $\{\phi_i\}_{i=1}^r$ is used to estimate the media function σ . With all these preparations, we recall the notation defined in (7) for the following Hessian matrix:

$$\begin{aligned} \mathbf{H}^N &= \Phi^N \Phi^{N\top}, \quad \text{where } \Phi_{i,j} = \left\langle \phi_i, \frac{\delta \mathcal{M}_j}{\delta \sigma} \right\rangle, \\ &= \int_{\Omega} \partial_x \phi_i \nabla_x u_{j_1} \cdot \nabla_x h_{j_2} \, dx. \end{aligned} \quad (37)$$

Noting that the Fréchet derivative takes the form of $\nabla_x u \cdot \nabla_x h$, we then have

$$\Phi_{x,j} = \nabla_x u_{j_1}(x) \cdot \nabla_x h_{j_2}(x).$$

Numerically we set $\Omega = [0, 1]^2 \subset \mathbb{R}^2$, and the mesh size is set to be $\Delta x = \frac{1}{64}$ in both dimensions. So the total number of grid points is $N = (64 + 1)^2 = 4225$. As for the measurements, we consider placing sources and detectors in two types of subdomains $\mathcal{D}_1, \mathcal{D}_2 \subset \Omega$, see figure 5:

- (i) $\mathcal{D}_1 = [\frac{1}{8}, \frac{7}{8}]^2$. In this domain we have $r_1 = 384$ grid points for sample sources and detectors.
- (ii) $\mathcal{D}_2 = [\frac{1}{32}, \frac{31}{32}]^2 \setminus [\frac{3}{16}, \frac{13}{16}]^2$. In this domain, we have $r_2 = 334$ grid points for sample sources and detectors.

To select data, we randomly pick $\frac{1}{6}$ portion of the nodes in these subdomains $\mathcal{D}$ to serve as sources, and we place detectors within ± 5 spatial units from the source points. The generated discrete data matrices $\Phi^N, \mathbf{H}^N$ are presented in figure 6.

We perform the random sketching on these two established data matrices. In figures 7 and 8 we present the condition number of the subsampled matrices $\mathbf{H}_s^N \in \mathbb{R}^{m \times m}$ (9) as a function of m . In each scenario, 10 000 runs are performed and we document the 20%–80% quantile. As m increases, the condition number clearly increases as well. The dotted line in the plots are calculated using $\frac{L}{\ell}$ (14) as a threshold. The two tables below (tables 3 and 4) the plots count the probability of the condition values exceeding this threshold from all simulations. The failure probability gradually increases as m increases.

These findings validate the main theorem 1.

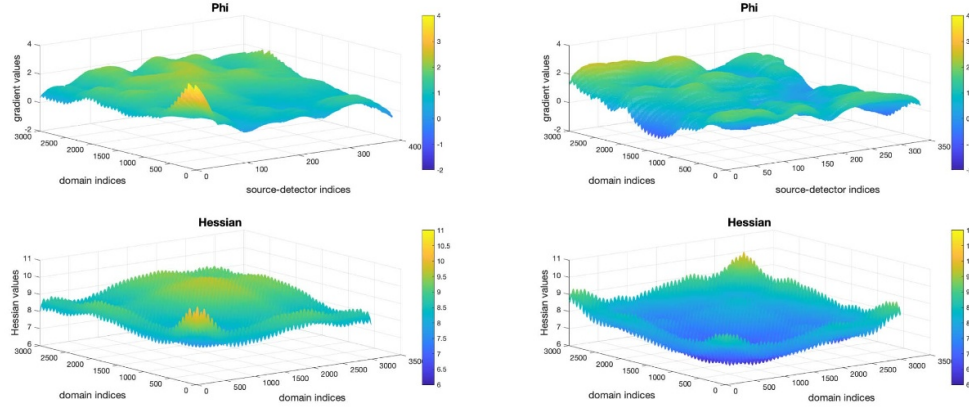
(a) Data $\Phi_1^N, \mathbf{H}_1^N$ with measurements from $\mathcal{D}_1$ (b) Data $\Phi_2^N, \mathbf{H}_2^N$ with measurements from $\mathcal{D}_2$

Figure 6. The entry magnitude of matrices Φ^N (top) and $\mathbf{H}^N$ (bottom). The left and right panels are results for $\mathcal{D}_1$ and $\mathcal{D}_2$ respectively. The magnitudes are showed in logarithm values, where the data matrices are first processed by Gaussian filters with std 20.

