# On the randomised stability constant for inverse problems

Giovanni S. Alberti, Yves Capdeboscq, Yannick Privat

▶ **To cite this version:**

Giovanni S. Alberti, Yves Capdeboscq, Yannick Privat. On the randomised stability constant for inverse problems. Mathematics in Engineering, 2020, 2 (2), pp.264-286. 10.3934/mine.2020013 . hal-02077786v2

**HAL Id: hal-02077786**
**https://hal.science/hal-02077786v2**

Submitted on 5 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ON THE RANDOMISED STABILITY CONSTANT FOR INVERSE PROBLEMS

GIOVANNI S. ALBERTI, YVES CAPDEBOSCQ, AND YANNICK PRIVAT

Abstract. In this paper we introduce the randomised stability constant for abstract inverse problems, as a generalisation of the randomised observability constant, which was studied in the context of observability inequalities for the linear wave equation. We study the main properties of the randomised stability constant and discuss the implications for the practical inversion, which are not straightforward.

## 1. Introduction

Inverse problems are the key to all experimental setups where the physical quantity of interest is not directly observable and must be recovered from indirect measurements. They appear in many different contexts including medical imaging, nondestructive testing, seismic imaging or signal processing. In mathematical terms, an inverse problem consists in the inversion of a linear or nonlinear map

$$T\colon X \to Y, \qquad x \mapsto T(x),$$

which models how the quantity of interest $x$ belonging to a space $X$ is related to the measurements $y = T(x)$ in the space $Y$. The reader is referred to the many books on inverse problems for a comprehensive exposition (see, e.g., [27, 3, 47, 30, 44, 2, 24, 4]).

Inverse problems can be ill-posed: the map $T$ may not be injective (i.e., two different $x_1$ and $x_2$ may correspond to the same measurement $T(x_1) = T(x_2)$) or, when injective, $T^{-1}\colon \operatorname{ran} T \subseteq Y \to X$ may not be continuous (i.e., two different and not close $x_1$ and $x_2$ may correspond to almost identical measurements $T(x_1) \approx T(x_2)$). Various strategies have been introduced to tackle the issue of inversion in this setting, Tykhonov regularisation being the most famous method [17, 29].

Our purpose is to investigate the role of randomisation in the resolution of inverse problems. By randomisation, we refer to the use of random measurements, or to the case of random unknowns. We do not mean the well-established statistical approaches in inverse problems, as in the Bayesian framework, where probability is used to assess the reliability of the reconstruction.

Even with this distinction, the wording "randomisation" may refer to many different concepts in the framework of inverse problems. A huge literature is devoted to the issue of randomisation in the measuring process: the unknown $x$ in $X$ is fixed and deterministic, and we choose the measurements randomly according to some suitable distribution. For example, compressed sensing [19] and passive imaging

2010 *Mathematics Subject Classification.* 65J22, 35R30.

*Key words and phrases.* Inverse problems, observability constant, compressed sensing, passive imaging, regularisation, randomisation, deep learning, electrical impedance tomography.

with ambient noise [20] belong to this class. Another very popular idea consists in randomising the unknown: in this scenario, we try to recover *most* unknowns $x$ in $X$, according to some distribution. For example, this is typically the situation when deep learning [21] is applied to inverse problems. We briefly review these instances in Appendix A.

This article deals with a more recent approach, in the framework of observability or control theory: the so-called randomised observability constants [38, 39, 40]. We argue that, in contrast with the initial motivation for its introduction, the randomised observability constant is not necessarily indicative of the likelihood for randomised unknowns to be observable.

In the next section, we recall the notion of randomised observability constant and comment on its use in optimal design problems. In section 3, we reformulate the randomised observability constant in an abstract setting as randomised stability constant. In section 4 we show that when the classical (deterministic) stability constant is null, a positive randomised stability constant need not imply, as one could hope, that the inverse problem can be solved for most unknowns. In the course of our study, we make several observations on the properties of the randomised stability constant. Section 5 contains several concluding remarks and discusses possible future directions.

## 2. Randomising initial data of PDEs

In this section we briefly review the randomised observability constant introduced in [38, 39, 40] and its main properties. This was motivated by the original idea of randomisation by Paley and Zygmund, which we now briefly discuss.

### 2.1. On randomisation processes.

In order to understand the principle of randomisation that will be at the core of this paper, it is useful to recall the historical result by Paley and Zygmund on Fourier series. Let $(c_n)_{n\in\mathbb{Z}}$ be an element of $\ell^2(\mathbb{C})$ and $f$ be the Fourier series given by

$$f : \mathbb{T} \ni \theta \mapsto \sum_{n\in\mathbb{Z}} c_n e^{in\theta},$$

where $\mathbb{T}$ denotes the torus $\mathbb{R}/(2\pi)$. According to the so-called Parseval identity, the function $f$ belongs to $L^2(\mathbb{T})$; furthermore, the coefficients $c_n$ can be chosen in such a way that $f$ does not belong to any $L^q(\mathbb{T})$ for $q > 2$. Some of the results obtained by Paley and Zygmund (see [36, 35, 37]) address the regularity of the Fourier series $f$. They show that if one changes randomly and independently the signs of the Fourier coefficients $c_n$, then the resulting random Fourier series belongs almost surely to any $L^q(\mathbb{T})$ for $q > 2$. More precisely, introducing a sequence $(\beta_n^\nu)_{n\in\mathbb{Z}}$ of independent Bernoulli random variables on a probability space $(A, \mathcal{A}, \mathbb{P})$ such that

$$\mathbb{P}(\beta_n^\nu = \pm 1) = \frac{1}{2},$$

then, the Fourier series $f^\nu$ given by

$$f^\nu : \mathbb{T} \ni \theta \mapsto \sum_{n\in\mathbb{Z}} \beta_n^\nu c_n e^{in\theta}$$

belongs almost surely to $L^q(\mathbb{T})$ for all $q < +\infty$.

In [13], the effect of the randomisation on the initial data of solutions to dispersive equations has been investigated. In that case, initial data and solutions of

the PDE considered are expanded in a Hilbert basis of $L^2(\Omega)$ made of eigenfunctions $(\phi_j)_{j\geq 1}$ of the Laplace operator. The randomisation procedure consists hence in multiplying all the terms of the series decomposition by well-chosen independent random variables. In particular, regarding for instance the homogeneous wave equation with Dirichlet boundary conditions, one can show that for all initial data $(y^0, y^1) \in H_0^1(\Omega) \times L^2(\Omega)$, the Bernoulli randomisation keeps the $H_0^1 \times L^2$ norm constant. It is observed that many other choices of randomisation are possible. For instance a positive effect of the randomisation can be observed by considering independent centred Gaussian random variables with variance 1 (Gaussian randomisation). Indeed, it allows to generate a dense subset of the space of initial data $H_0^1(\Omega) \times L^2(\Omega)$ through the mapping

$$R_{(y^0,y^1)}\colon A \to H_0^1(\Omega) \times L^2(\Omega), \qquad \nu \mapsto (y^0_\nu, y^1_\nu),$$

where $(y^0_\nu, y^1_\nu)$ denotes the pair of randomised initial data, provided that all the coefficients in the series expansion of $(y^0, y^1)$ are nonzero. Several other properties of these randomisation procedures are also established in [13].

2.2. **Randomised observability constant .** We now review how randomisation appeared in the framework of inverse problems involving an observability inequality. The property of observability of a system is related to the following issue: how to recover the solutions of a PDE from the knowledge of partial measurements of the solutions. In what follows, we will concentrate on wave models, having in particular photoacoustic/thermoacoustic tomography imaging in mind.

Let $T > 0$ and $\Omega \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain with outer unit normal $\nu$. We consider the homogeneous wave equation with Dirichlet boundary conditions

$$\begin{cases} \partial_{tt} y(t,x) - \Delta y(t,x) = 0 & (t,x) \in [0,T] \times \Omega, \\ y(t,x) = 0 & (t,x) \in [0,T] \times \partial\Omega. \end{cases} \tag{1}$$

It is well known that, for all $(y^0, y^1) \in H_0^1(\Omega) \times L^2(\Omega)$, there exists a unique solution $y \in C^0([0,T], H_0^1(\Omega)) \cap C^1((0,T), L^2(\Omega))$ of (1) such that $y(0,x) = y^0(x)$ and $\partial_t y(0,x) = y^1(x)$ for almost every $x \in \Omega$. Let $\Gamma$ be a measurable subset of $\partial\Omega$, representing the domain occupied by some sensors, which take some measurements over a time horizon $[0,T]$.

The inverse problem under consideration reads as follows.

> **Inverse problem:** reconstruct the initial condition $(y^0, y^1)$ from the knowledge of the partial boundary measurements
>
> $$\mathbb{1}_\Gamma(x) \frac{\partial y}{\partial \nu}(t,x), \qquad (t,x) \in [0,T] \times \partial\Omega.$$

To solve this problem, we introduce the so-called *observability constant*: $C_T(\Gamma)$ is defined as the largest non-negative constant $C$ such that

$$C \|(y(0,\cdot), \partial_t y(0,\cdot))\|^2_{H_0^1(\Omega)\times L^2(\Omega)} \leq \int_0^T \int_\Gamma \left| \frac{\partial y}{\partial \nu}(t,x) \right|^2 d\mathcal{H}^{d-1}\, dt, \tag{2}$$

for any solution $y$ of (1), where $H_0^1(\Omega)$ is equipped with the norm $\|u\|_{H_0^1(\Omega)} = \|\nabla u\|_{L^2(\Omega)}$. Then, the aforementioned inverse problem is well-posed if and only if $C_T(\Gamma) > 0$. In such a case, we will say that observability holds true in time $T$. Moreover, observability holds true within the class of $\mathcal{C}^\infty$ domains $\Omega$ if $(\Gamma, T)$ satisfies the *Geometric Control Condition (GCC)* (see [8]), and this sufficient condition is almost necessary.

