Data selection: at the interface of PDE-based inverse problem and randomized linear algebra

Kathrin Hellmuth[*], Ruhui Jin[†], Qin Li[‡], Stephen J. Wright[§]

Abstract

All inverse problems rely on data to recover unknown parameters, yet not all data are equally informative. This raises the central question of data selection. A distinctive challenge in PDE-based inverse problems is their inherently infinite-dimensional nature: both the parameter space and the design space are infinite, which greatly complicates the selection process. Somewhat unexpectedly, randomized numerical linear algebra (RNLA), originally developed in very different contexts, has provided powerful tools for addressing this challenge. These methods are inherently probabilistic, with guarantees typically stating that information is preserved with probability at least $1 - p$ when using N randomly selected, weighted samples. Here, the notion of "information" can take different mathematical forms depending on the setting. In this review, we survey the problem of data selection in PDE-based inverse problems, emphasize its unique infinite-dimensional aspects, and highlight how RNLA strategies have been adapted and applied in this context.

1 Introduction

Inverse problems based on partial differential equations (PDE) are relevant in such important and physically motivated applications as materials science, medical imaging, and geophysics. The fundamental task of the inverse problem is to reconstruct certain parameters (typically functions) in the PDEs that govern the dynamics of the system, using data gathered from the system. These parameters often encode critical information for downstream tasks. For example, recovering the optical properties of biological tissue may reveal abnormalities, recovering the refractive index of subsurface structures is essential in petroleum exploration, and recovering the band structure of a quantum system can help with material design.

PDE-based inverse problems have been studied extensively, from both analytical and algorithmic perspectives, across a wide range of PDE models (see, e.g., [85, 63, 97]). Most of

Funding: QL, RJ and SW acknowledge support from the NSF through the grant NSF-DMS-2023239. QL's work was also supported by NSF-DMS-2308440. KH acknowledges support from the Jet Propulsion Laboratory PDRDF 24AW0133.

[*]Department of Computing + Mathematical Sciences, Californian Institute of Technology, Pasadena, CA

[†]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD

[‡]Department of Mathematics and Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI

[§]Department of Computer Science, University of Wisconsin-Madison, Madison, WI

this work assumes that data is given, and focuses on the quality and stability of reconstruction. In practice, reconstruction is often preceded by a critical step: *experimental design*. Experimental design refers to the process of planning experiments so as to maximize the information content of the data collected for a given task. In PDE-based inverse problems, we usually cannot alter the governing equations themselves, but we do have control over experimental settings such as the source profile (which triggers the PDE dynamics) or the placement of detectors. This perspective is closely related to data selection: given the possibility of generating a large collection of data, the goal is to identify the most informative subset of experiments that yields the best possible reconstruction of the unknown parameters.

Classical experimental design is a rich and well-developed area of statistics, offering tools such as leverage scores, Fisher information, and optimality criteria including A-, D-, and E-optimality, which have been successfully applied across many scientific domains (see [65, 127]). In recent years, many of these techniques have been brought into the setting of PDE-based inverse problems, particularly within the framework of optimal design (see the reviews [13, 81]). Yet, the direct transfer of experimental design principles to PDE-based problems is not straightforward; it introduces several distinctive challenges.

- *Structure induced by PDEs.* Classical design methods are largely generic and independent of specific models, but PDE-constrained inverse problems possess rich structural features that impose constraints while also offering opportunities. Effective experimental design must respect and, if possible, exploit these structures.

- *Efficiency over optimality.* Many design strategies seek an "optimal" subset of experiments. In PDE-based settings, efficiency often takes precedence: Given the high computational and experimental costs, the aim is to obtain informative data quickly rather than achieve exact optimality.

- *Cost of data acquisition.* Traditional experimental design often assumes that all candidate data are already available for evaluation. By contrast, in PDE-based inverse problems, each data point may be expensive to generate, whether computationally or experimentally. This raises issues of sample complexity and calls for design strategies that account for acquisition costs.

These considerations indicate that classical methods for experimental design needed to be adapted in order to make them effective for PDE-based inverse problems.

The apparently unrelated research direction of *randomized numerical linear algebra* (RNLA) has developed rapidly in recent years [115, 113]. Originally motivated by large-scale data problems in machine learning and internet applications such as recommendation systems, RNLA methods emphasize efficiency, trading some accuracy for scalability through randomization. The RNLA philosophy resonates with the goals of experimental design for PDE inverse problems, where efficiency is a more pressing concern than exact optimality.

One of the classical contributions of RNLA to experimental design is the fast computation of leverage scores in linear regression [61, 113], which serves as a cornerstone for randomized data selection. Over the years, these techniques have been adapted and extended, and are now used for experimental design in PDE-based inverse problems.

This review summarizes these recent advances at the interface of data selection for PDE-based inverse problems and RNLA. The field is still developing: Many algorithms are being

actively refined, and nonlinear extensions remain largely open. Our aim is to provide a coherent overview of what has been achieved so far, while highlighting key challenges and promising directions for future research.

The remainder of the paper is structured as follows. Section 2 gives background on PDE-based inverse problems, highlighting structural features that are particularly relevant for experimental design. While reconstruction solvers are typically viewed as downstream tasks, the way they are formulated determines how much information can be extracted from data and thus directly guides data selection. Section 3 reviews two classical approaches: PDE-constrained optimization and Bayesian formulations. Both solvers rely critically on linearization; we also provide technical preparation in this context. Section 4, the main pillar of this review, surveys recent advances in RNLA methods and their applications to experimental design for PDE inverse problems. Section 5 concludes with a discussion of nonlinear extensions and directions for future research.

2 PDE based inverse problems

Among inverse problems, those governed by PDEs form a distinct and important subclass, marked by unique structural and analytical challenges. While different PDEs give rise to diverse inversion properties, they also share fundamental traits that permit a unified treatment. A particularly notable feature of PDE-based inverse problems is their inherently infinite-dimensional nature, which contrasts with the finite-dimensional representations required for computation. It is not trivial to bridge this disparity. Theoretical analysis is typically carried out in infinite-dimensional function spaces, whereas numerical algorithms must ultimately operate in finite dimensions. This bridge lies at the core of the present review and directly motivates our focus on data selection and experimental design.

In the subsections that follow, we introduce a generic PDE-based inverse problem in its infinite-dimensional form and describe the hierarchy of steps leading to the finite-dimensional numerical treatment. We discuss issues of well-posedness and the implications for experimental design, and we distinguish between quantitative and qualitative design goals. Finally, we conclude with representative examples of PDE-based inverse problems, grounding the abstract framework in concrete scientific applications.

2.1 Infinite-dimensional nature

Every PDE-based inverse problem begins with an underlying partial differential equation of the form:

$$\mathcal{L}_{\boldsymbol{\sigma}}[u] = S_{\theta_1}(x) \quad \Rightarrow \quad \mathrm{d} = M_{\theta_2}[u]\,. \tag{1}$$

Here, $\mathcal{L}_{\boldsymbol{\sigma}}$ denotes the PDE operator that models the physical or biological phenomenon under investigation, defined over a Euclidean domain with spatial variable $x \in \mathbb{R}^d$. The operator $\mathcal{L}_{\boldsymbol{\sigma}}$ depends on a parameter $\boldsymbol{\sigma}$, which encodes certain characteristics of the system. The input term $S_{\theta_1}(x)$ may represent a source term, boundary condition, or initial condition. The system is fully specified by fixing $\boldsymbol{\sigma}$, and a given input $S_{\theta_1}(x)$ leads to a solution u of the PDE.

The state u is typically not observable in full. Instead, measurements are obtained through an observation operator M_{θ_2}, producing ground-truth data $\mathrm{d}(\theta_1, \theta_2) = M_{\theta_2}[u]$. The pair

$\theta = (\theta_1, \theta_2) \in \Theta$ specifies the experimental design configuration, with θ_1 and θ_2 respectively describing the design of source and detector. Varying these configurations provides information about the unknown parameter $\boldsymbol{\sigma}$, and the collection of all such configurations Θ is referred to as the design space.

This entire process defines what we refer to as the **Input-to-Output (ItO)** operator:

$$\Lambda_{\boldsymbol{\sigma}} : \ (S_{\theta_1}, M_{\theta_2}) \mapsto \mathrm{d}(\theta_1, \theta_2) \, . \tag{2}$$

Throughout this paper, we assume for all admissible $\boldsymbol{\sigma} \in \Sigma$, the PDE (1) is well-posed under a prescribed metric, and the solution is measurable using the observation operator M_{θ_2} for all values of θ_2 under consideration. These assumptions guarantee that the ItO operator (2) is well-defined for each $\boldsymbol{\sigma}$. Moreover, we equip the domain and codomain of $\Lambda_{\boldsymbol{\sigma}}$ with metrics inherited from the input space $S_{\theta_1}(x)$ and the data space $\mathrm{d}(\theta_1, \cdot)$, respectively. Within this framework, the forward problem can be summarized as follows: for a fixed system parameter $\boldsymbol{\sigma}$, the operator $\Lambda_{\boldsymbol{\sigma}}$ maps each input $S_{\theta_1}(x)$ to a corresponding data function $\mathrm{d}(\theta_1, \cdot)$.

The inverse problem is to recover the parameter $\boldsymbol{\sigma}$ from knowledge of the operator $\Lambda_{\boldsymbol{\sigma}}$, or equivalently, from the full collection of data values $\mathrm{d}(\theta_1, \theta_2)$ over $\theta = (\theta_1, \theta_2) \in \Theta$. To formalize this, we define the data map:

$$\boldsymbol{\mathcal{M}}(\theta; \boldsymbol{\sigma}) : \Theta \otimes \Sigma \to \mathbb{R}, \quad \text{with} \quad \boldsymbol{\mathcal{M}}(\theta, \boldsymbol{\sigma}) = \Lambda_{\boldsymbol{\sigma}}[S_{\theta_1}, M_{\theta_2}] = \mathrm{d}(\theta_1, \theta_2) \, . \tag{3}$$

The task of the inverse problem is then to infer $\boldsymbol{\sigma}$ from complete knowledge of $\boldsymbol{\mathcal{M}}(\cdot \, ; \boldsymbol{\sigma})$.

A defining characteristic of this class of problems is their infinite-dimensional nature. The unknown parameter $\boldsymbol{\sigma}$ is typically a function and thus infinite-dimensional, while the data function d is indexed continuously over the design space Θ. Consequently, both the parameter space Σ and the design space Θ reside in infinite-dimensional settings, distinguishing these problems from standard finite-dimensional inverse problems.

2.2 Finite-dimensional computation

While the natural formulation of PDE-based inverse problems is often infinite-dimensional, we need to reduce to finite dimensions for experimental and numerical reasons. This reduction arises from two primary sources:

- **Limited data:** Only finitely many experiments can be carried out in practice, which restricts the design space to a finite subset $\Theta_c \subset \Theta$. If c experiments are conducted, and each experiment requires a specific value for $\theta = (\theta_1, \theta_2)$, then we are taking $|\Theta_c| = c$.
- **Reduced parameterization:** $\boldsymbol{\sigma}$ is typically represented with a finite number of degrees of freedom, yielding a finite-dimensional approximation Σ_n of dimension n. Any $\boldsymbol{\sigma}$ living in this space Σ_n is then automatically a finite-dimensional object.

Under these approximations, the data map (3) becomes

$$\boldsymbol{\mathcal{M}}(\theta; \boldsymbol{\sigma}) : \ \Theta_c \otimes \Sigma_n \to \mathbb{R} \, . \tag{4}$$

It is tempting to view the finite-dimensional formulation as simply an approximation of the infinite-dimensional one. Indeed, as $n, c \to \infty$, the two settings coincide. However, theoretical

results established in the infinite-dimensional framework do not in general translate directly to the finite n, c setting. Bridging these regimes requires a hierarchy of intermediate analyses, which we summarize as follows.

- **Well-posedness at the PDE level:** In the infinite-dimensional setting, well-posedness concerns uniqueness and stability of the inverse map. We seek an estimate of the form

$$\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\| \leq w(\|\boldsymbol{\mathcal{M}}(\cdot, \boldsymbol{\sigma}_1) - \boldsymbol{\mathcal{M}}(\cdot, \boldsymbol{\sigma}_2)\|) \,, \quad \boldsymbol{\sigma} \in \Sigma \cap B_R(\boldsymbol{\sigma}_*) \,, \tag{5}$$

  where w quantifies the stability of the inverse map and $B_R(\boldsymbol{\sigma}_*)$ denotes the neighborhood of radius R around the ground truth $\boldsymbol{\sigma}_*$. Here R may be infinite, or may encode problem-specific constraints such as positivity. The following two properties of w capture the essence of well-posedness.

  - **Uniqueness:**

$$w(0) = 0 \,, \tag{6}$$

    implying that complete knowledge of $\boldsymbol{\mathcal{M}}(\cdot, \boldsymbol{\sigma})$ uniquely determines $\boldsymbol{\sigma}$. Uniqueness has been established for many PDE-based inverse problems, the most celebrated being the Calderón problem [39], with pioneering results in [149, 100] and subsequent extensions in [124, 56, 84, 121].
  - **Stability:** We seek constants $C > 0$ and $\beta > 0$ such that

$$w(t) \leq C t^{\beta} \,, \tag{7}$$

    ensuring that small perturbations in data lead to controlled (e.g., Hölder/Lipschitz continuous) perturbations in the reconstruction. It is also possible for the upper bound to take on a logarithmic or exponential form. Foundational results in this direction include inverse conductivity problems [9] and inverse scattering problems [75, 145, 26].

  These results are highly problem-specific, relying on PDE-dependent techniques. For instance, complex geometrical optics (CGO) solutions are central in Schrödinger-type and elliptic problems [66], Carleman estimates play a key role in transport and hyperbolic equations [99, 98, 25], and singular decomposition or averaging-lemma techniques are employed for kinetic equations [52, 108].

- **Well-posedness in the finite-dimensional setting:** After discretization, we ask analogous questions of uniqueness and stability with Σ replaced by Σ_n and Θ by Θ_c. The relevant estimate becomes

$$\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\| \leq w_{n,c}(\|\boldsymbol{\mathcal{M}}(\cdot|_{\Theta_c}, \boldsymbol{\sigma}_1) - \boldsymbol{\mathcal{M}}(\cdot|_{\Theta_c}, \boldsymbol{\sigma}_2)\|) \,, \quad \sigma \in \Sigma_n \cap B_R(\boldsymbol{\sigma}_*) \,. \tag{8}$$

  That is:

  - If $w_{n,c}(0) = 0$, then $\boldsymbol{\mathcal{M}}(\cdot|_{\Theta_c}, \boldsymbol{\sigma})$ uniquely determines $\boldsymbol{\sigma}$ within Σ_n.
  - If an inequality of the form (7) holds for $w_{n,c}$, then small data perturbations lead to bounded reconstruction error.

It is worth noting that well-posedness in the infinite-dimensional setting does not directly translate to the finite-dimensional case. The properties of w in (5) are only suggestive; they do not guarantee corresponding properties for $w_{n,c}$ in (8).

- **Numerical implementation:** Ultimately, we need to implement a computational scheme for the reconstruction. This task requires algorithmic pipelines that recover $\boldsymbol{\sigma}$ from observed data. Two broad approaches dominate: (i) PDE-constrained optimization methods [80, 74], and (ii) Bayesian inference frameworks [96, 147]. When prior information is available, regularization methods are widely used to stabilize reconstruction, including classical Tikhonov regularization [152, 86], total variation (TV) [135], and ℓ_1-based sparsity methods [58, 90].