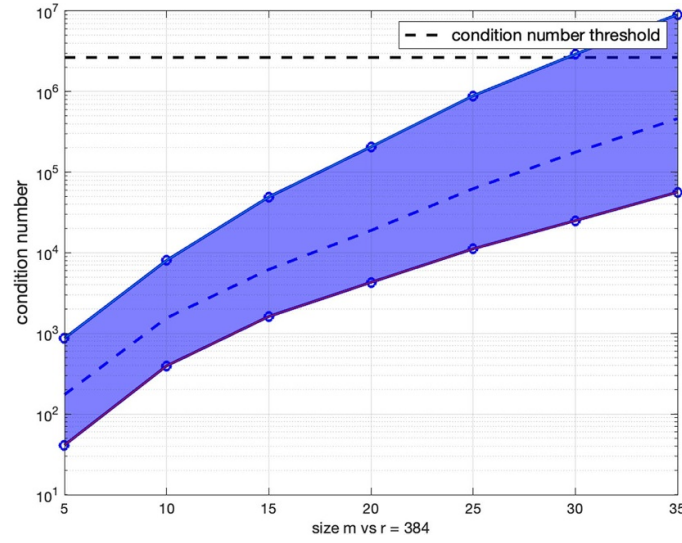


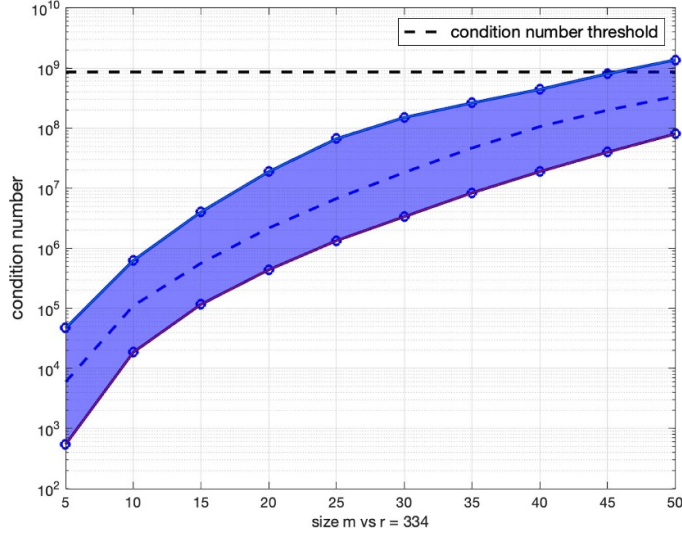
Figure 7. Conditioning concentration of subsampled $\mathbf{H}_1^N$.

Table 3. LHS probability in (14) of $\mathbf{H}_1^N$.

size m (vs $r = 384$)	5	10	15	20	25	30	35
failure probability (%)	0.13	0.95	3.28	7.05	13.39	20.69	29.76

Table 4. LHS probability in (14) of $\mathbf{H}_2^N$.

size m (vs $r = 334$)	10	15	20	25	30	35	40	45	50
failure probability (%)	0.13	0.2	0.47	1.5	3.27	6.00	11.13	18.83	27.93

**Figure 8.** Conditioning concentration of subsampled $\mathbf{H}_2^N$

5. Conclusions

In a numerical inverse problem, we are usually only presented with a finite amount of data. It is unrealistic to expect that the limited information from the data can be used to fully reconstruct high dimensional parameters in large-scale physical systems, and thus one has to narrow down the goal to inferring a few subsampled parameters. In contrast to the unique reconstruction theory used to infer parameters in the infinite-dimensional setting, this paper studies the unique reconstruction in the finite dimensional setting, and spells out the number of reconstructable parameters (m) when the data given is fixed (r). In particular, we deploy random sketching techniques to examine the conditioning of the Hessian matrix that is $m \times m$ with rank r . When the ‘success’ is quantified by the Hessian’s condition number, we establish the success probability. The theory is validated for cases of both synthetic data and PDE simulation data from an elliptic inverse problem. The numerical results confirm the theoretical findings: As long as the number of samples m is within a constant factor of the full Hessian rank r , the sampled Hessian is well-conditioned with high probability.

There are many related research directions that we would like to explore in the future. One such direction is to integrate the greedy approach, and to use the selected rows to eliminate the rows of similar information getting repeated. Furthermore, in our entire derivation, we have assumed the size of the Hessian is finite. We have noticed that the main theorem 1 has no explicit dependence on the ambient dimension N but only on the rank and sampling size, r and m . This is a strong indication that the result can be pushed to justify the scenario where $N = \infty$, in which case the Hessian is a bilinear functional, mapping $L^2 \times L^2$ to a real number. This would allow us to fully come back to the continuous PDE setting. Our analysis strongly

depends on existing theory developed in [28]. We believe with refined probabilistic arguments, the error bound can be tightened further as well.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

The authors thank the editor and anonymous reviewers for helpful comments that improved the presentation of the paper. R J and Q L acknowledge the support from NSF-DMS-2308440. The research of S S is partially supported by NSF-DMS-2324368 and ONR Grant N00014-21-1-2119. R J is partially supported by NSF-DMS-2023239.