Let us express the observability constant more explicitly. Fix an orthonormal basis (ONB) $(\phi_j)_{j\geq 1}$ of $L^2(\Omega)$ consisting of (real-valued) eigenfunctions of the Dirichlet-Laplacian operator on $\Omega$, associated with the negative eigenvalues $(-\lambda_j^2)_{j\geq 1}$. Then, any solution $y$ of (1) can be expanded as

$$y(t,x) = \sum_{j=1}^{+\infty} y_j(t)\phi_j(x) = \frac{1}{\sqrt{2}} \sum_{j=1}^{+\infty} \left( \frac{a_j}{\lambda_j} e^{i\lambda_j t} + \frac{b_j}{\lambda_j} e^{-i\lambda_j t} \right) \phi_j(x), \tag{3}$$

where the coefficients $a_j$ and $b_j$ account for initial data. More precisely, we consider the ONB of $H_0^1(\Omega) \times L^2(\Omega)$ given by $\{\psi_j^+, \psi_j^- : j \geq 1\}$, where

$$\psi_j^+ = \frac{1}{\sqrt{2}}\left(\frac{\phi_j}{\lambda_j}, i\phi_j\right), \qquad \psi_j^- = \frac{1}{\sqrt{2}}\left(\frac{\phi_j}{\lambda_j}, -i\phi_j\right). \tag{4}$$

Expanding now the initial data with respect to this basis we can write $(y^0, y^1) = \sum_{j=1}^{+\infty} a_j\psi_j^+ + b_j\psi_j^-$, namely,

$$y^0 = \sum_{j=1}^{+\infty} \frac{a_j + b_j}{\sqrt{2}\,\lambda_j}\,\phi_j, \qquad y^1 = \sum_{j=1}^{+\infty} i\frac{a_j - b_j}{\sqrt{2}}\,\phi_j.$$

The corresponding solution to (1) is given by (3). In addition, Parseval's identity yields $\|(y^0, y^1)\|^2_{H_0^1(\Omega)\times L^2(\Omega)} = \sum_{j=1}^{+\infty} |a_j|^2 + |b_j|^2$.

Then, the constant $C_T(\Gamma)$ rewrites

$$C_T(\Gamma) = \inf_{\substack{(a_j),(b_j)\in\ell^2(\mathbb{C}) \\ \sum_{j=1}^{+\infty}(|a_j|^2+|b_j|^2)=1}} \int_0^T \int_\Gamma \left| \frac{\partial y}{\partial \nu}(t,x) \right|^2 d\mathcal{H}^{d-1}\,dt,$$

where $y(t,x)$ is given by (3).

The constant $C_T(\Gamma)$ is deterministic and takes into account any $(a_j), (b_j) \in \ell^2(\mathbb{C})$, including the worst possible cases. Interpreting $C_T(\Gamma)$ as a quantitative measure of the well-posed character of the aforementioned inverse problem, one could expect that such worst cases do not occur too often; thus it would appear desirable to consider a notion of observation in average.

Motivated by the findings of Paley and Zygmund (see §2.1) and its recent use in another context [11, 13], making a random selection of all possible initial data for the wave equation (1) consists in replacing $C_T(\Gamma)$ with the so-called *randomised observability constant* defined by

$$C_{T,\mathrm{rand}}(\Gamma) = \inf_{\substack{(a_j),(b_j)\in\ell^2(\mathbb{C}) \\ \sum_{j=1}^{+\infty}(|a_j|^2+|b_j|^2)=1}} \mathbb{E}\left( \int_0^T \int_\Gamma \left| \frac{\partial y^\nu}{\partial \nu}(t,x) \right|^2 d\mathcal{H}^{d-1}\,dt \right), \tag{5}$$

where

$$y^\nu(t,x) = \frac{1}{\sqrt{2}} \sum_{j=1}^{+\infty} \left( \frac{\beta_{1,j}^\nu a_j}{\lambda_j} e^{i\lambda_j t} + \frac{\beta_{2,j}^\nu b_j}{\lambda_j} e^{-i\lambda_j t} \right) \phi_j(x) \tag{6}$$

and $(\beta_{1,j}^\nu)_{j\in\mathbb{N}}$ and $(\beta_{2,j}^\nu)_{j\in\mathbb{N}}$ are two sequences of independent random variables of Bernoulli or Gaussian type, on a probability space $(A, \mathcal{A}, \mathbb{P})$ with mean 0 and variance 1. Here, $\mathbb{E}$ is the expectation in the probability space, and runs over all possible events $\nu$. In other words, we are randomising the Fourier coefficients $\{a_j, b_j\}_{j\geq 1}$ of the initial data $(y^0, y^1)$ with respect to the basis $\{\psi_j^\pm\}_{j\geq 1}$.

The randomised observability constant was introduced in [38, 39, 40, 41, 42]. It can be expressed in terms of deterministic quantities (see [40, Theorem 2.2]).

**Proposition 1.** *Let $\Gamma \subset \partial\Omega$ be measurable. We have*

$$C_{T,\mathrm{rand}}(\Gamma) = \frac{T}{2} \inf_{j\in\mathbb{N}} \frac{1}{\lambda_j^2} \int_\Gamma \left(\frac{\partial \phi_j}{\partial \nu}(x)\right)^2 d\mathcal{H}^{d-1}. \tag{7}$$

*Proof.* In view of Lemma 1 (see below), we have that

$$C_{T,\mathrm{rand}}(\Gamma) = \inf\{\|\mathbb{1}_\Gamma \partial_\nu y_j^+\|^2_{L^2([0,T]\times\partial\Omega)}, \|\mathbb{1}_\Gamma \partial_\nu y_j^-\|^2_{L^2([0,T]\times\partial\Omega)} : j \geq 1\},$$

where $y_j^\pm$ is the solution to (1) with initial condition $\psi_j^\pm$. Thus

$$y_j^\pm(t,x) = \frac{1}{\sqrt{2}\,\lambda_j} e^{\pm i\lambda_j t}\phi_j(x),$$

and in turn

$$\|\mathbb{1}_\Gamma \partial_\nu y_j^\pm\|^2_{L^2([0,T]\times\partial\Omega)} = \frac{1}{2\lambda_j^2} \int_{[0,T]\times\Gamma} |e^{\pm i\lambda_j t}\partial_\nu \phi_j(x)|^2 \, dtdx,$$

which leads to our thesis. □

We have $C_{T,\mathrm{rand}}(\Gamma) \geq C_T(\Gamma)$ (see Proposition 4 below). It has been noted in [41] that the observability inequality defining $C_{T,\mathrm{rand}}(\Gamma)$ is associated to a deterministic control problem for the wave equation (1), where the control has a particular form but acts in the whole domain $\Omega$.

Regarding $C_{T,\mathrm{rand}}(\Gamma)$, we refer to [42, Section 4] for a discussion on the positivity of this constant. The authors show that if $\Omega$ is either a hypercube or a disk, then $C_{T,\mathrm{rand}}(\Gamma) > 0$ for every relatively non-empty open subset $\Gamma$ of $\partial\Omega$. In particular, in some cases $C_T(\Gamma) = 0$ while $C_{T,\mathrm{rand}}(\Gamma) > 0$. This raised hopes that, even if recovering *all* unknowns is an unstable process, recovering *most* unknowns could be feasible, since apparently most unknowns are observable. This heuristic argument, mentioned amongst the motivations for the study of the optimisation of the randomised observability constant, was not investigated further in the aforementioned papers. This matter will be studied in the following sections, dedicated more generally on the possible use of such a constant for investigating the well-posed character of general inverse problems.

*Applications to optimal design problems.* A larger observability constant $C_T(\Gamma)$ in (2) leads to a smaller Lipschitz norm bound of the inverse map. Therefore $C_T(\Gamma)$ can be used as the quantity to maximise when searching for optimal sensors' positions. However, this turns out to be somewhat impractical. When implementing a reconstruction process, one has to carry out in general a very large number of measures; likewise, when implementing a control procedure, the control strategy is expected to be efficient in general, but possibly not for all cases. Thus, one aims at exhibiting an observation domain designed to be the best possible in average, that is, over a large number of experiments. Adopting this point of view, it appeared relevant to consider an *average over random initial data*. In [38, 39, 40], the best observation is modelled in terms of maximising a *randomised observability constant*, which coincides with $C_{T,\mathrm{rand}}(\Gamma)$ when dealing with the boundary observation of the wave equation.

When dealing with internal observation of the wave equation on a closed manifold, it has been shown in [26] that the related observability constant reads as the

minimum of two quantities: the infimum of the randomised observability constants over every orthonormal eigenbasis and a purely geometric criterion standing for the minimal average time spent by a geodesic in the observation set.

However, one should keep in mind that a large randomised constant may not be associated with a reconstruction method (see section 4).

## 3. The randomised stability constant for abstract inverse problems

It is convenient to generalise the construction of the previous section to an abstract setting. In what follows, none of the arguments require the precise form of the forward operator related to the wave equation, as they rely solely on the structure of the randomised constant.