Throughout this paper, we assume that PDE-level well-posedness results (5)–(7) are available. Our focus is on transferring these theoretical guarantees to the finite-dimensional setting, establishing such finite-dimensional counterparts as (8).

2.3 Well-posedness for the finite-dimensional problem and design framework

It is tempting to transfer the infinite-dimensional well-posedness estimate (5) directly to its finite-dimensional counterpart (8). The passage, however, is not straightforward. A conceptual illustration appears in Figure 1.

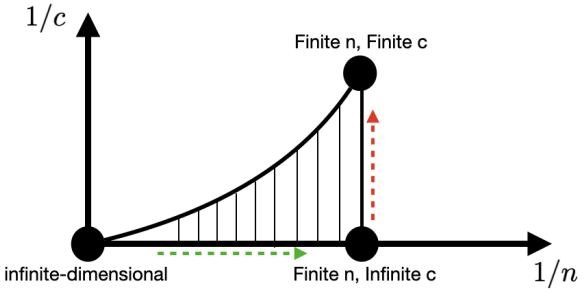


Figure 1: Conceptual pathway from infinite- to finite-dimensional formulations. The origin $(n, c) = (\infty, \infty)$ represents the infinite-dimensional setting, where well-posedness is established. The green arrow reduces the parameter space to finite dimension while keeping data abundant, leading to the semi-infinite case. The red arrow then reduces the data to finitely many experiments, posing the key challenges of identifiability and data selection.

At the origin $(n, c) = (\infty, \infty)$, the problem is posed in the infinite-dimensional setting, where well-posedness is assumed to hold. The most natural route to the fully finite-dimensional case proceeds first along the green arrow, followed by the red arrow.

The green arrow. This step reduces the problem to a semi-infinite setting: the unknown parameter is finite-dimensional, while data remains abundant. This reduction is essentially free. Once well-posedness (5) is established in the infinite-dimensional formulation, uniqueness readily follows in the semi-infinite case, with

$$|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2| \leq w_n \left(|\boldsymbol{\mathcal{M}}(\cdot, \boldsymbol{\sigma}_1) - \boldsymbol{\mathcal{M}}(\cdot, \boldsymbol{\sigma}_2)| \right), \quad \text{for } \sigma \in \Sigma_n \cap B_R(\boldsymbol{\sigma}_*), \tag{9}$$

for some function w_n. Since the amount of data used in the same, but the to-be-reconstructed parameter $\boldsymbol{\sigma}$ now lives in a smaller space: $\Sigma_n \subset \Sigma$, the stability function w_n inherits $w_n(0) = 0$ from w. Moreover, because Σ_n is finite dimensional and the analysis is restricted to a neighborhood of the ground truth, the domain under consideration is essentially compact, and the mapping is automatically Lipschitz [36].

Well-posedness in this semi-infinite regime has been studied extensively. For example, under the a priori assumption of piecewise-constant parameters, Alessandrini and Vessella [12] established Lipschitz stability for the Calderón problem [39]. Subsequent work has extended these results to more general settings [6, 30, 72, 136, 6, 10, 77, 22], as well as to Schrödinger and Helmholtz equations [29, 28, 11]. However, the Lipschitz stability constants may grow potentially exponentially with the parameter dimension n [132, 107].

The red arrow. In this (far more challenging) step, the available data is reduced from unlimited ($c = \infty$) to a finite c, while one seeks to maintain a comparable quality of reconstruction. This reduction raises two fundamental questions.

Q1. What is the minimal number of experiments c required to guarantee identifiability?

Q2. How should the subset $\Theta_c \subset \Theta$ be chosen?

These questions lie at the heart of data selection, which is essential for bridging continuous formulations with practical, finite data. Addressing them is central to PDE-based inverse problems and is the primary focus of this review.

Finally, one may also define a particular structure for Θ_c and examine only the scaling of c in terms of n. This approach was taken in a recent paper [8], where Θ_c is assumed to be a c-dimensional projection with a generic assumption.

2.4 Quantitative vs. Qualitative

The task of selecting data from a potentially infinite pool of experimental configurations is classical, and many approaches have been developed to address it. Broadly speaking, these approaches can be grouped into two categories.

– Quantitative design. This perspective is rooted in the Bayesian framework of *optimal experimental design*. Here, "design" refers to selecting, from all possible configurations, the experiments that most effectively extract information about the unknown parameter $\boldsymbol{\sigma}$. A quantitative design requires specifying a criterion Φ that measures the information content of a chosen subset $\Theta_c \subset \Theta$, then choosing the design that optimizes Φ, that is

$$\Theta_c^* = \arg\min_{\Theta_c \subset \Theta} \Phi(\Theta_c)\,.$$

Over the past decades, a wide variety of criteria ϕ and associated computational strategies have been developed. Among them, criteria based on minimizing uncertainty are particularly prominent: one models perturbations in the data and studies how they propagate into the reconstruction of $\boldsymbol{\sigma}$, then selects experiments that minimize the resulting uncertainty. See Section 3.2 for further details. For broader treatments of experimental design we refer to classical solvers [161, 20, 119, 162] and to the summaries in [65, 127, 138, 128]. Design used in PDE-based inverse problems context is reviewed in [13, 81, 46].

- Qualitative design. More recent approaches move away from strict optimality criteria. Instead, the goal is to secure reconstructions of reasonable quality [156, 157]. This relaxed perspective broadens the toolbox: techniques from randomized numerical linear algebra, gradient flows, and related areas that are traditionally viewed as separate from optimal design can now be applied effectively.

While quantitative design has long been the dominant paradigm, qualitative design has recently gained traction, particularly in connection with randomized numerical linear algebra (RNLA). As we discuss in Section 4, RNLA prioritizes efficiency while tolerating modest losses in precision. This philosophy aligns naturally with qualitative design, where "good enough" data selection suffices for practical purposes. For this reason, we are chiefly interested in qualitative approaches in this review.

2.5 Some representative PDE-based inverse problems

Several inverse problems arising from physical and biological applications have become canonical in the study of PDE-based inverse problems. We briefly summarize five such examples here. They are the inverse Lorenz 63 system, the inverse elliptic equation (also known as electrical impedance tomography, or the Calderón problem), the inverse Schrödinger equation, the inverse transport equation (as in optical tomography), and the inverse Helmholtz equation (arising in acoustic imaging).

E1. Inverse Lorenz 63 system. The Lorenz 63 system [111], originally introduced as a simplified model of atmospheric convection, describes the temporal evolution of three state variables:

$$\begin{cases} \dot{x} = \sigma(y - x)\,, \\ \dot{y} = x(\rho - z) - y\,, \\ \dot{z} = xy - \beta z\,. \end{cases} \tag{10}$$

Here, (x, y, z) correspond to convective flow rate, horizontal temperature difference, and vertical deviation of the temperature profile, while the parameter vector $\boldsymbol{\sigma} = (\sigma, \rho, \beta)$ encodes the Prandtl number, the Rayleigh number, and a geometric factor. Owing to its chaotic dynamics and bifurcation structure, the Lorenz system is a classical model in nonlinear dynamics.

In the inverse setting, given partial or full time-series observations of (x, y, z), the task is to reconstruct the parameter vector $\boldsymbol{\sigma}$. The associated ItO map is:

$$\Lambda_{\boldsymbol{\sigma}} : (x, y, z)(t = 0) = \theta_1 \mapsto (x, y, z)(t = \theta_2)\,.$$

Parameter identification for the Lorenz system has been widely studied; see, for instance, [83, 103, 142].

E2. Inverse elliptic equation. Elliptic equations provide the foundation of steady-state modeling. A typical formulation is:

$$\nabla \cdot (\boldsymbol{\sigma}(x)\nabla u) = 0\,, \quad \text{with} \quad u|_{\partial\Omega} = \phi\,. \tag{11}$$

The forward problem maps a prescribed Dirichlet boundary condition ϕ to the solution u. A canonical inverse problem is electrical impedance tomography (EIT) [50],

a noninvasive medical imaging technique in which one seeks to infer the conductivity distribution $\boldsymbol{\sigma}$ of tissue from boundary voltage–current measurements. In this case, the measured Neumann data $\boldsymbol{\sigma}\partial_n u$ defines the Dirichlet-to-Neumann (DtN) map:

$$\Lambda_{\boldsymbol{\sigma}} : \phi = \phi_{\theta_1} \mapsto \boldsymbol{\sigma}\, \partial_n u(\theta_2)\,.$$

Physically, this corresponds to the voltage-to-current relation. The EIT problem has a rich history; see surveys [33, 155, 66, 4]. Foundational works established uniqueness and stability results [149, 100, 9, 121, 19].

E3. Inverse Schrödinger equation The Schrödinger equation is a fundamental model for describing wave fields in quantum systems. Its steady-state form is a semilinear elliptic equation:

$$\Delta_x u + \boldsymbol{\sigma}(x)u = 0\,, \quad \text{with} \quad u|_{\partial\Omega} = \phi\,. \tag{12}$$

where u is the wave field and $\boldsymbol{\sigma}$ the potential. The inverse problem seeks to reconstruct $\boldsymbol{\sigma}$ from boundary measurements of u. Applications include material characterization and design for achieving prescribed properties [116, 140, 71]. The corresponding boundary map is:

$$\Lambda_{\boldsymbol{\sigma}} : u|_{\partial\Omega} \to \partial_n u\big|_{\partial\Omega}\,.$$

From a mathematical perspective, the inverse Schrödinger problem can be related to EIT: Under smoothness assumptions on the medium, the Schrödinger potential ($\boldsymbol{\sigma}$ in (12)) can be written as $\frac{\Delta\sqrt{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{\sigma}}}$ from (11), linking the two formulations [121, 66].

E4. Inverse transport equation. Transport equations are widely used in optical and radar imaging. The underlying principle is to infer the optical properties of a medium from measurements of incoming and outgoing photon fluxes [27, 67, 32, 120]. These material properties are encoded in the coefficient $\boldsymbol{\sigma}$. The governing PDE is

$$v \cdot \nabla_x f = \boldsymbol{\sigma}(x)\,\mathcal{L}[f]\,, \quad \text{with} \quad f|_{\Gamma_-} = \phi(x, v)\,,$$

where $f(x, v)$ denotes the photon distribution at spatial location x and velocity v, and $\mathcal{L}$ is a prescribed scattering operator. The domain is $x \in \Omega$, with $\Gamma_\pm = \{(x, v) : x \in \partial\Omega, \pm v \cdot n_x > 0\}$ representing inflow and outflow boundaries. The associated inverse problem seeks to reconstruct $\boldsymbol{\sigma}$ from boundary measurements, which defines the ItO map:

$$\Lambda_{\boldsymbol{\sigma}} : \phi_{\theta_1} \mapsto f|_{\Gamma_+}(\theta_2)\,.$$

Inverse transport problems are fundamental in optical tomography and have been studied extensively from both analytical and computational perspectives [104, 52, 158, 23, 17, 130].

E5. Inverse Helmholtz equation. The Helmholtz equation models time-harmonic wave propagation and forms the foundation of acoustic imaging, with applications in geophysics [31, 106, 148] and medical imaging [133]. The PDE takes the form

$$\Delta_x u + \omega^2 \boldsymbol{\sigma}(x)u = 0\,,$$

where ω is the prescribed frequency. The incoming wave is specified as a plane wave $u_{\theta_1}^{\text{in}} = e^{i\omega\theta_1\cdot x}$. Measurements are collected on the boundary of a large domain (typically modeled as a ball of radius R, denoted B_R). The inverse problem consists of reconstructing the coefficient $\boldsymbol{\sigma}$ using the ItO map:

$$\Lambda_{\boldsymbol{\sigma}} : u_{\theta_1}^{\text{in}} \mapsto u|_{\partial B_R}(\theta_2)\,.$$

The inverse Helmholtz scattering problem has been the subject of extensive research from mathematical, statistical, and numerical perspectives. Works that summarize the progress and highlight key advances in theory and algorithm design include the books [97, 69, 55, 16, 34, 85] and landmark works [150, 35, 87, 24, 64, 151, 34].

3 Technical preparation and classical tools

Experimental design is an upstream task that precedes the implementation and execution of inverse algorithms. The value of the collected data ultimately depends on how effectively it can be exploited by the chosen solver. Two major algorithmic pipelines have been particularly influential: *PDE-constrained optimization* and the *Bayesian formulation*, both of which we summarize in the following subsections. A recurring theme in both frameworks is the central role of *linearization* that requires gradient computation. We therefore devote Section 3.3 to surveying the linearization techniques.

3.1 PDE-constrained optimization

When data is abundant, PDE-constrained optimization is one of the most widely used methods for inferring unknown parameters in PDEs. The formulation is straightforward: Subject to the PDE being satisfied, we seek the parameter $\boldsymbol{\sigma}$ that produces the best match to the measurement data:

$$\min_{\boldsymbol{\sigma}} \sum_{\theta_1,\theta_2} \frac{1}{2}|\mathrm{d}(\theta) - M_{\theta_2}[u_{\theta_1}]|^2\,, \quad \text{s.t.} \quad \mathcal{L}_{\boldsymbol{\sigma}}[u_{\theta_1}] = S_{\theta_1}\,. \tag{13}$$

If the data d is noiseless (as when it is generated by the PDE solution using a ground-truth parameter $\boldsymbol{\sigma}^*$) and the problem is well-posed, then the optimizer is unique and yields a zero objective value. Since the PDE itself is well-posed, u_{θ_1} is uniquely determined; we therefore write it as $u(\boldsymbol{\sigma};\theta_1)$ to emphasize its dependence on $\boldsymbol{\sigma}$. The optimization problem can then be recast in an unconstrained form:

$$\min_{\boldsymbol{\sigma}} L(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{\theta_1,\theta_2} |\mathrm{d}(\theta) - M_{\theta_2}[u(\boldsymbol{\sigma};\theta_1)]|^2 = \frac{1}{2}\sum_{\theta_1,\theta_2} |\mathrm{d}(\theta) - \mathrm{pde}(\boldsymbol{\sigma},\theta))|^2\,,$$

where we adopt the shorthand $\mathrm{pde}(\boldsymbol{\sigma},\theta) = M_{\theta_2}[u(\boldsymbol{\sigma};\theta_1)]$.

In practice, gradient-based solvers are most commonly used to minimize $L(\boldsymbol{\sigma})$, updating $\boldsymbol{\sigma}$ by making use of the gradient $\frac{d}{d\boldsymbol{\sigma}}L(\boldsymbol{\sigma})$. The conditioning of the problem, which governs both stability and convergence properties, depends critically on the Hessian $\mathrm{Hess}_{\boldsymbol{\sigma}}L$. Calculation of the gradient and Hessian can be summarized as follows.

1. **Gradient** $\frac{d}{d\boldsymbol{\sigma}}L$. By the chain rule, we have

$$\frac{d}{d\boldsymbol{\sigma}}L = \sum_{\theta} \left(\mathrm{pde}(\boldsymbol{\sigma},\theta) - \mathrm{d}(\theta)\right) \frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma},\theta)\,.$$

   Since the dependence on $\boldsymbol{\sigma}$ appears only through $\mathrm{pde}(\boldsymbol{\sigma},\theta)$, computing the gradient amounts to evaluating $\frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}$.