ORCID iDs

Ruhui Jin  <https://orcid.org/0009-0004-2611-0889>

Qin Li  <https://orcid.org/0000-0001-9210-8948>

Samuel N Stechmann  <https://orcid.org/0000-0001-7188-7946>

References

- [1] Alberti G S and Santacesaria M 2022 Infinite-dimensional inverse problems with finite measurements *Arch. Ration. Mech. Anal.* **243** 1–31
- [2] Alessandrini G and Vessella S 2005 Lipschitz stability for the inverse conductivity problem *Adv. Appl. Math.* **35** 207–41
- [3] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41
- [4] Bhatia R 2001 Linear algebra to quantum cohomology: the story of Alfred Horn's inequalities *Am. Math. Mon.* **108** 289–318
- [5] Bui-Thanh T, Li Q and Zepeda-Núñez L 2022 Bridging and improving theoretical and computational electrical impedance tomography via data completion *SIAM J. Sci. Comput.* **44** B668–93
- [6] Castellanos C, Métivier L, Operto S, Brossier R and Virieux J 2015 Fast full waveform inversion with source encoding and second-order optimization methods *Geophys. J. Int.* **200** 720–44
- [7] Engl H W, Kunisch K and Neubauer A 1989 Convergence rates for Tikhonov regularisation of non-linear ill-posed problems *Inverse Problems* **5** 523
- [8] Engl H W and Zou J 2000 A new approach to convergence rate analysis of Tikhonov regularization for parameter identification in heat conduction *Inverse Problems* **16** 1907
- [9] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Springer) p 375
- [10] Engquist B and Yang Y 2022 Optimal transport based seismic inversion: beyond cycle skipping *Commun. Pure Appl. Math.* **75** 2201–44
- [11] Epanomeritakis I, Akçınelik V, Ghattas O and Bielak J 2008 A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion *Inverse Problems* **24** 034015
- [12] Frieze A, Kannan R and Vempala S 1998 Fast Monte-Carlo algorithms for finding low-rank approximations *Proc. 39th Annual Symp. on Foundations of Computer Science (Cat. No.98cb36280)* pp 370–8
- [13] Halko N, Martinsson P G and Tropp J A 2011 Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions *SIAM Rev.* **53** 217–88
- [14] Harrach B 2019 Uniqueness and Lipschitz stability in electrical impedance tomography with finitely many electrodes *Inverse Problems* **35** 024005
- [15] Hinze M, Pinnau R, Ulbrich M and Ulbrich S 2008 *Optimization With PDE Constraints* (Springer) p 23

- [16] Horn R A and Johnson C R 1991 *Matrix Analysis* (Cambridge University Press)
- [17] Ito K and Jin B 2014 *Inverse Problems: Tikhonov Theory and Algorithms* (World Scientific) p 22
- [18] Johnson W and Lindenstrauss J 1984 Extensions of Lipschitz maps into a Hilbert space *Contemp. Math.* **26** 189–206
- [19] Leugering G, Engell S, Griewank A, Hinze M, Rannacher R, Schulz V, Ulbrich M and Ulbrich S 2012 *Constrained Optimization and Optimal Control for Partial Differential Equations* (Springer) p 160
- [20] Mahoney M W and others 2011 Randomized algorithms for matrices and data *Found. Trends[®] Mach. Learn.* **3** 123–224
- [21] Martinsson P-G and Tropp J A 2020 Randomized numerical linear algebra: foundations and algorithms *Acta Numer.* **29** 403–572
- [22] Nelsen N H and Stuart A M 2024 Operator learning using random features: a tool for scientific computing *SIAM Rev.* **66** 535–71
- [23] Oldenburg D W, Haber E and Shekhtman R 2013 Three dimensional inversion of multisource time domain electromagnetic data *Geophysics* **78** E47–E57
- [24] Rudelson M 1999 Random vectors in the isotropic position *J. Funct. Anal.* **164** 60–72
- [25] Rudelson M and Vershynin R 2007 Sampling from large matrices: an approach through geometric functional analysis *J. ACM* **54** 21–es
- [26] Shepp L A and Logan B F 1974 The Fourier reconstruction of a head section *IEEE Trans. Nucl. Sci.* **21** 21–43
- [27] Tropp J A 2008 On the conditioning of random subdictionaries *Appl. Comput. Harmon. Anal.* **25** 1–24
- [28] Tropp J A 2012 User-friendly tail bounds for sums of random matrices *Found. Comput. Math.* **12** 389–434
- [29] Vershynin R 2018 *High-Dimensional Probability: An Introduction With Applications in Data Science* (Cambridge university press) p 47
- [30] Weyl H 1912 Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung *Math. Ann.* **71** 441–79
- [31] Woodruff D P and others 2014 Sketching as a tool for numerical linear algebra *Found. Trends[®] Theor. Comput. Sci.* **10** 1–157
- [32] Zhang B, Zepeda-Nunez L and Li Q 2024 Solving the wide-band inverse scattering problem via equivariant neural networks *J. Comput. Appl. Math.* **451** 116050