For the remainder of this paper, we let $X$ and $Y$ be separable infinite-dimensional Hilbert Spaces, and $P\colon X \to Y$ be an injective bounded linear operator.

If $P^{-1}\colon \operatorname{ran} P \to X$ is a bounded operator, the inverse problem of finding $x$ from $P(x)$ can be solved in a stable way for all $x \in X$, without the need for randomisation. This can be measured quantitatively by the constant

$$C_{\mathrm{det}} = \inf_{x \in X \setminus \{0\}} \frac{\|Px\|_Y^2}{\|x\|_X^2} > 0.$$

The smaller $C_{\mathrm{det}}$ is, the more ill-conditioned the inversion becomes.

On the other hand, when $P^{-1}$ is unbounded, the situation is different, and the inverse problem is ill-posed [30, 24]. In this case, although the kernel of $P$ reduces to $\{0\}$, we have

$$C_{\mathrm{det}} = \inf_{x \in X \setminus \{0\}} \frac{\|Px\|_Y^2}{\|x\|_X^2} = 0. \tag{8}$$

Examples of such maps abound. The example that motivated our study was that introduced in section 2.2, with $X = H_0^1(\Omega) \times L^2(\Omega)$, $Y = L^2([0,T] \times \partial\Omega)$ and

$$P(y^0, y^1) = y|_{[0,T]\times\Gamma},$$

where $\Gamma \subseteq \partial\Omega$ and $y$ is the solution to (1) with initial condition $(y^0, y^1)$: if $\Gamma$ is not large enough, $P$ is still injective but $P^{-1}$ is unbounded [31]. Any injective compact linear operator satisfies (8).

Let us now introduce the *randomised stability constant*, which generalises the randomised observability constant to this general setting. We consider the class of random variables introduced in the last section. Choose an ONB $\mathrm{e} = \{\mathrm{e}_k\}_{k\in\mathbb{N}}$ of $X$ and write $x = \sum_{k=1}^{\infty} x_k \mathrm{e}_k \in X$. We consider random variables of the form

$$x^\nu = \sum_{k=1}^{\infty} \beta_k^\nu x_k \mathrm{e}_k,$$

where $\beta_k^\nu$ are i.i.d. complex-valued random variables on a probability space $(A, \mathcal{A}, \mathbb{P})$ with vanishing mean and variance 1, so that $\mathbb{E}(|\beta_k^\nu|^2) = 1$ for every $k$. These include the Bernoulli and Gaussian random variables considered in the previous section. It is worth observing that, in the case of Bernoulli random variables, we have $|\beta_k^\nu|^2 = 1$ for every $k$, so that $\|x^\nu\|_X = \|x\|_X$.

**Definition 1.** The *randomised stability constant* is defined as

$$C_{\mathrm{rand}}(\mathrm{e}) = \inf_{x\in X\setminus\{0\}} \mathbb{E}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2}\right).$$

As in the previous section, this constant should represent stability of the inverse problem for *most* unknowns. By definition, we have $C_{\mathrm{rand}}(\mathrm{e}) \geq C_{\mathrm{det}}$. In general, this is a strict inequality: we will provide examples in section 4. This can be heuristically seen also by the following deterministic expression for the randomised stability constant.

**Lemma 1.** *There holds*

$$C_{\mathrm{rand}}(\mathrm{e}) = \inf_k \|P(\mathrm{e}_k)\|_Y^2. \tag{9}$$

*Proof.* Since $(Y, \|\cdot\|_Y)$ is also a Hilbert space, we find

$$\begin{aligned}\|P(x^\nu)\|_Y^2 &= \left\langle \sum_{k=1}^\infty \beta_k^\nu x_k P(\mathrm{e}_k), \sum_{l=1}^\infty \beta_l^\nu x_l P(\mathrm{e}_k)\right\rangle_Y \\ &= \sum_{k=1}^\infty |\beta_k^\nu|^2 |x_k|^2 \|P(e_k)\|_Y^2 + \sum_{\substack{k,l\\ k\neq l}} \beta_k^\nu \overline{\beta_l^\nu} x_k \overline{x_l} \langle P(\mathrm{e}_k), P(\mathrm{e}_k)\rangle_Y.\end{aligned} \tag{10}$$

Since $\beta_k^\nu$ are i.i.d. with vanishing mean and such that $\mathbb{E}(|\beta_k^\nu|^2) = 1$, we obtain

$$\mathbb{E}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2}\right) = \frac{\mathbb{E}\left(\|P(x^\nu)\|_Y^2\right)}{\sum_{k=1}^\infty |x_k|^2} = \frac{\sum_{k=1}^\infty |x_k|^2 \|P(\mathrm{e}_k)\|_Y^2}{\sum_{k=1}^\infty |x_k|^2} \geq \inf_k \|P(\mathrm{e}_k)\|_Y^2, \tag{11}$$

which means $C_{\mathrm{rand}}(\mathrm{e}) \geq \inf_k \|P(\mathrm{e}_k)\|_Y^2$. Choosing $x = \mathrm{e}_k$, we obtain (9). □

## 4. Exploiting $C_{\mathrm{rand}}$ in inverse problems

The aim of this section is to discuss the impact of the randomised observability constant in inverse problems. In other words, we would like to address the following question: how can the positivity of $C_{\mathrm{rand}}$ be exploited in the solution of an inverse ill-posed problem? We will not fully address this issue, but rather provide a few positive and negative partial results.

We remind the reader that the randomisation introduced in the last section when (8) holds, was based on the point of view that the ratio in (8) is not "usually" small, and that, hopefully, in most cases inversion "should" be possible. It is worthwhile to observe that the subset of the $x \in X$ such that $\|Px\|_Y \geq c\|x\|_X$ for some fixed $c > 0$ is never generic in $X$, since $\{x \in X : \|Px\|_Y < c\|x\|_X\}$ is open and non empty by (8). This caveat in mind, we nevertheless wish to test if some evidence can be given to support our optimistic approach that in most cases, inversion should be possible.

**Proposition 2.** *For every $\epsilon > 0$ and $x \in X$, there exists $c > 0$ such that*

$$\mathbb{P}\left(\|Px^\nu\|_Y \geq c\|x\|_X\right) > 1 - \epsilon. \tag{12}$$

*Proof.* Take $x \in X$. Define the real-valued map $g(c) = \mathbb{P}\left(\|Px^\nu\|_Y \geq c\|x\|_X\right)$. Take a sequence $c_n \searrow 0$. It is enough to show that

$$\lim_{n\to+\infty} g(c_n) = 1.$$

We write

$$g(c_n) = \int_A f_n(\nu)\, d\mathbb{P}(\nu), \text{ where } f_n(\nu) = \begin{cases} 1 & \text{if } \|Px^\nu\|_Y \geq c_n \|x\|_X, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $f_n$ is monotone increasing and, since $\ker P = \{0\}$, $\lim_{n\to\infty} f_n(\nu) = 1$ for every $\nu$. Thus by the Monotone Convergence Theorem,

$$\lim_{n\to\infty} \int_A f_n(\nu)\, d\mathbb{P}(\nu) = \int_A \left(\lim_{n\to\infty} f_n(\nu)\right) d\mathbb{P}(\nu) = 1. \qquad \square$$

This thus shows that, for a fixed $x$, our intuition is vindicated: in the vast majority of cases, the inequality $\|Px^\nu\|_Y \geq c\|x\|_X$ holds true. This is true independently of $C_{\text{rand}}(\mathrm{e})$; we now investigate whether the positivity of $C_{\text{rand}}(\mathrm{e})$ may yield a stronger estimate.

4.1. **Large deviation inequalities.** The next step is to estimate the probability that, for a given $x \in X \setminus \{0\}$, the square of the ratio $\frac{\|P(x^\nu)\|_Y}{\|x\|_X}$ used in Definition 1 is close to its mean value. The large deviation result we could derive describes the deviation from an upper bound to $C_{\text{rand}}(\mathrm{e})$, namely the constant $K_{\text{rand}}(\mathrm{e})$ defined by

$$K_{\text{rand}}(\mathrm{e}) = \sup_k \|P(\mathrm{e}_k)\|_Y^2. \tag{13}$$

**Theorem 1** (large deviation estimate)**.** *Assume that $Y = L^2(\Sigma, \mu)$, where $(\Sigma, S, \mu)$ is a measure space. Let $(\beta_k^\nu)_{k\in\mathbb{N}}$ be a sequence of independent random variables of Bernoulli type, on a probability space $(A, \mathcal{A}, \mathbb{P})$ with mean 0 and variance 1. Let $x \in X \setminus \{0\}$ and $x^\nu = \sum_{k=1}^\infty \beta_k^\nu \mathrm{e}_k x_k$. Then, for every $\delta > 0$ we have*

$$\mathbb{P}\left(\|Px^\nu\|_Y \geq \delta \|x^\nu\|_X\right) \leq \exp\left(2 - \frac{1}{e}\frac{\delta}{\sqrt{K_{\text{rand}}(\mathrm{e})}}\right).$$

The proof of Theorem 1 is postponed to Appendix B. The argument follows the same lines as the one of [11, Theorem 2.1] and the general method introduced in [13].