2. **Hessian** $\mathrm{Hess}_{\boldsymbol{\sigma}}L$**.** Differentiating once more gives

$$\mathrm{Hess}_{\boldsymbol{\sigma}}L \;=\; \underbrace{\sum_{\theta} \frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma},\theta) \otimes \frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma},\theta)}_{\text{Fisher Information Matrix}} + \sum_{\theta} \big(\mathrm{pde}(\boldsymbol{\sigma},\theta) - \mathrm{d}(\theta)\big)\, \mathrm{Hess}_{\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma},\theta)\,. \tag{14}$$

   The Hessian thus consists of two components. The first is the outer product of the pde gradients, usually referred to as the Fisher Information Matrix (FIM) or the Gauss-Newton matrix. Near the ground truth, where $\mathrm{pde}(\boldsymbol{\sigma},\theta) \approx \mathrm{d}(\theta)$, the second term vanishes, leaving the Hessian dominated by the FIM. Note that when $\boldsymbol{\sigma} \in \Sigma$, a function space, the Hessian is an operator, whereas for $\boldsymbol{\sigma} \in \Sigma_N$ it reduces to a finite-dimensional matrix in $\mathbb{R}^{N\times N}$.

Both the gradient and the Fisher Information Matrix (FIM) will reappear as key quantities in experimental design, since they provide the sensitivity information that guides the selection of data so as to maximize its impact on parameter inference.

3.2 Bayesian formulation

Bayes' formula provides a widely used alternative to PDE-constrained optimization, particularly when we need to quantify uncertainty of reconstruction. Instead of seeking a single best-fit parameter, the Bayesian perspective treats a range of parameters as plausible, each weighted by its posterior probability.

In this framework, prior knowledge is blended with observed data. For PDE-based inverse problems, we model the data as

$$\mathrm{d}(\theta) = \mathrm{pde}(\boldsymbol{\sigma},\theta) + \eta(\theta)\,,$$

where η represents measurement error. The posterior distribution of $\boldsymbol{\sigma}$ — which is also the conditional distribution of $\boldsymbol{\sigma}$ given the data d — takes the form

$$p^{\mathrm{d}}(\boldsymbol{\sigma}) = p(\boldsymbol{\sigma}|\mathrm{d}) \propto p(\mathrm{d}|\boldsymbol{\sigma})p(\boldsymbol{\sigma}) = p_{\eta}\left(\mathrm{d}(:) - \mathrm{pde}(\boldsymbol{\sigma},:)\right) p(\boldsymbol{\sigma})\,,$$

where $p(\boldsymbol{\sigma})$ is the *marginal* or *prior* distribution distribution and p_{η} is the noise distribution.

As in optimization, the posterior distribution has a few key quantities, one of which is the variance, which measures the uncertainty of the reconstruction. In general, the variance is difficult to compute for nonlinear PDE operators. However, in the linearized setting, the posterior distribution is well understood. The following assumptions are standard.

- The PDE operator is linear, e.g. $\mathrm{pde}(\boldsymbol{\sigma},\theta) = \mathbf{A}_{\theta,:}\boldsymbol{\sigma}$, or locally linearizable around $\boldsymbol{\sigma}^*$ as $\mathrm{pde}(\boldsymbol{\sigma},\theta) \approx \mathbf{A}_{\theta,:}\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}^*\right) + \mathrm{pde}(\boldsymbol{\sigma}^*,\theta)$, with data adjusted accordingly.

- Noise is Gaussian: $\eta \sim \mathcal{N}(0, \Gamma)$;
- The prior is Gaussian: $\boldsymbol{\sigma} \sim \mathcal{N}(\boldsymbol{\sigma}^\dagger, \Gamma_0)$.

Under these assumptions, the posterior remains Gaussian. When the data $\mathrm{d}(\theta)$ is assigned weights w_θ, the posterior distribution is

$$p^{\mathrm{d}}(\boldsymbol{\sigma}) \sim \mathcal{N}(\hat{\boldsymbol{\sigma}}_W, \hat{\Gamma}_W) \quad \text{with} \quad \hat{\Gamma}_W = (\mathbf{A}^\top W^{1/2} \Gamma^{-1} W^{1/2} \mathbf{A} + \Gamma_0^{-1})^{-1}\,, \tag{15}$$

and the mean is $\hat{\boldsymbol{\sigma}}_W = \hat{\Gamma}_W \left(\mathbf{A}^\top W^{1/2} \Gamma^{-1} W^{1/2} \mathrm{d} + \Gamma_0^{-1} \boldsymbol{\sigma}^\dagger\right)$. Here $W = \mathrm{diag}\{w_\theta\}$. Smaller posterior variance $\hat{\Gamma}_W$ corresponds to greater certainty and hence more accurate reconstruction. Thus, experimental design is done in a way that minimizes some scalar measure of $\hat{\Gamma}_W$.

Because $\hat{\Gamma}_W$ is a matrix, different scalarizations yield different design criteria. The following are most prominent.

- E-optimal design: $\Phi_E = \lambda_{\max}(\hat{\Gamma}_W)$.
- D-optimal design: $\Phi_D(\theta) = \det\Big(\hat{\Gamma}_W\Big)$.
- A-optimal design: $\Phi_A(\theta) = \mathrm{tr}(\hat{\Gamma}_W)$.

The experimental design problem is to select weights $\{w_\theta\}$ that minimize one of these criteria:

$$\min_{\{w_\theta\}} \ \Phi_{E,D,A}\,, \tag{16}$$

assignoing higher weights to data most effective at reducing uncertainty. In the semi-infinite case, where $\boldsymbol{\sigma} \in \Sigma_N$ and $\hat{\Gamma}_W \in \mathbb{R}^{N \times N}$, the sequence w_θ may still be infinite in size. Hence, the optimization is conducted over the entire design function space Θ.

Bayesian formulation, together with the associated design principle (16) based on uncertainty reduction, is one of the leading approaches to data selection. As in the optimization framework, the matrix $\mathbf{A}$, which captures the linear component of $\mathrm{pde}(\boldsymbol{\sigma}, \theta)$, plays a central role in quantifying the information content of the data θ.

3.3 Linearization

Linearization plays a central role in PDE-based inverse problems. As seen above, derivatives such as $\frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}$ arise repeatedly in both optimization and Bayesian formulations. Theoretically, linearization is also tied to the notation $B_R(\boldsymbol{\sigma}_*)$ in the well-posedness estimates (5), (9), and (8). In many cases, well-posedness can be guaranteed only within a neighborhood of the ground truth $\boldsymbol{\sigma}_*$. This locality is especially relevant in the finite-dimensional setting: Close to two different underlying truth $\boldsymbol{\sigma}_{*,1}$ and $\boldsymbol{\sigma}_{*,2}$, the same measurement $\mathrm{d}(\theta_1, \theta_2)$ may exhibit very different sensitivities to parameter variations. Consequently, the type and amount of data needed for stable reconstruction depend on the specific ground truth $\boldsymbol{\sigma}_*$.

To address this issue, one must examine how the data $\mathrm{d}(\theta_1, \theta_2)$ varies with perturbations of the parameter $\boldsymbol{\sigma}$ near a given $\boldsymbol{\sigma}_*$. Linearization and functional gradient computations provide precisely this information [146]. For PDE-based inverse problems, these tools can be

formulated systematically using the calculus of variations. Following the notation of (1), the linearized problem typically reduces to a Fredholm integral equation of the first kind:

$$\left\langle \frac{d}{d\boldsymbol{\sigma}} \mathrm{pde}(\boldsymbol{\sigma}, \theta), \delta\boldsymbol{\sigma} \right\rangle = \left\langle \frac{d}{d\boldsymbol{\sigma}} M_{\theta_2}[u(\boldsymbol{\sigma}; \theta_1)], \delta\boldsymbol{\sigma} \right\rangle = \delta \mathrm{d}_\theta \,, \tag{17}$$

where

$$\delta\boldsymbol{\sigma} = \boldsymbol{\sigma} - \boldsymbol{\sigma}_* \quad \text{and} \quad \delta \mathrm{d} = \mathrm{d}(\theta) - M_{\theta_2}[u_{\theta_1}(\boldsymbol{\sigma}_*)] = \mathrm{d}(\theta) - \mathrm{pde}(\boldsymbol{\sigma}_*, \theta) \,.$$

The inverse problem in this linearized form is to use pairs of $\left(\frac{d}{d\boldsymbol{\sigma}} M[u(\boldsymbol{\sigma}; \theta)] \,, \delta \mathrm{d}\right)$, collected over many θ, to reconstruct $\delta\boldsymbol{\sigma}$.

While the computation of δd is straightforward, the central challenge lies in evaluating the functional gradient $\frac{d}{d\boldsymbol{\sigma}} \mathrm{pde}(\boldsymbol{\sigma}, \theta) = \frac{d}{d\boldsymbol{\sigma}} M_{\theta_2}[u(\boldsymbol{\sigma}; \theta_1)]$, subject to the PDE constraint $\mathcal{L}_{\boldsymbol{\sigma}}[u] = S_{\theta_1}$. Different PDEs yield different functional gradients, but a common structure emerges:

$$\frac{d}{d\boldsymbol{\sigma}} M_{\theta_1}[u(\boldsymbol{\sigma}; \theta_2)] = f_{\theta_1} \cdot g_{\theta_2}$$

where f_{θ_1} is associated with the forward PDE solution u (driven by source θ_1), and g_{θ_2} is associated with the adjoint PDE solution that is driven by detector θ_2. Bu substituting this form into (17), renaming $\mathbf{b} := \delta \mathrm{d}$, and denoting $\boldsymbol{\sigma}$ as $\delta\boldsymbol{\sigma}$, we obtain the following compact expression:

$$\langle f_{\theta_1} g_{\theta_2}, \boldsymbol{\sigma} \rangle = \mathbf{b}(\theta_1, \theta_2) \,. \tag{18}$$

This tensorized structure in the Fredholm integral is a distinctive feature of PDE-based inverse problems, and it will serve as the foundation for the derivations that follow. While linearization itself is often treated as a black-box solver and may be of secondary importance for the purposes of this review, the tensorized structure plays a central role in data selection. In particular, effective designs must be compatible with this structure, a requirement that distinguishes PDE-based inverse problems from other experimental design settings.

We now describe the main techniques for computing functional gradients. (Readers already familiar with adjoint-based methods may skip this discussion.) For simplicity, we fix $\theta = (\theta_1, \theta_2)$ throughout this subsection and omit the subscript.

Linearized PDE operator via calculus of variations. The general recipe is straightforward. Given the PDE constraint $\mathcal{L}_{\boldsymbol{\sigma}}[u] = S$, we regard $u = u(\boldsymbol{\sigma})$ as a function of the parameter $\boldsymbol{\sigma}$. Differentiation of the observation functional then yields

$$\frac{d}{d\boldsymbol{\sigma}} M[u(\boldsymbol{\sigma})] = \left\langle \frac{\delta M}{\delta u} \,, \frac{du}{d\boldsymbol{\sigma}} \right\rangle . \tag{19}$$

Typically, the variational derivative $\frac{\delta M}{\delta u}$ is explicit; the main difficulty lies in evaluating $\frac{du}{d\boldsymbol{\sigma}}$.

To compute this term, we differentiate the PDE constraint. Since $\mathcal{L}_{\boldsymbol{\sigma}}[u(\boldsymbol{\sigma})] = S$ holds identically, we have

$$\left.\frac{d\mathcal{L}_{\boldsymbol{\sigma}}}{d\boldsymbol{\sigma}}\right|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} = \left.\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}\right|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} + D_u \mathcal{L}_{\boldsymbol{\sigma}}|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} \cdot \frac{du}{d\boldsymbol{\sigma}} = 0 \,,$$

or equivalently,

$$D_u \mathcal{L}_{\boldsymbol{\sigma}}|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} \cdot \frac{du}{d\boldsymbol{\sigma}} = - \left.\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}\right|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} .$$

Here both $D_u\mathcal{L}_{\boldsymbol{\sigma}}$ and $\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}$ are explicit from the PDE. Direct inversion of $D_u\mathcal{L}_{\boldsymbol{\sigma}}$, however, is often prohibitively expensive.

A standard alternative is the **adjoint method**. Returning to (19), we introduce an adjoint variable λ defined by

$$(D_u\mathcal{L}_{\boldsymbol{\sigma}})^* \lambda = \frac{\delta M}{\delta u}. \tag{20}$$

Multiplying the differentiated PDE constraint by λ and integrating by parts yields:

$$\begin{aligned} \frac{d}{d\boldsymbol{\sigma}} M[u(\boldsymbol{\sigma})] &= \left\langle \frac{\delta M}{\delta u}, \frac{du}{d\boldsymbol{\sigma}} \right\rangle = \left\langle (D_u\mathcal{L}_{\boldsymbol{\sigma}})^* \lambda, \frac{du}{d\boldsymbol{\sigma}} \right\rangle \\ &= \left\langle \lambda, D_u\mathcal{L}_{\boldsymbol{\sigma}} \cdot \frac{du}{d\boldsymbol{\sigma}} \right\rangle = - \left\langle \lambda, \left. \frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}} \right|_{\boldsymbol{\sigma}, u(\boldsymbol{\sigma})} \right\rangle. \end{aligned} \tag{21}$$

This adjoint formulation is highly efficient: only one forward PDE solve and one adjoint PDE solve (20) are required, independent of the dimension of $\boldsymbol{\sigma}$.

Since $M[u(\boldsymbol{\sigma})]$ maps a function $\boldsymbol{\sigma}$ to a scalar, its derivative with respect to $\boldsymbol{\sigma}$ is itself a function. Consequently, the terms $\frac{du}{d\boldsymbol{\sigma}}$ and $\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}$ in (21) should be interpreted as operators, with the inner product taken in the operator–function sense. This viewpoint naturally motivates a projected formulation, which is often more convenient in practice. We have

$$\left\langle \tfrac{d}{d\boldsymbol{\sigma}} M[u(\boldsymbol{\sigma})], \delta\boldsymbol{\sigma} \right\rangle = - \left\langle \lambda, \tfrac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}[\delta\boldsymbol{\sigma}] \right\rangle, \tag{22}$$

where $\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}[\delta\boldsymbol{\sigma}]$ denotes the action of $\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}$ on $\delta\boldsymbol{\sigma}$. This projected formulation avoids explicit construction of the full functional derivative and directly yields directional gradients, making it highly practical.

Example 1 (Linearized EIT **E2.**)**.** Consider the elliptic PDE $\mathcal{L}_{\boldsymbol{\sigma}}[u] = \nabla \cdot \big(\boldsymbol{\sigma}(x)\nabla u\big)$ corresponding to (11). The ingredients of the functional derivative are as follows.

1. **Operator derivative** $\frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}}$. By definition, we have

$$\left. \frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}} \right|_{\boldsymbol{\sigma}, u} [\delta\boldsymbol{\sigma}] = \left. \frac{d}{d\epsilon} \mathcal{L}_{\boldsymbol{\sigma}+\epsilon\delta\boldsymbol{\sigma}}[u] \right|_{\epsilon=0},$$

   where the expression is evaluated at $(\boldsymbol{\sigma}, u)$ and applied in the direction $\delta\boldsymbol{\sigma}$. Observing that

$$\mathcal{L}_{\boldsymbol{\sigma}+\epsilon\delta\boldsymbol{\sigma}}[u] = \nabla \cdot \big((\boldsymbol{\sigma} + \epsilon\delta\boldsymbol{\sigma})\nabla u\big) = \nabla \cdot (\boldsymbol{\sigma}\nabla u) + \epsilon\, \nabla \cdot (\delta\boldsymbol{\sigma}\nabla u),$$

   we obtain

$$\left. \frac{\partial \mathcal{L}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\sigma}} \right|_{\boldsymbol{\sigma}, u} [\delta\boldsymbol{\sigma}] = \nabla \cdot (\delta\boldsymbol{\sigma}\nabla u).$$

2. **Fréchet derivative in** u, $D_u\mathcal{L}_{\boldsymbol{\sigma}}$. We know that

$$D_u\mathcal{L}_{\boldsymbol{\sigma}}[\delta u] = \nabla \cdot (\boldsymbol{\sigma}\nabla \delta u).$$

   This operator is self-adjoint, so we have $(D_u\mathcal{L}_{\boldsymbol{\sigma}})^* \lambda = \nabla \cdot (\boldsymbol{\sigma}\nabla\lambda)$.