*Remark* 1 (Application to a wave system). Considering the wave equation (1) and adopting the framework of section 2.2 leads to choose $\{\psi_j^\pm\}_{j\geq 1}$ defined by (4) as the orthonormal basis e. In that case, $X = H_0^1(\Omega) \times L^2(\Omega)$, $\Sigma = [0,T] \times \partial\Omega$, $d\mu = dt\, d\mathcal{H}^{d-1}$ and $Y = L^2(\Sigma)$. Following the discussion in section 2.2, the map $P$ is given by $P(y^0, y^1) = \mathbb{1}_\Gamma \frac{\partial y}{\partial \nu}$, where $y$ is the unique solution of (1). Further, we have

$$C_{\text{rand}}(\mathrm{e}) = \frac{T}{2} \inf_{j\in\mathbb{N}} \frac{1}{\lambda_j^2} \int_\Gamma \left(\frac{\partial \phi_j}{\partial \nu}\right)^2 d\mathcal{H}^{d-1},$$

$$K_{\text{rand}}(\mathrm{e}) = \frac{T}{2} \sup_{j\in\mathbb{N}} \frac{1}{\lambda_j^2} \int_\Gamma \left(\frac{\partial \phi_j}{\partial \nu}\right)^2 d\mathcal{H}^{d-1},$$

where the first equality is given in Proposition 1, and the second one follows by applying the same argument.

Note that according to the so-called Rellich identity[1], we have $0 < K_{\mathrm{rand}}(\mathrm{e}) \leq \frac{T}{2}\operatorname{diam}(\Omega)$, under additional mild assumptions on the domain $\Omega$.

The estimate given in Theorem 1 is on the "wrong side", since we show that the ratio related to the inversion is much bigger than $K_{\mathrm{rand}}(\mathrm{e})$ with low probability. The issue is not the boundedness of $P$, which is given a priori, but of its inverse. This would correspond to a result of the type

$$\mathbb{P}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} < C_{\mathrm{rand}}(\mathrm{e}) - \delta\right) \leq \text{small constant}, \tag{14}$$

namely, a quantification of the estimate given in Proposition 2, uniform in $x$. If such a bound held, it would show that $C_{\mathrm{rand}}(\mathrm{e})$ is a reliable estimator of the behaviour of the ratio $\frac{\|P(x)\|_Y^2}{\|x\|_X^2}$ in general. Notice that, in the favourable case when $P^{-1}$ is bounded, there exists $\delta_0 \in [0, C_{\mathrm{rand}}(\mathrm{e}))$ such that

$$\mathbb{P}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} < C_{\mathrm{rand}}(\mathrm{e}) - \delta\right) = 0$$

for all $\delta \in [\delta_0, C_{\mathrm{rand}}(\mathrm{e}))$.

In this general framework, estimate (14) does not hold, see Example 2. Using a concentration inequality, a weaker bound can be derived.

**Proposition 3.** *Assume that $Y = L^2(\Sigma, \mu)$, where $(\Sigma, S, \mu)$ is a measure space. Let $(\beta_k^\nu)_{k\in\mathbb{N}}$ be a sequence of independent random variables of Bernoulli type, on a probability space $(A, \mathcal{A}, \mathbb{P})$ with mean 0 and variance 1. Let $x \in X \setminus \{0\}$ and $x^\nu = \sum_{k=1}^\infty \beta_k^\nu \mathrm{e}_k x_k$. Then, for every $\delta > 0$ we have*

$$\mathbb{P}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} - \mathbb{E}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2}\right) < -\delta\right) < \exp\left(-\frac{\delta^2}{4K_{\mathrm{rand}}(\mathrm{e})^2}\right). \tag{15}$$

This result is based on an appropriate Hoeffding inequality; its proof is postponed to Appendix B. Note that (15) is not a large deviation result: the quantity under consideration is bounded between 0 and 1, and the upper bound obtained is not small. This is unavoidable, see Example 2.

4.2. **Can you reconstruct two numbers from their sum?** We collect here several observations that suggest that the positivity of the randomised stability constant may not be helpful for solving the inverse problems, not even for most unknowns.

4.2.1. *Instability arises for every $x \in X$.* We remind the reader why (8) renders inversion unstable. Hypothesis (8) implies that there exists a sequence $(x_n)_{n\in\mathbb{N}}$ such that

$$\|x_n\|_X = 1 \text{ and } \|Px_n\|_Y < \frac{1}{n} \text{ for all } n \in \mathbb{N}. \tag{16}$$

[1]This identity, discovered by Rellich in 1940 [43], reads

$$2\lambda^2 = \int_{\partial\Omega} \langle x, \nu\rangle \left(\frac{\partial\phi}{\partial\nu}\right)^2 d\mathcal{H}^{d-1}$$

for every eigenpair $(\lambda, \phi)$ of the Laplacian-Dirichlet operator, $\Omega$ being a bounded connected domain of $\mathbb{R}^n$ either convex or with a $C^{1,1}$ boundary.

Suppose that our measurements are not perfect, and are affected by a low level of noise $\delta$, $\|\delta\|_Y \leq \epsilon$, with $\epsilon > 0$. Then, for every $n$ such that $n\epsilon > 1$, we have

$$\|P(x + x_n) - P(x)\|_Y < \epsilon,$$

hence $x$ and $x+x_n$ correspond to the same measured data, even if $\|(x + x_n) - x\|_X = 1$. This is an unavoidable consequence of the unboundedness of $P^{-1}$, and is true for every $x \in X$, even if the randomised stability constant were positive (and possibly large).

4.2.2. *The dependence of $C_{\text{rand}}$ on the basis.* Lemma 1 shows that the ONB used to randomise our input plays a role, as it appears explicitly in the formula (9). The following proposition underscores that point. Namely, if we consider all possible randomisations with respect to all ONB of $X$ we recover the deterministic stability constant $C_{\text{det}}$.

**Proposition 4.** *We have*

$$\inf_{\mathrm{e}} C_{\text{rand}}(\mathrm{e}) = C_{\text{det}},$$

*where the infimum is taken over all ONB of $X$. In particular, if $P^{-1}$ is unbounded then $\inf_{\mathrm{e}} C_{\text{rand}}(\mathrm{e}) = 0$.*

*Proof.* By definition of $C_{\text{rand}}(\mathrm{e})$, we have that $C_{\text{rand}}(\mathrm{e}) \geq C_{\text{det}}$ for every ONB e, and so it remains to prove that

$$\inf_{\mathrm{e}} C_{\text{rand}}(\mathrm{e}) \leq C_{\text{det}}.$$

By definition of $C_{\text{det}}$, we can find a sequence $x_n \in X$ such that $\|x_n\|_X = 1$ for every $n$ and $\|Px_n\|_Y^2 \to C_{\text{det}}$. For every $n$, complete $x_n$ to an ONB of X, which we call $\mathrm{e}^{(n)}$. By Lemma 1 we have $C_{\text{rand}}(\mathrm{e}^{(n)}) \leq \|Px_n\|_Y^2$, and so

$$\inf_{\mathrm{e}} C_{\text{rand}}(\mathrm{e}) \leq \inf_{n} C_{\text{rand}}(\mathrm{e}^{(n)}) \leq \inf_{n} \|Px_n\|_Y^2 \leq \lim_{n\to+\infty} \|Px_n\|_Y^2 = C_{\text{det}}. \qquad \square$$

This result shows that, in general, the randomised stability constant strongly depends on the choice of the basis. There will always be bases for which it becomes arbitrarily small when $P^{-1}$ is unbounded.

It is also worth observing that for compact operators, which arise frequently in inverse problems, the randomised stability constant is always zero.

**Lemma 2.** *If $P$ is compact then $C_{\text{rand}}(\mathrm{e}) = 0$ for every ONB* e *of $X$.*

*Proof.* Since $\mathrm{e}_k$ tends to zero weakly in $X$, by the compactness of $P$ we deduce that $P(\mathrm{e}_k)$ tends to zero strongly in $Y$. Thus, by Lemma 1 we have

$$C_{\text{rand}}(\mathrm{e}) = \inf_{k} \|P(\mathrm{e}_k)\|_Y^2 \leq \lim_{k\to+\infty} \|P(\mathrm{e}_k)\|_Y^2 = 0,$$

as desired. $\square$

4.2.3. *Examples.* Let us now consider some examples. The first example is finite dimensional and the kernel of the operator is not trivial. Note that the definition of $C_{\text{rand}}(\mathrm{e})$ and all results above, except Proposition 2 and Lemma 2, are valid also in this case, with suitable changes due to the finiteness of the ONB.

**Example 1.** Choose $X = \mathbb{R}^2$ and $Y = \mathbb{R}$, and consider the map

$$\begin{aligned} S\colon \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto x + y. \end{aligned}$$

The associated inverse problem can be phrased: find the two numbers whose sum is given. This problem is ill-posed and impossible to solve. The deterministic stability constant vanishes

$$\inf_{(x_1,x_2)\in\mathbb{R}^2\setminus\{(0,0)\}} \frac{|x_1 + x_2|^2}{x_1^2 + x_2^2} = 0,$$

and $S^{-1}$ does not exist. However, the randomised constant obtained using the canonical basis is positive. Indeed, $|P(1,0)| = |P(0,1)| = 1$, therefore

$$C_{\mathrm{rand}}(\{(1,0),(0,1)\}) = \inf\{1,1\} = 1.$$

The positivity of this constant does not imply the existence of any useful method to perform the reconstruction of $x$ and $y$ from $x + y$, even for most $(x, y) \in \mathbb{R}^2$.

Had we chosen as orthonormal vectors $\frac{1}{\sqrt{2}}(1,1)$ and $\frac{1}{\sqrt{2}}(1,-1)$, since $|P(1,-1)| = 0$, we would have found

$$C_{\mathrm{rand}}\left(\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\right) = 0.$$

One may wonder whether the features highlighted above are due to the fact that the kernel is not trivial. That is not the case, as the following infinite-dimensional generalisation with trivial kernel shows.

**Example 2.** Consider the case when $X = Y = \ell^2$, equipped with the canonical euclidean norm. Let $\mathrm{e} = \{\mathrm{e}_k\}_{k=0}^{+\infty}$ denote the canonical ONB of $\ell^2$. Take a sequence $(\eta_n)_{n\in\mathbb{N}_0}$ such that $\eta_n > 0$ for all $n$, and $\lim_{n\to\infty} \eta_n = 0$. We consider the operator $P$ defined by

$$P(\mathrm{e}_{2n}) = \mathrm{e}_{2n} + \mathrm{e}_{2n+1}, \qquad P(\mathrm{e}_{2n+1}) = \mathrm{e}_{2n} + (1 + \eta_n)\mathrm{e}_{2n+1}.$$

The operator $P$ may be represented with respect to the canonical basis e by the block-diagonal matrix

$$P = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots \\ 1 & 1+\eta_0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 1 & 1+\eta_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

In other words, $P$ may be expressed as

$$P(x) = (x_0 + x_1, x_0 + (1 + \eta_0)x_1, x_2 + x_3, x_2 + (1 + \eta_1)x_3, \dots), \qquad x \in \ell^2.$$

We note that $\ker T = \{0\}$ and that its inverse is given by

$$P^{-1}(y) = ((1+\eta_0^{-1})y_0 - \eta_0^{-1}y_1, \eta_0^{-1}(y_1 - y_0), (1+\eta_1^{-1})y_2 - \eta_1^{-1}y_3, \eta_1^{-1}(y_3 - y_2), \dots),$$

which is an unbounded operator since $\eta_n^{-1} \to +\infty$. Given the block diagonal structure of this map, the inversion consists of solving countably many inverse problems (i.e., linear systems) of the form

$$\begin{cases} x_{2n} + x_{2n+1} = y_{2n}, \\ x_{2n} + (1 + \eta_n)x_{2n+1} = y_{2n+1}. \end{cases}$$

As soon as $\tilde{n}$ is such that $\eta_{\tilde{n}}$ becomes smaller than the noise level, all the following inverse problems for $n \geq \tilde{n}$ are impossible to be solved, since they reduce to the "sum of two numbers" discussed in Example 1.

Note that

$$\|P(\mathrm{e}_{2n})\|_2^2 = 2, \qquad \|P(\mathrm{e}_{2n+1})\|_2^2 = 1 + (1+\eta_n)^2,$$

therefore

$$C_{\mathrm{rand}}(\mathrm{e}) = 2.$$

If we choose instead the rotated orthonormal basis,

$$v_{2n} = \frac{1}{\sqrt{2}}(\mathrm{e}_{2n} + \mathrm{e}_{2n+1}), \quad v_{2n+1} = \frac{1}{\sqrt{2}}(\mathrm{e}_{2n} - \mathrm{e}_{2n+1}),$$

then $P(v_{2n+1}) = -\eta_n \frac{1}{\sqrt{2}} \mathrm{e}_{2n+1}$, and so

$$C_{\mathrm{rand}}(\{v_k\}_k) = \inf_k \|P(v_k)\|_2^2 \leq \lim_{n\to\infty} \|P(v_{2n+1})\|_2^2 = 0.$$

We now turn to (14) and (15). For some $k \geq 0$, consider $x = \mathrm{e}_{2k}$. Then $P(x^\nu) = \beta_{2k}^\nu(\mathrm{e}_{2k} + \mathrm{e}_{2k+1})$ and therefore $\|P(x^\nu)\|_Y = \|P(x)\|_Y$: there is no deviation as

$$\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} = \mathbb{E}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2}\right) = C_{\mathrm{rand}}(\mathrm{e}) = 2.$$

Thus, the probabilities in (14) and (15) are equal to 0, and the inequalities are trivial.

However, alternatively, consider $x = \mathrm{e}_{2k} + \mathrm{e}_{2k+1}$. Then $\|P(x^\nu)\|_Y^2 = 4 + (2+\eta_{2k})^2$ if $\beta_1^\nu \beta_2^\nu = 1$ and $\|P(x^\nu)\|_Y^2 = \eta_{2k}^2$ if $\beta_1^\nu \beta_2^\nu = -1$. Therefore

$$\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} - C_{\mathrm{rand}}(\mathrm{e}) = \begin{cases} \frac{(2+\eta_{2k})^2}{2} & \text{with probability } \frac{1}{2}, \\ \frac{\eta_{2k}^2}{2} - 2 & \text{with probability } \frac{1}{2}. \end{cases}$$

As a consequence, (14) cannot be true in general for every $x$. Similarly, we have

$$\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} - \mathbb{E}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2}\right) = \begin{cases} 2 + \eta_{2k} & \text{with probability } \frac{1}{2}, \\ -2 - \eta_{2k} & \text{with probability } \frac{1}{2}, \end{cases}$$

and the left-hand side of (15) can indeed be large for some $x$.

It is worth observing that a very similar example is considered in [32] to show that particular complications may arise when using neural networks for solving some inverse problems, even naive and small scale (cfr. §A.3).

These examples show that considering the observability constant for a particular basis sheds little light on a potential stable inversion of the problem in average, and that considering all possible randomisations leads to the same conclusion as the deterministic case (confirming Proposition 4).

4.3. **Linear versus nonlinear problems.** The pitfalls we encountered when we tried to make use of the randomised stability constant all stem from the linearity of the problems we are considering. The seminal work of Burq and Tzvetkov [12], which showed existence of solutions in super-critical regimes for a semilinear problem did not involve tinkering with associated linear operator (the wave equation); it is the nonlinearity that controlled the critical threshold. In both compressed sensing and passive imaging with random noise sources, nonlinearity plays a key role; further, deep networks are nonlinear maps (cfr. Appendix A).

The naive intuition we discussed earlier, namely, that extreme situations do not occur often, is more plausible for nonlinear maps where pathological behaviour is local.

**Example 3.** As a toy finite-dimensional example, consider the map

$$T\colon \mathbb{R} \to \mathbb{R}, \qquad T(x) = x(x-\epsilon)(x+\epsilon),$$

for some small $\epsilon > 0$. Then $T$ can be stably inverted outside of a region with size of order $\epsilon$, since there the inverse is continuous. Thus, a random initial data has little chance of falling precisely in the problematic region. Such a case cannot happen with linear maps.

**Example 4.** Let $A\colon H \to H$ be an unbounded linear operator on a Hilbert space with compact resolvent, so that the spectrum of $A$ is discrete. Define the nonlinear map

$$T\colon H \times [0,1] \to H \times [0,1], \qquad T(x,\lambda) = (Ax + \lambda x, \lambda).$$

Note that $A + \lambda I$ is invertible with probability 1 if $\lambda$ is chosen uniformly in $[0,1]$. Thus, if $H \times [0,1]$ is equipped with a product probability measure whose factor on $[0,1]$ is the uniform probability, then $x$ may be reconstructed from $T(x)$ with probability 1.

## 5. Concluding remarks

In this paper we focused on the randomised stability constant for linear inverse problems, which we introduced as a generalisation of the randomised observability constant.

We argue that, despite its intuitive and simple definition, the randomised stability constant has no implications in the practical solution of inverse problems, even for most unknowns. As the examples provided show, this may be due to the linearity of the problem. With nonlinear problems, the situation is expected to be completely different. It could be that the randomised stability constant is meaningful in the context of a nonlinear inversion process, involving for example a hierarchical decomposition [46, 34], but we do not know of results in that direction: this is left for future research.

## References


[1] N. Akhtar and A. Mian. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 6:14410–14430, 2018.

[2] Giovanni S. Alberti and Yves Capdeboscq. *Lectures on elliptic methods for hybrid inverse problems*, volume 25 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2018.

[3] Habib Ammari. *An introduction to mathematics of emerging biomedical imaging*, volume 62 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Berlin, 2008.

[4] Habib Ammari, Josselin Garnier, Hyeonbae Kang, Loc Hoang Nguyen, and Laurent Seppecher. *Multi-Wave Medical Imaging*. WORLD SCIENTIFIC (EUROPE), 2017.

[5] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. Deep learning for photoacoustic tomography from sparse data. *Inverse Problems in Science and Engineering*, 0(0):1–19, 2018.

[6] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction – Does AI come at a cost? *arXiv preprint arXiv:1902.05300*, 2019.