By combining these results with (22), we obtain

$$\left\langle \frac{d}{d\boldsymbol{\sigma}} M[u(\boldsymbol{\sigma})]\,, \delta\boldsymbol{\sigma} \right\rangle = -\left\langle \lambda,\, \nabla\cdot(\delta\boldsymbol{\sigma}\nabla u)\right\rangle = \left\langle \nabla\lambda\cdot\nabla u,\, \delta\boldsymbol{\sigma}\right\rangle .$$

where we used integration-by-parts. By reinstating the dependence on $\theta = (\theta_1, \theta_2)$, we have

$$\tfrac{d}{d\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma}, \theta) = \nabla\lambda_{\theta_2}\cdot\nabla u_{\theta_1},, \tag{23}$$

where u_{θ_1} and λ_{θ_2} are solutions of the forward and adjoint problems, respectively. The linearized inverse problem becomes

$$\left\langle \nabla\lambda_{\theta_2}\cdot\nabla u(\boldsymbol{\sigma}^*;\theta_1), \delta\boldsymbol{\sigma}\right\rangle = \mathrm{d}(\theta) - M[u(\boldsymbol{\sigma}^*;\theta_1)]\,. \tag{24}$$

This example highlights the decoupling of u and λ associated with source θ_1 and detector θ_2, respectively, in consistency with the tensorized structure discussed in (18).

4 Randomized linear algebra techniques

Randomized numerical linear algebra (RNLA) incorporates randomization into the design of algorithms for numerical linear algebra algorithms. In contrast to classical deterministic methods, randomized algorithms deliberately trade accuracy for efficiency. They are particularly effective for large-scale problems or for settings that require repeated linear operations, where computational cost rather than a need for high precision is the primary bottleneck. Early applications of RNLA arose in such data-science domains as imaging, signal processing, and internet-scale problems, where speed is of paramount importance [115, 113, 159]. More recently, applications in the physical sciences have emerged, where accuracy demands must be balanced against computational feasibility, making RNLA both powerful and practical.

PDE-based experimental design aligns naturally with this perspective. Inverse problems governed by PDEs ultimately prioritize reconstruction accuracy. Yet before reconstruction can take place, one must often examine a large pool of potential data to identify informative subsets. This selection step is typically repeated many times and, crucially, it does not require exact precision. RNLA algorithms are well-suited to this step, as they accelerate the repeated linear algebra computations underlying data selection while maintaining enough accuracy to guide the process.

A classical entry point for RNLA in experimental design is through leverage scores. Leverage scores are statistical quantities that measure the importance of each constraint in a linear regression problem, and they have long been used to guide data selection. However, they are expensive to compute exactly, which is why randomization becomes valuable. To illustrate, consider the linear regression problem:

$$\min_{\boldsymbol{\sigma}} \|\mathbf{A}\boldsymbol{\sigma} - \mathbf{b}\|_2\,.$$

The optimal solution and its projection are explicitly given by

$$\boldsymbol{\sigma}^* = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}\,, \quad \text{and} \quad \mathbf{b}^* = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}\,.$$

Here $\mathbf{A} \in \mathbb{R}^{C\times n}$ is the design matrix, with each row encoding one constraint, and $\mathbf{b}$ collects the data. The influence of the data $\mathbf{b}$ on the fitted output $\mathbf{b}^*$ can be quantified through the Jacobian:

$$\nabla_{\mathbf{b}}\mathbf{b}^* = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top \doteq \mathbf{H}\,,$$

known as the hat matrix (as it is often, though not here, denoted by $\hat{\mathsf{H}}$). Its diagonal entries, $l_i = \mathbf{H}_{ii}$ define the leverage scores, which measure the relative importance of the i-th observation. High-leverage points correspond to experiments with strong influence on the regression fit. Leverage scores are widely used, for example, in outlier detection in economics and medical imaging [163, 118], kernel methods [21, 105] and graph learning [144].

Computing $\mathbf{H}$ presents challenges in computation and data access. Forming this matrix directly requires $O(Cn^2)$ flops, which is costly for large problems. Even more problematic is the sample complexity: Evaluating leverage scores requires explicit knowledge of the entire matrix $\mathbf{A}$, meaning all experiments must be conducted and formulated before any selection is possible. Randomized methods offer a way to address both challenges: They reduce the computational cost to near-linearity in C and n and allow approximate leverage scores to be estimated from only partial or sampled information about $\mathbf{A}$ [109, 61, 115].

While these techniques were not originally developed for PDE-based inverse problems, they represent one of the earliest demonstrations of how RNLA can be useful for experimental design. Over the past decade, this perspective has broadened: Methods such as matrix sketching, randomized matrix–matrix multiplication, column selection, randomized singular value decomposition (SVD), compressed sensing, and matrix completion have all been adapted to PDE settings. Collectively, they provide ways to accelerate and guide data selection. In what follows, we first review these methods in their general linear algebraic form, and then describe how they can be specialized for PDE-based experimental design.

4.1 Matrix sketching

Matrix sketching is an important problem class in RNLA. Conceptually, it can be understood in analogy to regression: in classical least-squares regression, one may "sketch" by reducing the row dimension of the system in a principled way, hoping to obtain a solution that is nearly as good as the full system.

Formally, consider the overdetermined linear system

$$\mathbf{A}\,\boldsymbol{\sigma} \approx \mathbf{b}\,, \quad \mathbf{A} \in \mathbb{R}^{C\times n}\,, C \gg n\,,$$

whose least-squares solution is

$$\boldsymbol{\sigma}^* = \arg\min_{\boldsymbol{\sigma}\in\mathbb{R}^n} \|\mathbf{A}\boldsymbol{\sigma} - \mathbf{b}\|_2^2\,. \tag{25}$$

Since the system is highly overdetermined, a reasonable approximation can often be obtained by using only a reduced set of rows from $\mathbf{A}$. To achieve this, one introduces a sketching matrix $\mathbf{S} \in \mathbb{R}^{c\times C}$ with $c \ll C$ and solves the reduced problem:

$$\boldsymbol{\sigma}^*_{\mathbf{S}} \;=\; \arg\min_{\boldsymbol{\sigma}\in\mathbb{R}^n} \|\mathbf{S}\mathbf{A}\boldsymbol{\sigma} - \mathbf{S}\mathbf{b}\|_2^2\,, \tag{26}$$

with the goal that $\boldsymbol{\sigma}^*_{\mathbf{S}} \approx \boldsymbol{\sigma}^*$. A typical guarantee is:

$$\|\mathbf{A}\boldsymbol{\sigma}^*_{\mathbf{S}} - \mathbf{b}\|_2^2 \le (1+\epsilon)\|\mathbf{A}\boldsymbol{\sigma}^* - \mathbf{b}\|_2^2, \quad \text{with probability } 1-\delta. \tag{27}$$

The mechanism is illustrated in Figure 2.

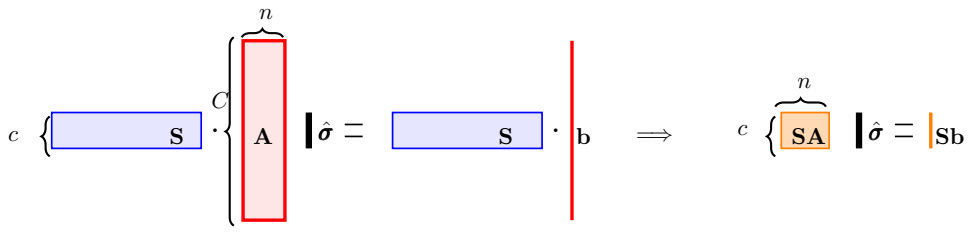


Figure 2: Sketching mechanism.

The advantages of sketching are twofold. From a **computational perspective**, solving (25) requires $O(Cn^2)$ flops, while (26) requires only $O(\min(cn^2, c^2n))$ with $c \ll C$, yielding significant savings. From an **experimental design** perspective, the savings are even more pronounced in sample complexity: the full problem requires C measurements (the rows of $\mathbf{A}$ and entries of $\mathbf{b}$), whereas the sketched problem reduces this to c measurements, substantially cutting the effort in data acquisition.

The main challenge is to construct a sketching matrix $\mathbf{S}$ that ensures the guarantee (27) while maintaining the property $c \ll C$. Without entering technical details, we highlight several widely used strategies (see [159, 112]):

- SubGaussian matrices [134]. The sketching matrix is generated by i.i.d. zero-mean, isotropic subGaussian entries. Common subGaussian types include Gaussian, Bernoulli, uniform distributions.
- Fast Johnson-Lindenstrauss transform (FJLT) [5, 154]. The sketching matrix is constructed as $\mathbf{\Pi}\boldsymbol{\mathcal{F}}$, where $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{C\times C}$ represents a discrete Fourier or Hadamard transform with randomly sign-flipped columns and $\mathbf{\Pi} \in \mathbb{R}^{c\times C}$ denotes a random row sampling operation.
- Sparse sketching matrices [2, 45]. Particularly useful in streaming settings, these matrices have entries in $\{0, +1, -1\}$, chosen according to a suitable distribution. Examples include CountSketch [45] and sparse random projection [2].

The three design principles are visualized in Figure 3.

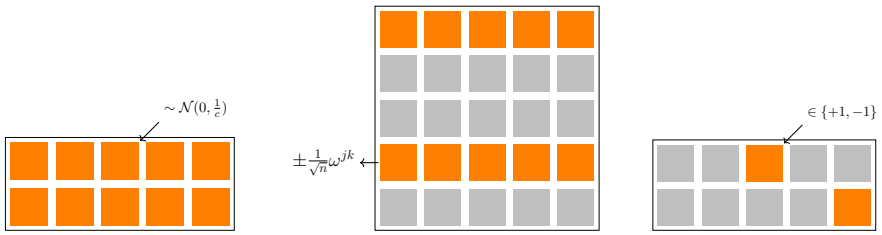
While the precise dependence of c on ϵ and δ varies across these constructions, their theoretical justification is unified by the celebrated Johnson–Lindenstrauss lemma [95], which guarantees the existence of low-dimensional embeddings that preserve geometry with high probability.

4.1.1 Matrix sketching for PDE based experimental design

Sketching aims to provide an almost equally good reconstruction to a regression problem using only a subsampled system. This perspective aligns naturally with experimental design. However, PDE-based inverse problems exhibit a special structure that prevents a direct use of classical sketching. Most notably, as shown in (18), the linearized PDE-based inverse problem typically takes a tensorized form.

We rewrite the linearized system in algebraic form:

$$\mathbf{A} \cdot \boldsymbol{\sigma} = \mathbf{b}\,, \tag{28}$$



(a) subGaussian (b) FJLT (c) sparse

Figure 3: Major classes of sketching matrix choices. Orange squares indicate nontrivial entries. Gray squares represent unsampled entries in (b), and zero-valued entries in (c). In (b) FJLT, the (j, k)-th entry of $\boldsymbol{\mathcal{F}}$ is given by $\pm\frac{1}{\sqrt{n}}\omega^{jk}$, where $\omega = \exp(-\frac{2\pi \mathrm{i}}{n})$ is the n-th root of unity, and jk is the exponent. The sign (positive or negative) is determined by a random sign-flip applied to its column.

where $\mathbf{b}$ encodes the right-hand side of (17) (with entries indexed by θ), and $\mathbf{A}$ is essentially the Jacobian matrix. Unlike in classical regression, $\mathbf{A}$ is infinite in size: its rows are indexed by $\theta = (\theta_1, \theta_2) \in \Theta$, the entire design space, while its columns are indexed by the discretization of $\boldsymbol{\sigma}$. Restricting to $\boldsymbol{\sigma} \in \Sigma_n$ with dimension n, each row of $\mathbf{A}_{\theta,:}$ lies in $\mathbb{R}^n$.

The key distinction from standard regression is the tensorized structure of $\mathbf{A}$. In numerical setting of sketching, we focus on the finite but possibly large data space as Θ_C. By (17), $\mathbf{A}$ can be expressed columnwise as a Khatri–Rao product. Let

$$\mathbf{F} \in \mathbb{R}^{C_1 \times n}, \quad \mathbf{G} \in \mathbb{R}^{C_2 \times n} \quad \text{with} \quad \mathbf{F}_{\theta_1,:} = f_{\theta_1}, \quad \mathbf{G}_{\theta_2,:} = g_{\theta_2},$$

where $\theta_i \in \Theta_i$ and $|\Theta_i| = C_i$ and $C = C_1 C_2$. Then

$$\mathbf{A} = \mathbf{F} * \mathbf{G} \in \mathbb{R}^{C_1 C_2 \times n} \quad \text{with} \quad \mathbf{A}_{\theta,:} = \mathbf{F}_{\theta_1,:} \mathbf{G}_{\theta_2,:}.$$

This structure requires a tensor-aware sketching procedure, with the sketching matrix also having a row-wise tensorized form, so that $\mathbf{F}$ and $\mathbf{G}$ are sketched independently. Physically, this corresponds to the fact that the parameter specifying the PDE input (θ_1) and the parameter specifying the detector (θ_2) are independently configurable in a fully factorized design. Two representative constructions along these lines are the following.

- SubGaussian tensorizing. Each row of the sketching matrix $\mathbf{S}$ is written as $\mathbf{S}_{i,:} = \mathbf{S}^{\mathbf{F}}_{i,:} \otimes \mathbf{S}^{\mathbf{G}}_{i,:}$, where $\mathbf{S}^{\mathbf{F}}$ and $\mathbf{S}^{\mathbf{G}}$ are filled with i.i.d. zero-mean isotropic subGaussian entries.
- Kronecker FJLT. The transform $\boldsymbol{\Pi}\boldsymbol{\mathcal{F}}$ is modified to the tensorized form $\boldsymbol{\Pi}\big(\boldsymbol{\mathcal{F}}^{\mathbf{F}} \otimes \boldsymbol{\mathcal{F}}^{\mathbf{G}}\big)$, so that $\boldsymbol{\mathcal{F}}^{\mathbf{F}}$ and $\boldsymbol{\mathcal{F}}^{\mathbf{G}}$ act separately on $\mathbf{F}$ and $\mathbf{G}$.

These mechanisms are illustrated in Figure 4.

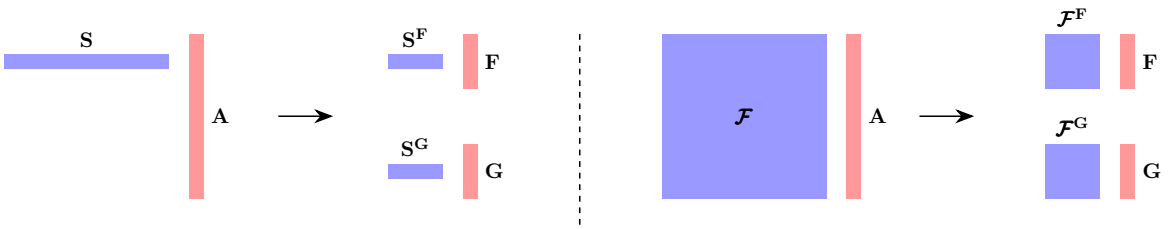


Figure 4: Tensor sketching acceleration: the complexity of computing $\mathbf{SA}$ breaks down to the scale of each factor matrix size.

Tensor-structured sketching has been extensively developed in recent years [117, 48, 92, 47, 88], often relying on the distributive property $(\mathbf{B}\otimes\mathbf{D})(\mathbf{C}\otimes\mathbf{E}) = (\mathbf{BC})\otimes(\mathbf{DE})$. In PDE-based inverse problems, the Khatri–Rao product induces the tensorized requirement, with notable works including [48, 47]. Likewise, [92] developed Kronecker Fourier-based random projections (Kronecker FJLT), extending standard FJLT.

Finally, we highlight the issue of sampling complexity: How many randomly sampled measurements are sufficient to ensure that the sketched solution $\boldsymbol{\sigma}_{\mathbf{S}}^*$ approximates the full solution? A trade-off emerges: The tensor structure accelerates computations but introduces higher-order Gaussian chaos, complicating probabilistic analysis. For instance, Theorem 1.1 in [47] establishes a nearly optimal bound:

$$c \geq \max\left(\mathcal{O}\left(\frac{\mathrm{rank}^2(\mathbf{A}) + \log^2(1/\delta)}{\epsilon}\right), \mathcal{O}\left(\frac{\mathrm{rank}(\mathbf{A}) + \log(1/\delta)}{\epsilon^2}\right)\right), \tag{29}$$

which ensures that $\boldsymbol{\sigma}_{\mathbf{S}}^*$ satisfies the guarantee (27) with high probability. The tensor sample complexity estimate (29) yields a slightly weaker bound compared to the optimal result regarding standard unstructured sketching [126], given by $c \geq \mathcal{O}(\frac{\mathrm{rank}(\mathbf{A})+\log(1/\delta)}{\epsilon^2})$, which corresponds to the second term in (29).

Due to the inherent source-detector structure in data arising from PDE inverse problems, tensor sketching is specially designed to efficiently process such structured data. Although great progress has been made in studying tensor sketching methods, several key open questions remain. A central question is to establish the optimal sample complexity bound, improving upon the sub-optimal result in (29)). Another crucial question concerns the practical scenario in which measurements are sparse or missing. In such cases, it may be beneficial to design the sketching operation in a correspondingly sparse format, allowing the use of even fewer measurements while still preserving the essential information of the inference problem.

4.2 Matrix-Matrix product

Matrix multiplication is one of the most fundamental tasks in numerical linear algebra. Given two matrices $\mathbf{A} \in \mathbb{R}^{m\times C}$ and $\mathbf{B} \in \mathbb{R}^{C\times n}$, we want to compute their product

$$\mathbf{P} = \mathbf{AB} \in \mathbb{R}^{m\times n}\,,$$

for which the formula is $\mathbf{P}_{ij} = \mathbf{A}_{i,:}\mathbf{B}_{:,j} = \sum_k \mathbf{A}_{ik}\mathbf{B}_{kj}$ for every entry (i,j). Randomization enters through an alternative but equivalent representation of the product:

$$\mathbf{P} = \sum_{k=1}^{C} \mathbf{A}_{:,k}\mathbf{B}_{k,:} = \mathbb{E}_k\left(\alpha_k \mathbf{P}_k\right)\,, \tag{30}$$

where $\mathbf{P}_k = \mathbf{A}_{:,k}\mathbf{B}_{k,:}$ is a rank-1 matrix formed from the outer product of k-th column of $\mathbf{A}$ and the k-th row of $\mathbf{B}$. The expectation is taken over a random index k, sampled according to a distribution $1/\alpha_k$ chosen by the user.

This rewriting expresses $\mathbf{P}$ as the expectation of random rank-1 matrices. Consequently, it suggests a Monte Carlo (MC) approximation:

$$\mathbf{P} = \mathbb{E}_k(\alpha_k \mathbf{P}_k) \;\approx\; \frac{1}{c}\sum_{i=1}^{c} \alpha_{k_i} \mathbf{P}_{k_i}, \tag{31}$$

where $\{k_i\}_{i=1}^{c}$ are i.i.d. samples drawn from the distribution $\{1/\alpha_j\}_j$. Equivalently, the approximation can be cast in linear algebra form:

$$\mathbf{P} = \mathbf{AB} \;\approx\; \mathbf{AS}^\top \mathbf{SB}, \tag{32}$$

where $\mathbf{S} \in \mathbb{R}^{c\times C}$ is a sampling matrix with rows

$$\mathbf{S}_{i,:} = \sqrt{\alpha_{k_i}}\, \mathbf{e}_{k_i}^\top, \qquad k_i \sim \{1/\alpha_j\}_j \,. \tag{33}$$

See the sampling illustration in Fig. 5 below.

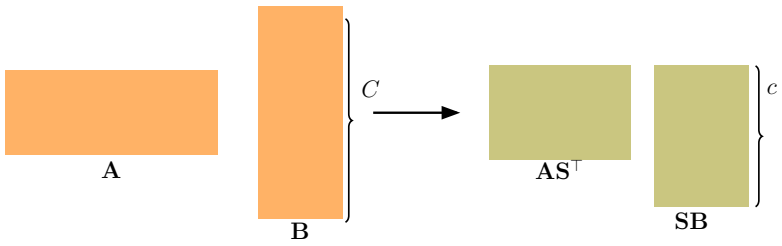


Figure 5: Matrix-matrix product sampling demonstration.

The quality and efficiency of this randomized solver depend on the choice of probabilities $1/\alpha_k$ and the number of samples c. Drineas et al.[60] showed that the optimal configuration is achieved when

$$\alpha_k \propto |\mathbf{A}_{:,k}||\mathbf{B}_{k,:}|\,, \qquad \text{and} \qquad c = O\left(\frac{1+\log\delta^{-1}}{\epsilon^2}\right) \tag{34}$$

which guarantees that the Monte Carlo approximation(31) satisfies

$$|\mathbf{P} - \mathbf{AS}^\top \mathbf{SB}|_F < \epsilon \qquad \text{with probability at least } 1-\delta. \tag{35}$$

This randomized viewpoint is mathematically elegant: the matrix product is reconstructed as the average of randomly sampled rank-1 contributions. In practice the method has computational benefits only when the required number of samples c is much smaller than C.

4.2.1 Matrix-Matrix product for PDE based experimental design

PDE-based experimental design provides a natural example of an effective application of the randomization approach for computing matrix-matrix product, through the computation of the Fisher Information Matrix (FIM) (14). Recall from (14) that, in regions where $\boldsymbol{\sigma}$ is close to the ground truth, the Hessian matrix of the loss functional is well approximated by the

FIM. A distinctive feature of this matrix is that it decomposes as a sum of many rank-1 contributions:

$$\mathrm{Hess}_{\boldsymbol{\sigma}} L \approx \mathbf{H} = \mathbf{A}^\top \mathbf{A} = \sum_{\theta} \mathbf{A}_{\theta,:}^\top \mathbf{A}_{\theta,:} \,. \tag{36}$$

Here, each row $\mathbf{A}_{\theta,:}$ corresponds to the Fréchet derivative $\frac{d}{d\boldsymbol{\sigma}}\mathrm{pde}(\boldsymbol{\sigma}, \theta)$ associated with a particular design θ. In the semi-infinite setting with $\boldsymbol{\sigma} \in \Sigma_n$ and $\theta \in \Theta$, this makes $\mathbf{A}$ a matrix with n columns and infinitely many rows, a scenario that fits seamlessly into the randomized matrix–matrix product framework of (31).

Following the sampling formulation (32)-(33), one selects a subset of designs $\Theta_c \subset \Theta$ to construct a down-sampled Hessian:

$$\mathbf{H}_S = \mathbf{A}^\top \mathbf{S}^\top \mathbf{S} \mathbf{A} \,,$$

where $\mathbf{S}$ encodes the randomized sampling. According to (34), the optimal strategy is to sample in proportion to $|\mathbf{A}_{\theta,:}|^2$, i.e., the contribution of each design to the FIM.

A corollary of the analysis in [62], as applied in [78], guarantees that with high probability:

$$\lambda(\mathbf{H}_S) > \lambda(\mathbf{H}) - \epsilon \quad \text{with probability} \geq 1 - \delta \,,$$

where λ denotes the smallest eigenvalue. This bound ensures the positivity of the sketched Hessian and thereby the well-posedness of the subsampled optimization problem.

This randomized FIM approximation has been applied, for instance, to the linearized stationary Schrödinger potential reconstruction problem (**E3.**). The results demonstrate that the optimal sensor locations are highly sensitive to the underlying ground truth $\boldsymbol{\sigma}_*$, as illustrated in Figure 6.

This sensitivity is not a desired feature for reconstruction, and can potentially be overcome by a natural extension of this perspective to a sequential selection of experimental designs that involves feedback from previous experiments, as described in Section 5.1, which is left for future work. Moreover, a combination with a matrix completion approach in Section 4.6 promises enhanced reconstruction efficiency and may help bridge the gap between the semi-infinite and finite data regime.

4.3 Column subset selection and conditioning guarantees

Identifying a well-conditioned subset of columns from a large, wide matrix is another fundamental question in numerical linear algebra: Given a "dictionary" matrix $\mathbf{A} \in \mathbb{R}^{c \times N}$ with $c \leq N$, how can one select columns from $\mathbf{A}$ to form a submatrix $\mathbf{A_S}$ so that $\mathbf{A_S}$ is well-conditioned? This problem is central to many applications in data science, including sparse signal and image reconstruction.

Mathematically, the selection process can be described using a sampling operator $\mathbf{S}$, and the conditioning is evaluated via the Gram matrix

$$\mathbf{H_S} \;=\; \mathbf{S}^\top \mathbf{A}^\top \mathbf{A} \mathbf{S} \;=\; \mathbf{A_S}^\top \mathbf{A_S} \;\in \mathbb{R}^{n \times n} \,, \quad \text{Where } \mathbf{A_S} := \mathbf{AS}. \tag{37}$$

The goal is to design $\mathbf{S}$ so that $\mathrm{cond}(\mathbf{H_S}) = \frac{\lambda_{\max}(\mathbf{H_S})}{\lambda_{\min}(\mathbf{H_S})}$ is small, where $\lambda_{\max,\min}$ denote the extreme eigenvalues. See illustration in Fig. 7.

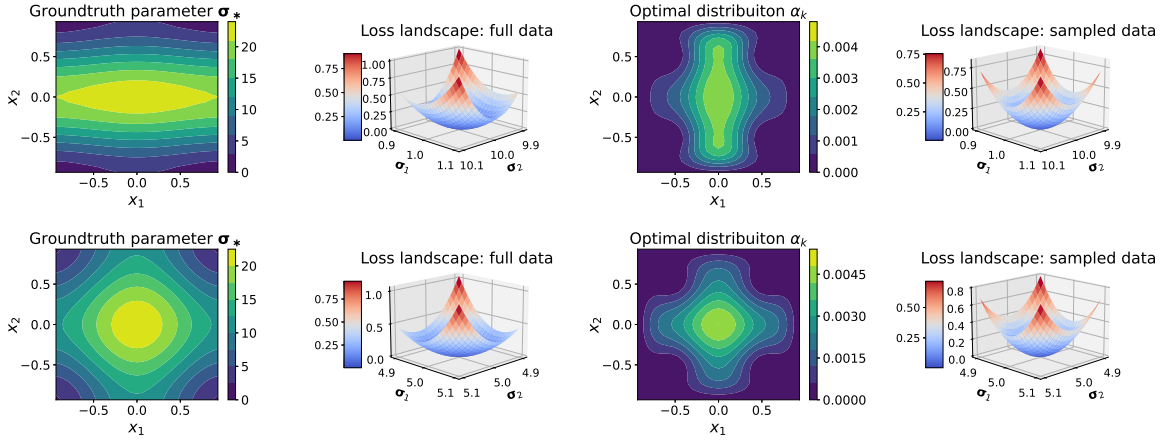


Figure 6: Loss landscapes $L(\boldsymbol{\sigma})$ for two different ground truth parameters $\boldsymbol{\sigma}_*(x)$ (left column) under full data (middle left), and sampled data (right column) according to the optimal sampling distribution (middle right).

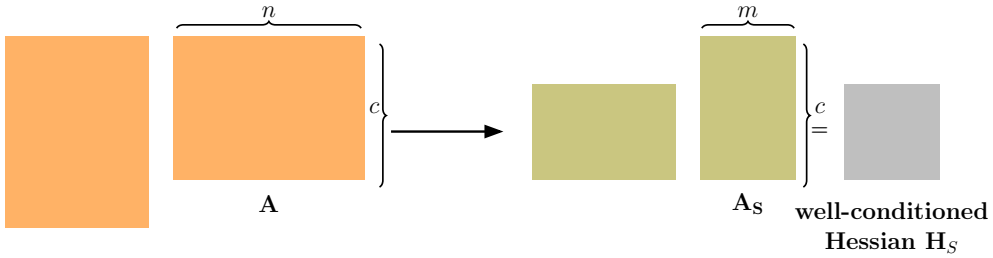


Figure 7: Hessian (column) sampling: $\mathbf{H} = \mathbf{A}^\top \mathbf{A} \in \mathbb{R}^{N\times N} \to \mathbf{H_S} = \mathbf{A_S}^\top \mathbf{A_S} \in \mathbb{R}^{n\times n}$.

Randomization provides powerful tools for this problem. A key result of [153] gives probabilistic guarantees on conditioning when columns are sampled uniformly. Assuming the columns of $\mathbf{A}$ are normalized, the number of columns n that can be selected while retaining good conditioning satisfies

$$n \;\le\; \min\left(o\left(\frac{1}{\mu^2 \log c}\right),\; o\left(\frac{N}{\|\mathbf{H}\|_2}\right)\right), \tag{38}$$

with high probability, where

$$\mathbb{P}\left(\mathrm{cond}(\mathbf{H_S}) \;\le\; \frac{1+\tau(n)}{1-\tau(n)}\right) \;\ge\; 1 - \frac{1}{c}. \tag{39}$$

Here, $\mu = \max_{i\neq j} |\langle \mathbf{A}_{:,i}, \mathbf{A}_{:,j}\rangle|$ is the coherence between columns, and $\tau(n)$ is a distortion parameter scaling as $\tau(n) = \mathcal{O}\left(\sqrt{\mu^2 n \log c} \;+\; n\frac{\|\mathbf{H}\|_2}{N}\right)$.

The coherence μ plays a decisive role: If μ is large, columns are highly overlapping, and only a relatively small number of them can be selected while preserving conditioning. Therefore, smaller μ enlarges the allowable range of n, consistent with the intuition that independence between columns enables more extensive sampling without loss of stability. The second term $\mathcal{O}\left(\frac{N}{\|\mathbf{H}\|_2}\right)$ can be misleading. Since $N = \mathrm{Tr}(\mathbf{H}) = \sum_i \lambda_i(\mathbf{H}) \;\le\; c\,\|\mathbf{H}\|_2$,

the ratio $\frac{N}{\|\mathbf{H}\|_2}$ is controlled by the matrix rank c and may not provide an informative bound in practice. In summary, the sampling dimension n (38) scales linearly with the matrix rank c and is influenced by the factor of $1/\mu^2$.

4.3.1 Column subset selection for PDE based experimental design

As discussed in (14) and Section 4.2, the Hessian of the objective near the ground truth is dominated by the FIM. Since the FIM has an outer-product structure closely resembling (37), it is natural to expect that column subset selection strategies can provide useful guidance for data selection in PDE-based inverse problems.