[7] Simon Arridge and Andreas Hauptmann. Networks for Nonlinear Diffusion Problems in Imaging. *arXiv preprint arXiv:1811.12084*, 2018.
[8] Claude Bardos, Gilles Lebeau, and Jeffrey Rauch. Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary. *SIAM J. Control Optim.*, 30(5):1024–1065, 1992.
[9] Jens Berg and Kaj Nyström. Neural network augmented inverse problems for PDEs. *arXiv preprint arXiv:1712.09685*, 2017.
[10] Tatiana A. Bubba, Gitta Kutyniok, Matti Lassas, Maximilian März, Wojciech Samek, Samuli Siltanen, and Vignesh Srinivasan. Learning the invisible: a hybrid deep learning–shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6):064002, 38, 2019.
[11] Nicolas Burq. Random data Cauchy theory for dispersive partial differential equations. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1862–1883. Hindustan Book Agency, New Delhi, 2010.
[12] Nicolas Burq and Nikolay Tzvetkov. Random data Cauchy theory for supercritical wave equations. II. A global existence result. *Invent. Math.*, 173(3):477–496, 2008.
[13] Nicolas Burq and Nikolay Tzvetkov. Probabilistic well-posedness for the cubic wave equation. *J. Eur. Math. Soc. (JEMS)*, 16(1):1–30, 2014.
[14] Emmanuel Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.
[15] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
[16] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.
[17] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
[18] Jinchao Feng, Qiuwan Sun, Zhe Li, Zhonghua Sun, and Kebin Jia. Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography. *J. Biomed. Opt.*, 24(5):051407, 2018.
[19] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.
[20] Josselin Garnier and George Papanicolaou. *Passive imaging with ambient noise*. Cambridge University Press, Cambridge, 2016.
[21] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.
[22] Markus Haltmeier, Stephan Antholzer, and Johannes Schwab. *Deep Learning for Image Reconstruction*. World Scientific, 2019.
[23] S. J. Hamilton and A. Hauptmann. Deep d-bar: Real-time electrical impedance tomography imaging with deep neural networks. *IEEE Transactions on Medical Imaging*, 37(10):2367–2377, Oct 2018.
[24] Alemdar Hasanov Hasanoğlu and Vladimir G. Romanov. *Introduction to inverse problems for differential equations*. Springer, Cham, 2017.
[25] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge. Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 37(6):1382–1393, June 2018.
[26] Emmanuel Humbert, Yannick Privat, and Emmanuel Trlat. Observability properties of the homogeneous wave equation on a closed manifold. *Communications in Partial Differential Equations*, 44(9):749–772, 2019.
[27] Victor Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006.
[28] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, Sep. 2017.
[29] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[30] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2011.
[31] Camille Laurent and Matthieu Léautaud. Quantitative unique continuation for operators with partially analytic coefficients. Application to approximate control for waves. *Journal of the European Mathematical Society*, to appear.
[32] Peter Maass. *Deep Learning for Trivial Inverse Problems*, pages 195–209. Springer International Publishing, Cham, 2019.
[33] Sébastien Martin and Charles T. M. Choi. A post-processing method for three-dimensional electrical impedance tomography. *Scientific Reports*, 7(1):7212, 2017.
[34] Klas Modin, Adrian Nachman, and Luca Rondi. A multiscale theory for image registration and nonlinear inverse problems. *Adv. Math.*, 346:1009–1066, 2019.
[35] R. E. A. C. Paley. On some series of functions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(4):458–474, 1930.
[36] R. E. A. C. Paley and A. Zygmund. On some series of functions, (1). *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(3):337–357, 1930.
[37] R. E. A. C. Paley and A. Zygmund. On some series of functions, (3). *Mathematical Proceedings of the Cambridge Philosophical Society*, 28(2):190–205, 1932.
[38] Yannick Privat, Emmanuel Trélat, and Enrique Zuazua. Optimal observation of the one-dimensional wave equation. *J. Fourier Anal. Appl.*, 19(3):514–544, 2013.
[39] Yannick Privat, Emmanuel Trélat, and Enrique Zuazua. Optimal shape and location of sensors for parabolic equations with random initial data. *Arch. Ration. Mech. Anal.*, 216(3):921–981, 2015.
[40] Yannick Privat, Emmanuel Trélat, and Enrique Zuazua. Optimal observability of the multi-dimensional wave and Schrödinger equations in quantum ergodic domains. *J. Eur. Math. Soc. (JEMS)*, 18(5):1043–1111, 2016.
[41] Yannick Privat, Emmanuel Trélat, and Enrique Zuazua. Randomised observation, control and stabilization of waves [Based on the plenary lecture presented at the 86th Annual GAMM Conference, Lecce, Italy, March 24, 2015]. *ZAMM Z. Angew. Math. Mech.*, 96(5):538–549, 2016.
[42] Yannick Privat, Emmanuel Trélat, and Enrique Zuazua. Spectral shape optimization for the neumann traces of the dirichlet-laplacian eigenfunctions. *Calculus of Variations and Partial Differential Equations*, 58(2):64, Mar 2019.
[43] Franz Rellich. Darstellung der eigenwerte von $\delta u+\lambda u= 0$ durch ein randintegral. *Mathematische Zeitschrift*, 46(1):635–636, 1940.
[44] Otmar Scherzer, editor. *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*. Springer, New York, second edition, 2015.
[45] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.
[46] Eitan Tadmor, Suzanne Nezzar, and Luminita Vese. A multiscale image representation using hierarchical $(BV, L^2)$ decompositions. *Multiscale Model. Simul.*, 2(4):554–579, 2004.
[47] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
[48] Z. Wei, D. Liu, and X. Chen. Dominant-current deep learning scheme for electrical impedance tomography. *IEEE Transactions on Biomedical Engineering*, pages 1–1, 2019.
[49] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, June 2018.
[50] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555:487 EP –, Mar 2018.

## Appendix A. Examples of techniques based on randomisation

In this appendix we briefly review three different techniques for solving inverse problems where randomisation plays a crucial role. We do not aim at providing an exhaustive overview, or at reporting on the most recent advances, or at discussing

the many variants that have been studied. The examples we present are used to contrast possible different approaches, and the level of mathematical understanding associated with them.

A.1. **Compressed sensing.** Since the seminal works [15, 16], compressed sensing (CS) has provided a theoretical and numerical framework to overcome Nyqvist criterion in sampling theory for the reconstruction of sparse signals. In other words, sparse signals in $\mathbb{C}^n$ may be reconstructed from $k$ discrete Fourier measurements, with $k$ smaller than $n$ and directly proportional to the sparsity of the signal (up to log factors, see eqn. (18) below). Let us give a quick overview of the main aspects of CS, in order to show how it fits in the general framework of section 1. For additional details, the reader is referred to the book [19], and to the references therein.

Given $s \in \mathbb{N} = \{1, 2, \dots\}$, let $X$ be the set of $s$-sparse signals in $\mathbb{C}^n$, namely

$$X = \{x \in \mathbb{C}^n : \# \operatorname{supp} x \leq s\}.$$

Let $F \colon \mathbb{C}^n \to \mathbb{C}^n$ denote the discrete Fourier transform. In fact, any unitary map may be considered, by means of the notion of incoherence [14]. In any case, the Fourier transform is a key example for the applications to Magnetic Resonance Imaging and Computerised Tomography (via the Fourier Slice Theorem). In order to subsample the Fourier measurements, we consider subsets $S_a$ of cardinality $k$ of $\{1, 2, \dots, n\}$ and parametrise them with $a \in \{1, 2, \dots, \binom{n}{k}\}$. Let $Y = \mathbb{C}^k$ and $P_a \colon \mathbb{C}^n \to \mathbb{C}^k$ be the projection selecting the entries corresponding to $S_a$. We then define the measurement map

$$T_a = P_a \circ F \colon X \to \mathbb{C}^k.$$

In other words, $T_a$ is the partial Fourier transform, since only the frequencies in $S_a$ are measured, and $\# S_a = k \leq n$.

Given an unknown signal $x_0 \in X$, we need to reconstruct it from the partial knowledge of its Fourier measurements represented by $y := T_a(x_0)$. The sparsity of $x_0$ has to play a crucial role in the reconstruction, since as soon as $k < n$ the map $P_a \circ F \colon \mathbb{C}^n \to \mathbb{C}^k$ necessarily has a non-trivial kernel. It is worth observing that sparsity is a nonlinear condition: if $X$ were a linear subspace of $\mathbb{C}^n$, the problem would be either trivial or impossible, depending on $\ker(P_a \circ F) \cap X$. Thus nonlinearity plays a crucial role here.

The simplest reconstruction algorithm is to look for the sparsest solution to $T_a x = y$, namely to solve the minimisation problem

$$\min_{x \in \mathbb{C}^n} \|x\|_0 \quad \text{subject to } T_a x = y,$$

where $\|x\|_0 = \# \operatorname{supp} x$. However, this problem is NP complex, and its direct resolution impractical. Considering the convex relaxation $\|\cdot\|_1$ of $\|\cdot\|_0$ leads to a well-defined minimisation problem

$$\min_{x \in \mathbb{C}^n} \|x\|_1 \quad \text{subject to } T_a x = y, \tag{17}$$

whose solution may be easily found by convex optimisation (in fact, by linear programming).

The theory of CS guarantees exact reconstruction. More precisely, if $\tilde{x}$ is a minimiser of (17), then $\tilde{x} = x_0$ with high probability, provided that

$$k \geq C s \log n, \tag{18}$$

and that $a$ is chosen uniformly at random in $\{1, 2, \dots, \binom{n}{k}\}$ (namely, the subset $S_a$ is chosen uniformly at random among all the subsets of cardinality $k$ of $\{1, 2, \dots, n\}$) [15]. In addition, in the noisy case, with measurements of the form $y = T_a(x_0) + \eta$ where $\|\eta\|_2 \leq \epsilon$, by relaxing the equality "$T_a x = y$" to the inequality "$\|T_a x - y\|_2 \leq \epsilon$" in (17), one obtains the linear convergence rate $\|x_0 - \tilde{x}\|_2 \leq C\epsilon$, namely, the solution is stable.

In summary, CS allows for the stable reconstruction of *all* sparse signals from partial Fourier measurements, for *most* choices of the measured frequencies. The corresponding forward map $T_a \colon X \to \mathbb{C}^k$ is *nonlinear*, simply because $X$ is not a vector space.