From (36), the Hessian is expressed as an outer product of $\mathbf{A}$, with rows $\mathbf{A}_{\theta,:} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\sigma}}\mathrm{pde}(\theta, \boldsymbol{\sigma}) \in \mathbb{R}^N$. Applying the column selection framework in (37) corresponds to fixing a set of θ samples and restricting $\boldsymbol{\sigma}$. This is the opposite setting from Section 4.2.1: Here, the experiments are already collected, and the task is to identify subsets of parameters $\boldsymbol{\sigma}$ that can be reconstructed in a well-conditioned manner. This strategy corresponds to *random sampling in parameter space*.

Direct application of prior results (38)–(39) is obstructed by the fact that the columns of $\mathbf{A}$ are not normalized in our application. Nevertheless, by adapting proof techniques from [134, 153], the work [94] generalizes the guarantees to PDE settings and establishes that

$$n \leq \min\left(o\left(\frac{\mathrm{Tr}(\mathbf{H})}{\|\mathbf{H}\|_2}\right), o\left(\frac{1}{\mu^2\ \log c}\frac{\mathrm{Tr}(\mathbf{H})^2}{\|\mathbf{H}\|_F^2}\right)\right), \tag{40}$$

with the probabilistic conditioning bound

$$\mathbb{P}\left(\mathrm{cond}(\mathbf{H_S}) \leq \frac{L+\tau(n)}{\ell-\tau(n)}\right) \geq 1 - \frac{1}{c}\,. \tag{41}$$

Here the distortion quantity is given by $\tau(n) = e^{\frac{1}{4}}\left(2n\frac{\|\mathbf{H}\|_2}{\mathrm{Tr}(\mathbf{H})} + 12\mu\sqrt{n\log c}\,\frac{\|\mathbf{H}\|_F}{\mathrm{Tr}(\mathbf{H})}\right)$, and the normalization constants account for extreme column norms:

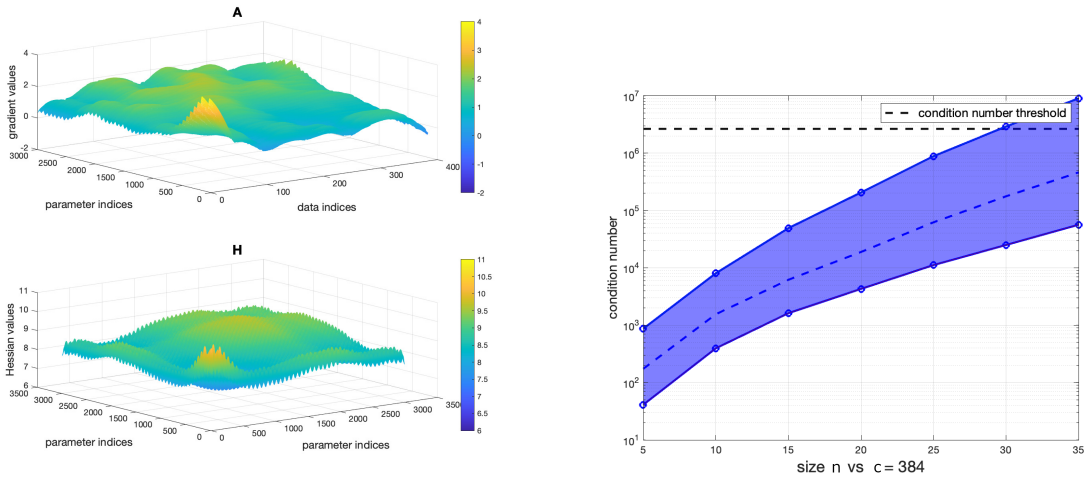
$$\ell := \frac{N}{\mathrm{Tr}(\mathbf{H})}\min_{i=1\ldots N}\|\mathbf{A}_{:,i}\|_2^2, \quad L := \frac{N}{\mathrm{Tr}(\mathbf{H})}\max_{i=1\ldots N}\|\mathbf{A}_{:,i}\|_2^2\,.$$

The coherence parameter must also be adjusted to account for non-uniform column magnitudes: $\mu := \frac{N}{\|\mathbf{H}\|_F}\max_{i\neq j=1\ldots N}|\langle\mathbf{A}_{:,i}, \mathbf{A}_{:,j}\rangle|$.

The second term in (40) highlights the role of coherence: if two parameter directions $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ yield nearly parallel sensitivity vectors $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\sigma}}\mathrm{pde}(:, \boldsymbol{\sigma}_1)$ and $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\sigma}}\mathrm{pde}(:, \boldsymbol{\sigma}_2)$, the effective number of reliably reconstructible parameters decreases.

Numerical experiments validate these results. In Fig.8, $\mathbf{A}$ is generated from the linearized EIT model **E2.** with $c = 384$. As n increases, the condition number of $\mathbf{H_S}$ grows consistently with the concentration inequality (41). For reference, the baseline threshold $\frac{L}{\ell}$ is also shown (dashed line).

To conclude, understanding unique reconstruction and well-posedness from limited data information is a key challenge in PDE inverse problems. Random sampling in the parameter space can be used to build a probabilistic framework and discern the number of parameters identifiable in a well-conditioned manner. Open research avenues for future study include: 1.



(a) Visualization of $\mathbf{A}$ and $\mathbf{H}$ from the EIT model.

(b) Sample complexity n (horizontal) vs. condition number (vertical). The shaded region shows the 20%–80% quantile range over 10,000 simulations.

Figure 8: Column subset selection in PDE-based inverse problems.

Incorporating greedy algorithm and importance sampling strategy to maximize the number of parameters recovered; 2. Extending the analysis to the infinite-dimensional PDE setting with full parameter space $\Sigma = \lim_{N\to\infty} \Sigma_N$, motivated by the concentration results (40)-(41) being independent from the ambient parameter dimension N.

4.4 Randomized SVD

In singular value decomposition (SVD), another classical task in numerical linear algebra, we start with a rectangular matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ of rank $d \leq \min(m, n)$, and find matrices $\mathbf{U}$, $\mathbf{S}$, $\mathbf{V}$ such that

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \sum_i s_i \mathbf{U}_{:,i} \left(\mathbf{V}_{:,i}\right)^\top ,$$

where $\mathbf{S} = \mathrm{diag}\{s_i\}$ collects singular values that is ordered in a descending manner, while $\mathbf{U} \in \mathbb{R}^{m\times d}$ and $\mathbf{V} \in \mathbb{R}^{n\times d}$ have orthonormal columns, which are the left and right singular vectors or $\mathbf{C}$, respectively, Classically, computation of the SVD starts with a bidiagonalization via Householder transformations, which transforms $\mathbf{C}$ into a matrix that has two diagonal components. The total cost is $O(\max(m, n) \cdot \min(m, n)^2)$.

Randomized SVD (RSVD) returns estimates to $\mathbf{U}$, $\mathbf{V}$ and $\{s_i\}$ with high accuracy and probability. As thoroughly analyzed in the foundational works [114, 109, 76], it is a much more cost-efficient algorithm than classical SVD for large, dense matrices, and is highly parallelizable and memory-efficient. The total cost is about $O(mnd)$, and in comparison with classical solvers, the saving is most pronounced when the large matrix $\mathbf{C}$ is known to be of low rank, that is, $d \ll \min\{m, n\}$. As a consequence, RSVD finds applications in domains that traditionally rely on Principal Component Analysis (PCA) [131], such as bioinformatics [57] and image processing [164].

RSVD has been successfully deployed for experimental design purposes for PDE-based

inverse problems as well, especially through the optimal design angle. Recalling (16), the task at hand it to find the weights $W = \mathrm{diag}\{w_\theta\}$ so that the induced variance matrix $\hat{\Gamma}_W$ is small, either in terms of largest eigenvalue (Φ_E), or the summation of all eigenvalues (Φ_A) or the multiplication (Φ_D), all of which require computation of eigenvalues. Considering the size of $\hat{\Gamma}_W$, RSVD that returns all eigenvalues efficiently becomes very useful. This perspective was adopted in [139, 1, 79, 13, 160, 14].

4.5 Compressed sensing

Compressed sensing (CS) is a signal processing technique that leverages randomized numerical linear algebra ideas, and it has promising applications in PDE-based inverse problems. As its name suggests, CS acquires information about a signal in a compressed manner. When the object to be reconstructed lies in a high-dimensional ambient space $\mathbb{R}^N$ but is known to be sparse (that is, supported on only a small fraction of the coordinates), the sparsity can be exploited to drastically reduce the number of measurements required. In particular, if the sensing mechanism satisfies certain structural properties, the number of "sensors" scales with the intrinsic information content n rather than with the ambient dimension N.

A standard reconstruction method takes the form of an optimization problem:

$$\min \|\mathbf{x}\|_1 \,, \quad \text{such that} \quad \mathbf{P}_\Omega \mathbf{U} \cdot \mathbf{x} = \mathbf{P}_\Omega \mathbf{y} \,.$$

Here $\mathbf{U}$ is an orthonormal sensing matrix collecting all possible measurements, and $\mathbf{P}_\Omega$ is a projection operator selecting entries according to the mask Ω of size $|\Omega| = c$. Because $\mathbf{x}$ is assumed sparse, the ℓ_1 norm promotes sparsity in the recovered solution. Although the system is highly underdetermined, posing $c \ll N$ constraints for N-entry of reconstruction, exact recovery is still possible when $\mathbf{U}$ is incoherent with the coordinate basis. Classical results show that if $c \gtrsim n \log N$, then exact recovery of $\mathbf{x}$ is achievable with high probability [40, 59, 41].

These foundational results, originally developed in finite-dimensional linear algebra, extend naturally to infinite-dimensional settings that arise in PDE-based inverse problems. In this context, the to-be-reconstructed object is a function $f(x)$ rather than a finite-dimensional vector, and the sensing matrix becomes a continuous operator $\mathcal{U}$. Nevertheless, if $f(x)$ is sparse in a basis system $\mathcal{D} = \{\phi_i\}$ and if $\mathcal{D}$ and $\mathcal{U}$ are mutually incoherent, compressed sensing theory ensures stable recovery even from finitely many measurements [7, 3].

4.6 Matrix completion

Matrix completion seeks to complete a dense matrix $\mathbf{A} \in \mathbb{R}^{C \times C}$, given certain entries $\mathbf{A}_{ij}$ for $(i, j) \in \Omega$, a sparse mask. The task is applicable to recommendation system [101] and genomics [38]. The problem itself does not necessarily belong to RNLA, but some solvers use random matrices and the associated techniques. In its generic form, there is no hope to reconstruct a matrix with a few entries, but if certain assumptions are satisfied, the reconstruction problem can be solved. Notably, when the matrix is known a-priori to be of low rank and its column and row space known to be decoherent, the following optimization problem returns an optimizer that is known to recover $\mathbf{A}$ with high probability[42, 43, 129].

$$\begin{array}{ll} \text{minimize} & \|\mathbf{X}\|_* \\ \text{subject to} & \mathbf{X}_{ij} = \mathbf{A}_{ij}, \quad \forall\ (i,j) \in \Omega \,. \end{array} \tag{42}$$

where Ω is a randomly sampled mask with each entry following the Bernoulli distribution at a low sampling rate, and $\|\cdot\|_*$ is the nuclear norm (the sum of all singular values). Intuitively, the objective function is the L_1 norm version of singular values, in analogy to compressed sensing, where in both context it serves as a proxy for the L_0 norm. The minimization of L_0 norm means one looks for a reconstruction that has smallest rank and agrees with the given data. A crude conclusion suggests a sampling complexity of $c \approx Cr$ where r is the approximate rank.

It is easy to imagine these completion solvers being useful for qualitative experimental design associated with inverse problems. An inverse problem is to use ItO data pairs to reconstruct unknown parameters. Restricted by sampling and experiment budget, one does not have the full set of input-to-output data pairs, but rather a subset of entries. Can one use this subset to reconstruct the full data set, which is then feed into the true reconstruction?

The approach is taken in [37] for the EIT problem, **E2.** from section 2.5, for which the ItO map is the Dirichlet-to-Neumann data. It is important to note that a well-posed inverse problem typically corresponds to a full-rank data matrix that stores the full ItO map, which prevents immediate application of matrix completion method (42). Indeed, in [37], recognizing the off-diagonally low rank feature of the DtN map, a dyadic decomposition on the diagonal blocks can be used to peel off the off-diagonal blocks as the partition level increases [110], yielding a block structure as displayed in illustration Fig. 9. Then matrix

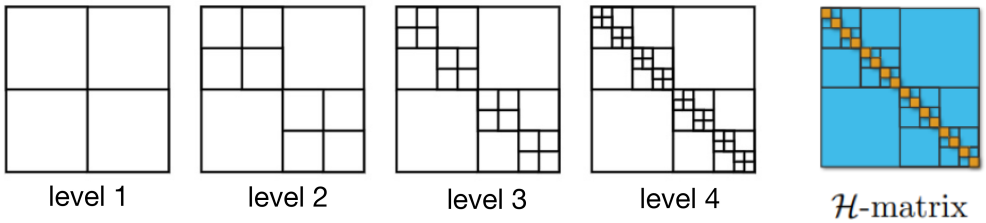


Figure 9: Hierarchical matrix block structure on varying partition levels and rank structure of the $\mathcal{H}$-matrix: the diagonal full-rank blocks (orange) vs. blue low-rank off diagonal blocks.

completion (42) is applied to each off-diagonal block. The recovered full dataset is then deployed in PDE-constrained optimization for the execution of the inverse problem.

5 Outlook

This review has concentrated on the role of randomized numerical linear algebra (RNLA) in data selection for PDE-based inverse problems. While powerful and increasingly influential, our focus is necessarily narrow. At its core, RNLA is a linear framework whereas data selection is intrinsically nonlinear. As such, relying exclusively on RNLA overlooks essential aspects of the problem. Even within the linear regime, alternative toolboxes from modern machine learning and applied mathematics offer complementary strategies that may also prove valuable. In what follows, we highlight two such directions and discuss how they may broaden the scope of data selection methodologies.

5.1 Nonlinear data selection and active learning

The primary limitation of RNLA-based approaches is that they operate strictly within linear spaces, while experimental design is fundamentally nonlinear. As discussed in Section 3.3, the information value of an experiment depends strongly on the underlying medium $\boldsymbol{\sigma}^*$. This dependence is evident in Figure 6, where the weights assigned to measurements vary dramatically with different $\boldsymbol{\sigma}^*$. A natural measure of data value is the sensitivity of the reconstruction to the data, often encoded in the Jacobian. In the RNLA framework, the Jacobian is treated as a fixed matrix, making most strategies effectively *passive*: The algorithm is prescribed once, and the user simply waits for the output. In the nonlinear regime, however, the Jacobian itself evolves as reconstruction improves, shifting the sensitivity landscape and altering the relative value of new data.

This observation motivates *interactive* or *adaptive* data selection: Begin with an initial dataset, perform a reconstruction, and then collect additional data informed by the current state of knowledge, repeating the process as needed.

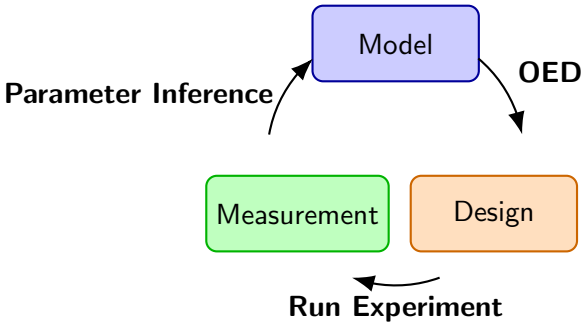


Figure 10: Greedy sequential optimal experimental design.