A.2. **Passive imaging with random noise sources.** The material presented in this part is taken from [20], to which the reader is referred for more detailed discussion on this topic.

A typical multistatic imaging problem is the recovery of some properties of a medium with velocity of propagation $c(x) > 0$ from some measurements at locations $x_j \in \mathbb{R}^3$ of the solution $u(t, x)$ of the wave equation

$$\partial_t^2 u(t, x) - c(x)^2 \Delta u(t, x) = f(t)\delta(x - y), \qquad (t, x) \in \mathbb{R} \times \mathbb{R}^3,$$

where $f(t)$ is the source pulse located at $y$. One of the major applications of this setup is geophysical imaging, where one wants to recover properties of the structure of the earth from measurements taken on the surface. Generating sources in this context is expensive and disruptive. Earthquakes are often used as sources, but they are rare and isolated events. Yet, noisy signals, as they may be recorded by seismographers, even if low in amplitude, may be relevant and useful even in absence of important events.

The key idea is to consider the data generated by random sources (e.g., in seismology, those related to the waves of the sea). The equation becomes

$$\partial_t^2 u(t, x) - c(x)^2 \Delta u(t, x) = n(t, x), \qquad (t, x) \in \mathbb{R} \times \mathbb{R}^3,$$

where the source term $n(t, x)$ is a zero-mean stationary random process that models the ambient noise sources. We assume that its autocorrelation function is

$$\mathbb{E}(n(t_1, y_1)n(t_2, y_2)) = F(t_2 - t_1)K(y_1)\delta(y_1 - y_2),$$

where $F$ is the time correlation function (normalised so that $F(0) = 1$) and $K$ characterises the spatial support of the sources. The presence of $\delta(y_1 - y_2)$ makes the process $n$ delta-correlated in space.

The reconstruction is based on the calculation of the empirical cross correlation of the signals recorded at $x_1$ and $x_2$ up to time $T$:

$$C_T(\tau, x_1, x_2) = \frac{1}{T} \int_0^T u(t, x_1)u(t + \tau, x_2)\, dt.$$

Its expectation is the statistical cross correlation

$$\mathbb{E}(C_T(\tau, x_1, x_2)) = C^{(1)}(\tau, x_1, x_2),$$

which is given by

$$C^{(1)}(\tau, x_1, x_2) = \frac{1}{2\pi} \int_{\mathbb{R}\times\mathbb{R}^3} \hat{F}(\omega)K(y)\overline{\hat{G}(\omega, x_1, y)}\hat{G}(\omega, x_2, y)e^{-i\omega\tau}\, dtdy, \tag{19}$$

where $\hat{\cdot}$ denotes the Fourier transform in time and $G(t,x,y)$ is the time-dependent Green's function. Moreover, $C_T$ is a self-averaging quantity, namely

$$\lim_{T\to+\infty} C_T(\tau,x_1,x_2) = C^{(1)}(\tau,x_1,x_2)$$

in probability.

The role of randomised sources is now clear: from the measured empirical cross correlation $C_T$ with large values of $T$ it is possible to estimate, with high probability, the statistical cross correlation $C^{(1)}$. Using (19), from $C^{(1)}(\tau,x_1,x_2)$ it is possible to recover (some properties of) the Green function $G$, which yield useful information about the medium, such as travel times.

A.3. **Deep Learning in inverse problems for PDE.** Convolutional Neural Networks have recently been used for a variety of imaging and parameter reconstruction problems [22], including Electrical Impedance Tomography (EIT) [33, 23, 48], optical tomography [18], inverse problems with internal data [9], diffusion problems in imaging [7], computerised tomography [28, 10], photoacoustic tomography [5, 25] and magnetic resonance imaging [50, 49]. In the following brief discussion, we decided to focus on inverse problems for partial differential equations (PDE), and in particular on EIT, but similar considerations are valid for most methods cited above.

Significant improvement has been observed in EIT with deep learning compared to previous imaging approaches. Let $\Omega \subseteq \mathbb{R}^d$, $d \geq 2$, be a bounded Lipschitz domain with outer unit normal $\nu$. The data in EIT is (a part of) of the Dirichlet-to-Neumann map

$$\begin{array}{rccc} \Lambda_\sigma\colon & H^{\frac{1}{2}}(\partial\Omega)/\mathbb{R} & \to & H^{-\frac{1}{2}}(\partial\Omega)/\mathbb{R} \\ & v & \mapsto & \sigma\nabla u\cdot\nu|_{\partial\Omega} \end{array}$$

where $u(x)$ denotes the unique solution of the elliptic problem

$$\begin{cases} \operatorname{div}(\sigma(x)\nabla u(x)) = 0 & x\in\Omega, \\ u(x) = v(x) & x\in\partial\Omega, \end{cases}$$

and $\sigma(x) > 0$ is the unknown conductivity. The experimental data is usually part of the inverse map, namely the Neumann-to-Dirichlet map $\Lambda_\sigma^{-1}$. In two dimensions, provided that the electrodes are equally separated on the unit disk, the data may be modelled by

$$T_N\Lambda_\sigma^{-1}T_N,$$

where $T_N$ is the $L^2$ projection on $\operatorname{span}\{\theta\mapsto\cos(n\theta) : 1\leq n\leq N\}$, where $N$ is the number of electrodes: it is the partial Fourier transform limited to the first $N$ coefficients.

Direct neural network inversion approaches suffer from drawbacks alike direct non-regularised inversion attempts: the output is very sensitive to measurement errors and small variations. Successful approaches to Deep Learning EIT [33, 23, 48], and to other parameter identification problems in PDE, often involve two steps.

The first step consists in the derivation of an approximate conductivity $\sigma$ by a stable, albeit blurry, regularised inversion method. For instance, in [23] the "D-bar" equation is used, while in [33] a one-step Gauss-Newton method is used. In both cases, the output of this step is a representation of the conductivity coefficient, which depends on the inversion method used. This first step is deterministic and its analysis is well understood. The forward problem, relating the conductivity to the

Dirichlet-to-Neumann map, is nonlinear, independently of the inversion algorithm used. Indeed, the map $\Lambda_\sigma$ is a linear operator, but $\sigma \mapsto \Lambda_\sigma$ is nonlinear.

The second, post-processing, step uses a neural network to "deblur" the image, and in fact restores details that were not identifiable after the first step.

The second step is not unlike other successful usage of deep-learning approaches for image classification; in general they are known to be successful only with very high probability (and in turn for random unknowns). More precisely, since the findings of [45], deep networks are known to be vulnerable to so-called "adversarial perturbations" (see the review article [1] and the references therein). Given an image $x$ that is correctly classified by the network with high confidence, an adversarial perturbation is a small perturbation $p$ such that the images $x$ and $y = x + p$ are visually indistinguishable but the perturbed image $y$ is misclassified by the network, possibly with high confidence too. State-of-the-art classification networks are successful for the vast majority of natural images, but are very often vulnerable to such perturbations.

These instabilities are not specific to image classification problems; they appear in the same way in image reconstruction [6]. In this case, given an image that is well-reconstructed by the network, it is possible to create another image that is visually indistinguishable from the original one, but that is not well-reconstructed by the network.

A full mathematical understanding of deep networks is still lacking, and the reasons of this phenomenon are not fully known. However, the large Lipschitz constant of the network certainly plays a role, since it is a sign of potential instability: in order for the network to be effective, the weights of its linear steps need to be chosen large enough, and the composition of several layers yields an exponentially large constant.

## Appendix B. Proofs of the large deviation estimates

The proofs of Theorem 1 and Proposition 3 rest upon a classical large deviation estimate, the so-called Hoeffding inequality, see e.g. [11, Prop. 2.2], whose proof is recalled for the convenience of the reader.

**Proposition 5.** *Let $(\alpha_n^\nu)_{n\geq 1}$ be a sequence of sequence of independent random variables of Bernoulli type, on a probability space $(A, \mathcal{A}, \mathbb{P})$ with mean 0 and variance 1. Then, for any $t > 0$ and any sequence $(u_n)_{n\geq 1} \in \ell^2(\mathbb{C})$, we have*

$$\mathbb{P}\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n < -t\right) = \mathbb{P}\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n > t\right) \leq \exp\left(-\frac{1}{2}\frac{t^2}{\sum_{n=1}^{+\infty} |v_n|^2}\right).$$

*Proof.* There holds $\mathbb{E}\left(\exp\left(\alpha_n^\nu v_n\right)\right) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{(\alpha_n^\nu v_n)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbb{E}\left((\alpha_n^\nu v_n)^k\right)$. All odd powers of $k$ vanish as $\alpha_n^\nu$ has zero mean and is symmetrical. Therefore for

any $\lambda > 0$

$$\begin{aligned}\mathbb{E}\left(\exp\left(\lambda\alpha_n^\nu v_n\right)\right) &= \sum_{k=0}^{\infty} \frac{1}{(2k)!}\mathbb{E}\left(\left(\lambda\alpha_n^\nu v_n\right)^{2k}\right)\\ &\leq \sum_{k=0}^{\infty} \frac{\lambda^{2k} v_n^{2k}}{2^k\,(k!)}\\ &= \exp\left(\frac{\lambda^2}{2} v_n^2\right).\end{aligned}$$

Applying Chernoff's inequality, we obtain for any $\lambda > 0$

$$\begin{aligned}\mathbb{P}\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n > t\right) &\leq \mathbb{E}[\exp\left(\lambda\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n\right)\right)]\exp\left(-\lambda t\right)\\ &\leq \exp\left(\lambda^2 \frac{1}{2}\sum_{n=1}^{+\infty} v_n^2 - \lambda t\right).\end{aligned}$$

Choosing $\lambda = \frac{t}{\sum v_n^2}$ this yields

$$\mathbb{P}\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n < -t\right) = \mathbb{P}\left(\sum_{n=1}^{+\infty} \alpha_n^\nu v_n > t\right) \leq \exp\left(-\frac{t^2}{2\sum v_n^2}\right),$$

as desired. □

We are now ready to prove Theorem 1.