Several concrete strategies capture this philosophy. One is to pose the problem as a bilevel optimization: The inner layer solves the reconstruction problem for a given dataset, while the outer layer selects the dataset itself [137, 93]. Another approach is greedy data acquisition [53, 161, 89, 160], in which data are acquired sequentially, each step exploiting information from previous ones, as described by Figure 10. More sophisticated approaches additionally allow for coordination of subsequent experiments and pose the problem in the reinforcement learning regime [82, 141].

Crucially, this challenge is not unique to PDE-based inverse problems. Similar difficulties arise across machine learning tasks constrained by costly data acquisition, where *active learning* has emerged as a central paradigm. We expect that cross-fertilization between active learning and nonlinear experimental design for PDEs will play an increasingly important role in the coming years [143].

5.2 Operating in infinite-dimensional settings

Even within the linear regime, RNLA methods ultimately reduce large *but finite* linear systems via sampling or sketching. The design space is therefore still finite, no matter how large it is taken. This discreteness is still incompatible with PDE-based inverse problems, where the design space is continuously indexed and thus infinite-dimensional.

The importance of continuous indexing is illustrated in Fig. 11, which simulates radiative transfer **E4.**. For each experiment (each column), only a handful of detector locations record nontrivial light intensity. The relevant information thus arises in a measure-zero subset of the design space. Any fixed discretization, no matter how fine, risks missing these measure-zero structures and therefore producing a poor design space for data selection.

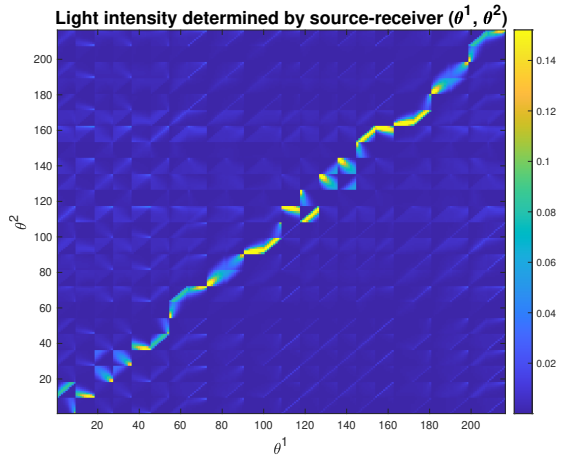


Figure 11: Simulation of the radiative transfer equation characterizing laser beam propagation. The column index θ_1 denotes the laser location and velocity direction. For each θ_1, the rows correspond to boundary measurements at detector locations θ_2. Only a few values of θ_2 produce nontrivial readings, forming a measure-zero set in the full design space.

Analogous phenomena arise in the Darcy flow example **E2.** and the Lorenz-63 problem **E1.**.

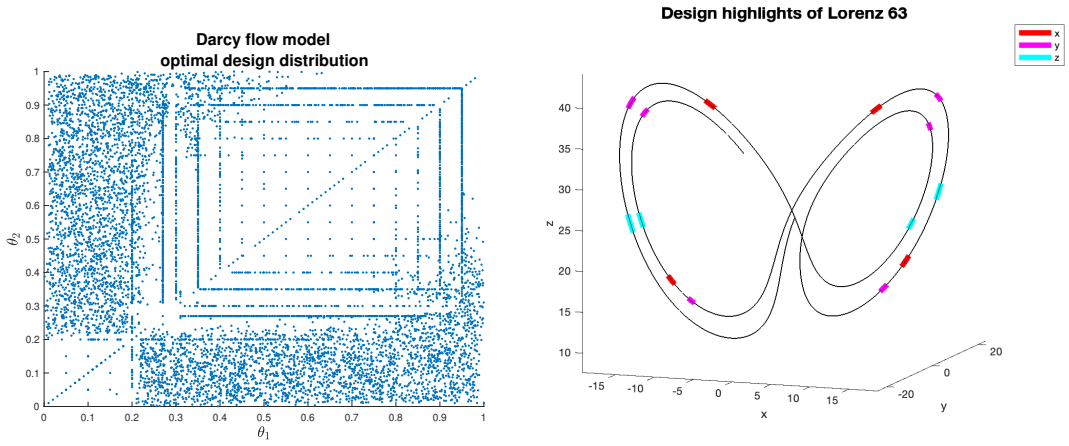


Figure 12: Left panel: Optimal distribution over the continuous design space for the Darcy flow problem, showing extreme concentration of the data importance. Right panel: Important observation times in colored slots for parameter inference in the Lorenz-63 model. Different colors indicate different measured components (x, y, z).

In such cases, methods that *directly* operate in infinite-dimensional spaces become essential. Traditionally, such an approach was regarded as impossible, since most optimization

solvers operate in $\mathbb{R}^d$ and exhibit complexity growing at least algebraically with d, rendering them infeasible in the infinite-dimensional limit. Recent advances have changed this landscape. Tools from optimal transport and gradient flows [15, 68, 125] have clarified the structure of optimization over probability measure spaces, which are intrinsically infinite-dimensional. Techniques such as the mean-field limit and propagation of chaos [44, 70] further enable practical computation by translating infinite-dimensional PDEs into coupled ODE systems. These ideas, widely used in machine learning [51, 123, 122, 73, 49], have only recently been adapted to experimental design problems [91].

One such approach replaces discrete weights $w(\theta)$ in the classical optimal design problem (16) with a probability measure $\rho(\theta)$:

$$\begin{aligned} &\text{A-optimal} : \Phi_A[\rho] = \operatorname{tr}\left(\left(\mathbf{A}^\top \mathbf{A}[\rho]\right)^{-1}\right), \\ &\text{D-optimal} : \Phi_D[\rho] = \det\left(\mathbf{A}^\top \mathbf{A}[\rho]\right), \end{aligned} \tag{43}$$

where

$$\mathbf{A}^\top \mathbf{A}[\rho] = \int_\Omega \mathbf{A}(\theta, :)^\top \mathbf{A}(\theta, :) \, \mathrm{d}\rho(\theta).$$

Optimization is then conducted directly over the probability measure space, bypassing the limitations of finite discretization.

Taken together, these directions highlight a broader landscape of experimental design beyond RNLA. In addition to embracing nonlinearity through adaptive and interactive strategies that tightly couple experimental design with reconstruction, and advancing toward infinite-dimensional formulations via tools from optimal transport, gradient flows, and mean-field theory, other approaches are also expected to play an important role. Vice versa, experimental design methods have successfully been customized to several machine learning tasks [102, 18, 54], prompting towards a more efficient use of data and training capacities. We anticipate that progress along these lines—at the interface of applied mathematics, machine learning, and computational science—will significantly broaden the scope and amplify the impact of experimental design methodologies in the coming decade.

References

[1] Christian Aarset. Global optimality conditions for sensor placement, with extensions to binary low-rank a-optimal designs. *Inverse Problems*, 41(6):065013, 2025.

[2] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003.

[3] Ben Adcock and Anders C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, 16, 2016.

[4] Andy Adler, Romina Gaburro, and William Lionheart. *Electrical Impedance Tomography*. Springer New York, New York, NY, 2020.

[5] Nir Ailon and Bernard Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009.

[6] Giovanni S Alberti and Matteo Santacesaria. Calderón's inverse problem with a finite number of measurements. In *Forum of mathematics, sigma*, volume 7, page e35. Cambridge University Press, 2019.

[7] Giovanni S. Alberti and Matteo Santacesaria. Infinite dimensional compressed sensing from anisotropic measurements and applications to inverse problems in PDE. *Applied and Computational Harmonic Analysis*, 50:105–146, 2021.

[8] Giovanni S. Alberti and Matteo Santacesaria. Infinite-dimensional inverse problems with finite measurements. *Archive for Rational Mechanics and Analysis*, 243, 2022.

[9] Giovanni Alessandrini. Stable determination of conductivity by boundary measurements. *Applicable Analysis*, 27(1-3):153–172, 1988.

[10] Giovanni Alessandrini, Maarten V de Hoop, Romina Gaburro, and Eva Sincich. Lipschitz stability for the electrostatic inverse boundary value problem with piecewise linear conductivities. *Journal de mathématiques pures et appliquées*, 107(5):638–664, 2017.

[11] Giovanni Alessandrini, Maarten V. de Hoop, Romina Gaburro, and Eva Sincich. Lipschitz stability for a piecewise linear Schroedinger potential from local cauchy data. *Asymptotic Analysis*, 108(3):115–149, 2018.

[12] Giovanni Alessandrini and Sergio Vessella. Lipschitz stability for the inverse conductivity problem. *Advances in Applied Mathematics*, 35(2):207–241, 2005.

[13] Alen Alexanderian. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review. *Inverse Problems*, 37(4):043001, 2021.

[14] Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized l_0-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

[15] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[16] Habib Ammari, Josselin Garnier, Wenjia Jing, Hyeonbae Kang, Mikyoung Lim, and Han Wang. *Mathematical and Statistical Methods for Multistatic Imaging*, volume 2098. Springer, 01 2013.

[17] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):123010, dec 2009.

[18] Jordan Ash, Surbhi Goel, Akshay Krishnamurthy, and Sham Kakade. Gone fishing: Neural active learning with fisher embeddings. *Advances in Neural Information Processing Systems*, 34:8927–8939, 2021.

[19] Kari Astala and Lassi Päivärinta. Calderón's inverse conductivity problem in plane. *Annals of mathematics*, 163:265–299, 01 2006.

[20] Corwin L Atwood. Sequences converging to d-optimal designs of experiments. *The Annals of Statistics*, pages 342–352, 1973.

[21] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. Random Fourier features for kernel ridge regression: Approximation bounds and statistical guarantees. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 253–262. PMLR, 06–11 Aug 2017.

[22] Valeria Bacchelli and Sergio Vessella. Lipschitz stability for a stationary 2D inverse problem with unknown polygonal boundary. *Inverse Problems*, 22(5):1627, aug 2006.

[23] Guillaume Bal. Inverse transport theory and applications. *Inverse Problems*, 25(5):053001, mar 2009.

[24] Gang Bao, Peijun Li, Junshan Lin, and Faouzi Triki. Inverse scattering problems with multi-frequencies. *Inverse Problems*, 31(9):093001, 2015.

[25] Mourad Bellassoued and Masahiro Yamamoto. *Carleman Estimates and Applications to Inverse Problems for Hyperbolic Systems*. Springer Monographs in Mathematics. Springer Tokyo, 2017.

[26] Hamid Bellout, Avner Friedman, and Victor Isakov. Stability for an inverse problem in potential theory. *Transactions of the American Mathematical Society*, 332(1):271–296, 1992.

[27] David A. Benaron, Susan R. Hintz, Arno Villringer, David Boas, Andreas Kleinschmidt, Jens Frahm, Christina Hirth, Hellmuth Obrig, John C. van Houten, Eben L. Kermit, Wai-Fung Cheong, and David K. Stevenson. Noninvasive functional imaging of human brain using light. *Journal of Cerebral Blood Flow & Metabolism*, 20(3):469–477, 2000.

[28] Elena Beretta, Maarten V De Hoop, Florian Faucher, and Otmar Scherzer. Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates. *SIAM Journal on Mathematical Analysis*, 48(6):3962–3983, 2016.

[29] Elena Beretta, Maarten V. de Hoop, and Lingyun Qiu. Lipschitz stability of an inverse boundary value problem for a Schroedinger-type equation. *SIAM Journal on Mathematical Analysis*, 45(2):679–699, 2013.

[30] Elena Beretta and Elisa Francini. Lipschitz stability for the electrical impedance tomography problem: the complex case. *Communications in Partial Differential Equations*, 36(10):1723–1749, 2011.

[31] Karianne J Bergen, Paul A Johnson, Maarten V de Hoop, and Gregory C Beroza. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433):eaau0323, 2019.

[32] Avraham Y. Bluestone, Gassan Abdoulaev, Christoph H. Schmitz, Randall L. Barbour, and Andreas H. Hielscher. Three-dimensional optical tomography of hemodynamics in the human head. *Opt. Express*, 9(6):272–286, Sep 2001.

[33] Liliana Borcea. Electrical impedance tomography. *Inverse problems*, 18(6):R99, 2002.

[34] Liliana Borcea. Imaging in random media. In *Handbook of Mathematical Methods in Imaging*, pages 1–54. Springer, 2014.

[35] Liliana Borcea, George Papanicolaou, Chrysoula Tsogka, and James Berryman. Imaging and time reversal in random media. *Inverse Problems*, 18(5):1247, 2002.

[36] Laurent Bourgeois. A remark on Lipschitz stability for inverse problems. *Comptes Rendus Mathematique*, 351(5):187–190, 2013.

[37] Tan Bui-Thanh, Qin Li, and Leonardo Zepeda-Núñez. Bridging and improving theoretical and computational electrical impedance tomography via data completion. *SIAM Journal on Scientific Computing*, 44(3):B668–B693, 2022.

[38] Tianxi Cai, T Tony Cai, and Anru Zhang. Structured matrix completion with applications to genomic data integration. *Journal of the American Statistical Association*, 111(514):621–633, 2016.

[39] Alberto Calderón. On inverse boundary value problem. *Computational & Applied Mathematics*, 25, 01 2006.

[40] E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[41] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969, apr 2007.

[42] Emmanuel J Candes and Benjamin Recht. Exact low-rank matrix completion via convex optimization. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 806–812. IEEE, 2008.

[43] Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE transactions on information theory*, 56(5):2053–2080, 2010.

[44] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: A review of models, methods and applications. II. Applications. *Kinetic and Related Models*, 15(6):1017–1173, 2022.

[45] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theoretical Computer Science*, 312(1):3–15, 2004.

[46] Gaoming Chen, Fadil Santosa, and Aseel Titi. Determination of a small elliptical anomaly in electrical impedance tomography using minimal measurements. *Inverse Problems*, 40(11):115006, oct 2024.

[47] Ke Chen and Ruhui Jin. Tensor-structured sketching for constrained least squares. *SIAM Journal on Matrix Analysis and Applications*, 42(4):1703–1731, 2021.

[48] Ke Chen, Qin Li, Kit Newton, and Stephen J Wright. Structured random sketching for PDE inverse problems. *SIAM Journal on Matrix Analysis and Applications*, 41(4):1742–1770, 2020.

[49] Shi Chen, Zhengjiang Lin, Yury Polyanskiy, and Philippe Rigollet. Quantitative clustering in mean-field transformer models. *arXiv preprint arXiv:2504.14697*, 2025.

[50] Margaret Cheney, David Isaacson, and Jonathan C. Newell. Electrical impedance tomography. *SIAM Review*, 41(1):85–101, 1999.

[51] Lénaïc Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 3040–3050, 2018.

[52] Mourad Choulli and Plamen Stefanov. An inverse boundary value problem for the stationary transport equation. *Osaka Journal of Mathematics*, 36(1):87 – 104, 1999.

[53] Emily Clark, Travis Askham, Steven L. Brunton, and J. Nathan Kutz. Greedy sensor placement with cost constraints. *IEEE Sensors Journal*, 19(7):2642–2656, 2019.

[54] David A Cohn, Zoubin Ghahramani, and Michael I Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

[55] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Applied Mathematical Sciences. Springer New York, 2012.

[56] David Colton and Lassi Päivärinta. The uniqueness of a solution to an inverse scattering problem for electromagnetic waves. *Archive for rational mechanics and analysis*, 119(1):59–70, 1992.

[57] Gregory Darnell, Stoyan Georgiev, Sayan Mukherjee, and Barbara E Engelhardt. Adaptive randomized dimension reduction on massive data. *Journal of Machine Learning Research*, 18(140):1–30, 2017.

[58] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[59] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[60] Petros Drineas, Ravi Kannan, and Michael W Mahoney. Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication. *SIAM Journal on Computing*, 36(1):132–157, 2006.

[61] Petros Drineas, Malik Magdon-Ismail, Michael W Mahoney, and David P Woodruff. Fast approximation of matrix coherence and statistical leverage. *The Journal of Machine Learning Research*, 13(1):3475–3506, 2012.

[62] Petros Drineas, Michael W Mahoney, Shan Muthukrishnan, and Tamás Sarlós. Faster least squares approximation. *Numerische mathematik*, 117(2):219–249, 2011.

[63] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands, 2000.

[64] Björn Engquist and Lexing Ying. Sweeping preconditioner for the Helmholtz equation: hierarchical matrix representation. *Communications on pure and applied mathematics*, 64(5):697–735, 2011.

[65] Valerii Vadimovich Fedorov. *Theory of Optimal Experiments*. Cellular Neurobiology. Academic Press, 1972.

[66] Joel Feldman, Mikko Salo, and Gunther Uhlmann. *The Calderón Problem: An Introduction*. Graduate Studies in Mathematics. AMS, 2025.

[67] Marco Ferrari and Valentina Quaresima. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921–935, 2012.

[68] A. Figalli and F. Glaudo. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. EMS textbooks in mathematics. European Mathematical Society, 2021.

[69] Jean-Pierre Fouque, Josselin Garnier, and Knut Solna. *Wave propagation and time reversal in randomly layered media*, volume 56. Springer Science & Business Media, 2007.

[70] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3):707–738, 2015.

[71] Alberto Franceschetti and Alex Zunger. The inverse band-structure problem of finding an atomic configuration with given electronic properties. *Nature*, 402, 1999.

[72] Romina Gaburro and Eva Sincich. Lipschitz stability for the inverse conductivity problem for a conformal class of anisotropic conductivities. *Inverse Problems*, 31(1):015008, 2015.

[73] Borjan Geshkovski, Cyril Letrouit, Yury Polyanskiy, and Philippe Rigollet. The emergence of clusters in self-attention dynamics. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[74] Eldad Haber, Uri M Ascher, and Doug Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse problems*, 16(5):1263, 2000.

[75] Peter Hähner and Thorsten Hohage. New stability estimates for the inverse acoustic inhomogeneous medium problem and applications. *SIAM Journal on Mathematical Analysis*, 33(3):670–685, 2001.

[76] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[77] Bastian Harrach. Uniqueness and Lipschitz stability in electrical impedance tomography with finitely many electrodes. *Inverse problems*, 35(2):024005, 2019.

[78] Kathrin Hellmuth, Christian Klingenberg, and Qin Li. Sensitivity-preserving of Fisher information matrix through random data down-sampling for experimental design. *arXiv preprint arXiv:2409.15906*, 2024.

[79] Elizabeth Herman, Alen Alexanderian, and Arvind K Saibaba. Randomization and reweighted l_1-minimization for a-optimal design of linear inverse problems. *SIAM Journal on Scientific Computing*, 42(3):A1714–A1740, 2020.

[80] Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*, volume 23. Springer Science & Business Media, 2008.

[81] Xun Huan, Jayanth Jagalur, and Youssef Marzouk. Optimal experimental design: Formulations and computations. *Acta Numerica*, 33:715–840, 2024.

[82] Xun Huan and Youssef M Marzouk. Sequential Bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

[83] Debin Huang and Rongwei Guo. Identifying parameter by identical synchronization between different systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 14(1):152–159, 03 2004.

[84] Victor Isakov. Uniqueness and stability in multi-dimensional inverse problems. *Inverse problems*, 9(6):579, 1993.

[85] Victor Isakov. *Inverse problems for partial differential equations*. Springer, 2006.

[86] Kazufumi Ito and Bangti Jin. *Inverse problems: Tikhonov theory and algorithms*, volume 22. World Scientific, 2014.

[87] Kazufumi Ito, Bangti Jin, and Jun Zou. A direct sampling method to an inverse medium scattering problem. *Inverse Problems*, 28(2):025003, 2012.

[88] Mark A Iwen, Deanna Needell, Elizaveta Rebrova, and Ali Zare. Lower memory oblivious (tensor) subspace embeddings with fewer random bits: modewise methods for least squares. *SIAM Journal on Matrix Analysis and Applications*, 42(1):376–416, 2021.

[89] Jayanth Jagalur-Mohan and Youssef Marzouk. Batch greedy maximization of non-submodular functions: Guarantees and applications to experimental design. *Journal of Machine Learning Research*, 22(252):1–62, 2021.

[90] Bangti Jin and Peter Maass. Sparsity regularization for parameter identification problems. *Inverse Problems*, 28(12):123001, 2012.

[91] Ruhui Jin, Martin Guerra, Qin Li, and Stephen Wright. Optimal experimental design for linear models via gradient flow. *arXiv preprint arXiv:2401.07806*, 2024.

[92] Ruhui Jin, Tamara G Kolda, and Rachel Ward. Faster Johnson–Lindenstrauss transforms via Kronecker products. *Information and Inference: A Journal of the IMA*, 10(4):1533–1562, 10 2020.

[93] Ruhui Jin, Qin Li, Stephen O. Mussmann, and Stephen J. Wright. Continuous nonlinear adaptive experimental design via gradient flow. *arXiv preprint arXiv:2411.14332*, 2024.

[94] Ruhui Jin, Qin Li, Anjali Nair, and Sam Stechmann. Unique reconstruction for discretized inverse problems: a random sketching approach via subsampling. *Inverse Problems*, 41(4):045010, 2025.

[95] William Johnson and Joram Lindenstrauss. Extensions of Lipschitz maps into a hilbert space. *Contemporary Mathematics*, 26:189–206, 01 1984.

[96] Jari P Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Springer, 2005.

[97] Andreas Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*, volume 120. Springer, 01 2011.

[98] M V Klibanov. Inverse problems and Carleman estimates. *Inverse Problems*, 8(4):575, aug 1992.

[99] Michael V. Klibanov and Sergey E. Pamyatnykh. Global uniqueness for a coefficient inverse problem for the non-stationary transport equation via Carleman estimate. *Journal of Mathematical Analysis and Applications*, 343(1):352–365, 2008.

[100] Robert Kohn and Michael Vogelius. Determining conductivity by boundary measurements. *Communications on pure and applied mathematics*, 37(3):289–298, 1984.

[101] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[102] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(2), 2008.

[103] H. Kunze and K. Heidler. The collage coding method and its application to an inverse problem for the Lorenz system. *Applied Mathematics and Computation*, 186(1):124–129, 2007.

[104] Edward W. Larsen. Solution of multidimensional inverse transport problems. *Journal of Mathematical Physics*, 25(1):131–135, 01 1984.

[105] Jason D Lee, Ruoqi Shen, Zhao Song, Mengdi Wang, and zheng Yu. Generalized leverage score sampling for neural networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10775–10787. Curran Associates, Inc., 2020.

[106] Matthew Li, Laurent Demanet, and Leonardo Zepeda-Núñez. Wide-band butterfly network: Stable and efficient inversion via multi-frequency neural networks. *Multiscale Modeling & Simulation*, 20(4):1191–1227, 2022.

[107] Peijun Li, Jian Zhai, and Yue Zhao. Lipschitz stability for an inverse source scattering problem at a fixed frequency. *Inverse Problems*, 37(2):025003, 2021.

[108] Qin Li and Weiran Sun. Applications of kinetic tools to inverse transport problems. *Inverse Problems*, 36(3):035011, feb 2020.

[109] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences*, 104(51):20167–20172, 2007.

[110] Lin Lin, Jianfeng Lu, and Lexing Ying. Fast construction of hierarchical matrix representation from matrix–vector multiplication. *Journal of Computational Physics*, 230(10):4071–4087, 2011.

[111] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130 – 141, 1963.

[112] Michael W Mahoney. Lecture notes on randomized linear algebra. *arXiv preprint arXiv:1608.04481*, 2016.

[113] Michael W Mahoney et al. Randomized algorithms for matrices and data. *Foundations and Trends® in Machine Learning*, 3(2):123–224, 2011.

[114] Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30(1):47–68, 2011.

[115] Per-Gunnar Martinsson and Joel A Tropp. Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 29:403–572, 2020.

[116] Charlie Mattschas, Marius Puplauskis, Chris Toebes, Violetta Sharoglazova, and Jan Klaers. Inverse solving the Schrodinger equation for precision alignment of a microcavity. *Phys. Rev. Res.*, 7:013296, Mar 2025.

[117] Michela Meister, Tamas Sarlos, and David Woodruff. Tight dimensionality reduction for sketching low degree polynomial kernels. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[118] Amanda F Mejia, Mary Beth Nebel, Ani Eloyan, Brian Caffo, and Martin A Lindquist. PCA leverage: outlier detection for high-dimensional functional magnetic resonance imaging data. *Biostatistics*, 18(3):521–536, 2017.

[119] Toby J Mitchell and FL Miller Jr. Use of design repair to construct designs for special linear models. *Math. Div. Ann. Progr. Rept.(ORNL-4661)*, 13, 1970.

[120] Judith R. Mourant, James P. Freyer, Andreas H. Hielscher, Angelia A. Eick, Dan Shen, and Tamara M. Johnson. Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. *Applied optics*, 37(16):3586–3593, Jun 1998.

[121] Adrian I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Annals of Mathematics*, 143(1):71–96, 1996.

[122] Naoki Nishikawa, Taiji Suzuki, Atsushi Nitanda, and Denny Wu. Two-layer neural network on infinite-dimensional data: global optimization guarantee in the mean-field regime. *Journal of Statistical Mechanics: Theory and Experiment*, 2023(11):114007, nov 2023.

[123] Atsushi Nitanda, Anzelle Lee, Damian Tan Xing Kai, Mizuki Sakaguchi, and Taiji Suzuki. Propagation of chaos for mean-field Langevin dynamics and its application to model ensemble. In *Forty-second International Conference on Machine Learning*, 2025.

[124] R.G. Novikov. Multidimensional inverse spectral problem for the equation—$\delta\ \psi + (v(x)|eu(x))\psi = 0$. *Functional Analysis and Its Applications*, 22(4):263–272, 1988.

[125] Gabriel Peyre and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[126] Mert Pilanci and Martin J Wainwright. Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory*, 61(9):5096–5115, 2015.

[127] Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[128] Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern Bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

[129] Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(12), 2011.

[130] Kui Ren. Recent developments in numerical techniques for transport-based medical imaging methods. *Commun. Comput. Phys*, 8(1):1–50, 2010.

[131] Vladimir Rokhlin, Arthur Szlam, and Mark Tygert. A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1100–1124, 2010.

[132] Luca Rondi. A remark on a paper by Alessandrini and Vessella. *Advances in Applied Mathematics*, 36(1):67–69, 2006.

[133] Amir Rosenthal, Vasilis Ntziachristos, and Daniel Razansky. Acoustic inversion in optoacoustic tomography: A review. *Current Medical Imaging*, 9(4):318–336, 2013.

[134] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(8):1025–1045, 2008.

[135] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[136] Angkana Ruland and Eva Sincich. Lipschitz stability for the finite dimensional fractional Calderón problem with finite cauchy data. *Inverse Problems and Imaging*, 13(5):1023–1044, 2019.

[137] Lars Ruthotto, Julianne Chung, and Matthias Chung. Optimal experimental design for inverse problems with state constraints. *SIAM Journal on Scientific Computing*, 40(4):B1080–B1100, 2018.

[138] Elizabeth G. Ryan, Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

[139] Arvind K Saibaba, Alen Alexanderian, and Ilse CF Ipsen. Randomized matrix-free trace and log-determinant estimators. *Numerische Mathematik*, 137(2):353–395, 2017.

[140] Benjamin Sanchez-Lengeling and Alán Aspuru-Guzik. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, 2018.

[141] Fadil Santosa and Loren Anderson. Bayesian sequential optimal experimental design for linear regression with reinforcement learning. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 641–646. IEEE, 2022.

[142] Michael Shatalov, Samuel Surulere, Lilies Phadime, and Phumezile Kama. Parameter identification of Lorenz system with incomplete information: Case of one unknown function. *International Journal of Engineering Research in Africa*, 55:82–102, 08 2021.

[143] Atsushi Shimizu, Xiaoou Cheng, Christopher Musco, and Jonathan Weare. Improved active learning via dependent leverage score sampling. *arXiv preprint arXiv:2310.04966*, 2023.

[144] Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 563–568, 2008.

[145] Plamen Stefanov. Stability of the inverse problem in potential scattering at fixed energy. In *Annales de l'institut Fourier*, volume 40, pages 867–884, 1990.

[146] Plamen Stefanov and Gunther Uhlmann. Linearizing non-linear inverse problems and an application to inverse backscattering. *Journal of Functional Analysis*, 256(9):2842–2866, 2009.

[147] Andrew Stuart. Inverse problems: a Bayesian perspective. *Acta numerica*, 19:451–559, 2010.

[148] Hongyu Sun, Yen Sun, Rami Nammour, Christian Rivera, Paul Williamson, and Laurent Demanet. Learning with real data without real labels: a strategy for extrapolated full-waveform inversion with field data. *Geophysical Journal International*, 235(2):1761–1777, 08 2023.

[149] John Sylvester and Gunther Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Annals of Mathematics*, 125(1):153–169, 1987.

[150] Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.

[151] Matthias Taus, Leonardo Zepeda-Núñez, Russell J Hewett, and Laurent Demanet. L-sweeps: A scalable, parallel preconditioner for the high-frequency Helmholtz equation. *Journal of Computational Physics*, 420:109706, 2020.

[152] A. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 5:1035–1038, 1963.

[153] Joel A Tropp. On the conditioning of random subdictionaries. *Applied and Computational Harmonic Analysis*, 25(1):1–24, 2008.

[154] Joel A Tropp. Improved analysis of the subsampled randomized hadamard transform. *Advances in Adaptive Data Analysis*, 3(01n02):115–126, 2011.

[155] Gunther Uhlmann. 30 years of Calderón's problem. *Séminaire Laurent Schwartz — EDP et applications*, pages 1–25, 01 2012.

[156] Éric Walter and Luc Pronzato. Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica*, 26(2):195–213, 1990.

[157] Eric Walter and Luc Pronzato. On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and computers in simulation*, 42(2-3):125–134, 1996.

[158] Jenn-Nan Wang. Stability estimates of an inverse problem for the stationary transport equation. *Annales de l'I.H.P. Physique théorique*, 70(5):473–495, 1999.

[159] David P Woodruff et al. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science*, 10(1–2):1–157, 2014.

[160] Keyi Wu, Peng Chen, and Omar Ghattas. A fast and scalable computational framework for large-scale high-dimensional Bayesian optimal experimental design. *SIAM/ASA Journal on Uncertainty Quantification*, 11(1):235–261, 2023.

[161] Henry P Wynn. The sequential generation of d-optimum experimental designs. *The Annals of Mathematical Statistics*, 41(5):1655–1664, 1970.

[162] Henry P Wynn. Results in the theory and construction of D-optimum experimental designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2):133–147, 1972.

[163] Alwyn Young. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598, 11 2018.

[164] Jiani Zhang, Jennifer Erway, Xiaofei Hu, Qiang Zhang, and Robert Plemmons. Randomized SVD methods in hyperspectral imaging. *Journal of Electrical and Computer Engineering*, 2012(1):409357, 2012.