*Proof of Theorem 1.* Fix $r \geq 2$ and set $\mathcal{Y}^\nu = \frac{P(x^\nu)}{\|x^\nu\|_X}$. Markov's inequality yields

$$\mathbb{P}\left(\|\mathcal{Y}^\nu\|_Y \geq \delta\right) = \mathbb{P}\left(\|\mathcal{Y}^\nu\|_Y^r \geq \delta^r\right) \leq \frac{1}{\delta^r}\mathbb{E}\left(\|\mathcal{Y}^\nu\|_Y^r\right). \tag{20}$$

Let us denote by $\mathcal{L}_\nu^r$ the standard Lebesgue space with respect to the probability measure $d\mathbb{P}$. Recall that $Y = L^2(\Sigma, \mu)$. To provide an estimate of the right-hand side, notice that

$$\begin{aligned}\mathbb{E}\left(\|\mathcal{Y}^\nu\|_Y^r\right) &= \int_A \|\mathcal{Y}^\nu\|_Y^r \, d\mathbb{P}(\nu)\\ &= \int_A \left(\int_\Sigma |\mathcal{Y}^\nu(s)|^2 \, d\mu(s)\right)^{r/2} d\mathbb{P}(\nu)\\ &= \left\| \int_\Sigma |\mathcal{Y}^\nu(s)|^2 \, d\mu(s)\right\|_{\mathcal{L}_\nu^{r/2}}^{r/2}\\ &\leq \left(\int_\Sigma \left\||\mathcal{Y}^\nu(s)|^2\right\|_{\mathcal{L}_\nu^{r/2}} d\mu(s)\right)^{r/2}\\ &= \left(\int_\Sigma \left\|\mathcal{Y}^\nu(s)\right\|_{\mathcal{L}_\nu^r}^2 d\mu(s)\right)^{r/2}\\ &= \left\|s \mapsto \|\mathcal{Y}^\nu(s)\|_{\mathcal{L}_\nu^r}\right\|_Y^r\end{aligned} \tag{21}$$

by using Jensen's inequality.

Furthermore, for a.e. $s \in \Sigma$, we have[2]

$$\|\mathcal{Y}^\nu(s)\|_{\mathcal{L}^r_\nu}^r = \int_A |\mathcal{Y}^\nu(s)|^r d\mathbb{P}(\nu) = \int_0^{+\infty} r u^{r-1} \mathbb{P}(|\mathcal{Y}^\nu(s)| > u)\, du$$

and by using Proposition 5 and the fact that $\mathcal{Y}^\nu(s)$ reads

$$\mathcal{Y}^\nu(s) = \frac{\sum_{k=1}^\infty \beta_k^\nu x_k \left(P\mathrm{e}_k\right)(s)}{\sqrt{\sum_{k=1}^\infty |x_k|^2}},$$

one gets

$$\|\mathcal{Y}^\nu(s)\|_{\mathcal{L}^r_\nu}^r \leq 2 \int_0^{+\infty} r u^{r-1} \exp\left( -\frac{1}{2} \frac{\sum_{k=1}^\infty |x_k|^2 u^2}{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(s)|^2 |x_k|^2} \right) du.$$

As a consequence, by using the change of variable

$$v = \frac{\sqrt{\sum_{k=1}^\infty |x_k|^2}\, u}{\sqrt{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(s)|^2 |x_k|^2}}$$

we get

$$\|\mathcal{Y}^\nu(s)\|_{\mathcal{L}^r_\nu}^r \leq C(r) \left( \frac{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(s)|^2 |x_k|^2}{\sum_{k=1}^\infty |x_k|^2} \right)^{r/2}$$

with

$$C(r) = 2 \int_0^{+\infty} r v^{r-1} e^{-\frac{1}{2}v^2}\, dv.$$

An elementary computation yields $C(r) < r^r$. Therefore,

$$\|\mathcal{Y}^\nu(s)\|_{\mathcal{L}^r_\nu} < \left( r^2 \frac{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(s)|^2 |x_k|^2}{\sum_{k=1}^\infty |x_k|^2} \right)^{1/2}.$$

According to (21), we infer that

$$\begin{aligned} \mathbb{E}\left(\|\mathcal{Y}^\nu\|_Y^r\right) &< \left\| \left( r^2 \frac{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(\cdot)|^2 |x_k|^2}{\sum_{k=1}^\infty |x_k|^2} \right)^{1/2} \right\|_Y^r \\ &= \left\| r^2 \frac{\sum_{k=1}^{+\infty} |\left(P\mathrm{e}_k\right)(\cdot)|^2 |x_k|^2}{\sum_{k=1}^\infty |x_k|^2} \right\|_{L^1(\Sigma)}^{r/2}. \end{aligned}$$

From (13) and estimate (20), we get

$$\mathbb{P}\left(\|\mathcal{Y}^\nu\|_Y \geq \delta\right) < \frac{1}{\delta^r} \left( r^2 \frac{\sum_{k=1}^{+\infty} \|P\mathrm{e}_k\|_Y^2 |x_k|^2}{\sum_{k=1}^\infty |x_k|^2} \right)^{r/2} \leq \left( \frac{K_{\mathrm{rand}}(\mathrm{e})}{\delta^2} r^2 \right)^{r/2},$$

using the triangle inequality. Minimising the upper bound with respect to $r$, that is, choosing

$$r^2 = \frac{e^{-2}\delta^2}{K_{\mathrm{rand}}(\mathrm{e})}$$

[2]Here, we use that if $X$ denotes a non-negative random variable and $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$, then

$$\mathbb{E}(\varphi(X)) = \int_0^{+\infty} \varphi'(u) \mathbb{P}(X > u)\, du.$$

in the inequality above, one finally obtains

$$\mathbb{P}\left(\|\mathcal{Y}^\nu\|_Y \geq \delta\right) \leq \exp\left(-\frac{e^{-1}\delta}{\sqrt{K_{\mathrm{rand}}(\mathrm{e})}}\right). \tag{22}$$

Note that we have assumed $r \geq 2$, thus implicitly posited that $e^{-2}\delta^2 \geq 4K_{\mathrm{rand}}(\mathrm{e})$; we multiply the bound by $\exp(2)$ to cover the other case. □

We conclude by proving Proposition 3.

*Proof of Proposition 3.* As in (10), we have

$$\|P\left(x^\nu\right)\|_Y^2 = \sum_{k=1}^{\infty} x_k^2 \left\|P\left(e_k\right)\right\|_Y^2 + 2\sum_{\substack{k,l \\ k<l}} \beta_k^\nu \beta_l^\nu x_k x_l \left\langle P\left(\mathrm{e}_k\right), P\left(\mathrm{e}_l\right)\right\rangle_Y. \tag{23}$$

Observing that the family $\alpha_{k,l}^\nu = \{\beta_k^\nu \beta_l^\nu\}_{\substack{k,l \\ k<l}}$ is made of independent Bernoulli variables, equal almost surely to $-1$ or $1$, with $1/2$ as probability of success, we apply Proposition 5 with

$$v_{k,l} = 2\frac{x_k}{\|x\|_X}\frac{x_l}{\|x\|_X}\left\langle P\left(\mathrm{e}_k\right), P\left(\mathrm{e}_l\right)\right\rangle_Y$$

and obtain for all $\delta > 0$

$$\mathbb{P}\left(\frac{\|P(x^\nu)\|_Y^2}{\|x\|_X^2} - \mathbb{E}\left(\frac{\|P\left(x^\nu\right)\|_Y^2}{\|x\|_X^2}\right) < -\delta\right) < \exp\left(-\frac{\delta^2}{2K}\right),$$

with

$$\begin{aligned}
K &= \sum_{\substack{k,l \\ k<l}} \left(2\frac{x_k x_l}{\|x\|_X^2}\left\langle P\left(\mathrm{e}_k\right), P\left(\mathrm{e}_l\right)\right\rangle_Y\right)^2 \\
&\leq 2\sum_{k,l}\left(\frac{|x_k|\,|x_l|}{\|x\|_X^2}\left\|P\mathrm{e}_k\right\|_Y\left\|P\mathrm{e}_l\right\|_Y\right)^2 \\
&= 2\left(\sum_k \frac{|x_k|^2}{\|x\|_X^2}\left\|P\mathrm{e}_k\right\|_Y^2\right)^2 \\
&\leq 2K_{\mathrm{rand}}(\mathrm{e})^2.
\end{aligned}$$

□

MaLGa Center, Department of Mathematics, University of Genoa, Via Dodecaneso 35, 16146 Genova, Italy.
*E-mail address*: giovanni.alberti@unige.it

Universit de Paris, CNRS, Sorbonne Universit, Laboratoire Jacques-Louis Lions, UMR 7598, Paris, France
*E-mail address*: yves.capdeboscq@sorbonne-universite.fr

IRMA, Universit de Strasbourg, CNRS UMR 7501, 7 rue Ren Descartes, 67084 Strasbourg, France
*E-mail address*: yannick.privat@unistra.fr