ÖAW AUSTRIAN ACADEMY OF SCIENCES

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

RICAM JOHANN-RADON-INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Complexity of linear ill-posed problems in Hilbert space

P. Mathe, S. Pereverzyev

RICAM-Report 2016-09

COMPLEXITY OF LINEAR ILL-POSED PROBLEMS IN HILBERT SPACE

PETER MATHÉ AND SERGEI V. PEREVERZEV

Abstract. Information complexity of ill-posed problems may be seen as controversial. On the one hand side there were pessimistic results stating that the complexity is infinite, while on the other hand side the theory of ill-posed problems is well developed. In contrast to well-posed problems (continuous solution operators) the complexity analysis of ill-posed problems (discontinuous solution operators) is impossible without taking into account the impact of noise in the information. Commonly used models consider bounded deterministic noise and unbounded stochastic (Gaussian white) noise. It is common belief that white noise makes ill-posed problems more complex than problems under bounded noise. In this study we shed light on a rigorous complexity analysis of ill-posed problems providing (tight) lower and upper bounds for both noise models. It will be shown that in contrast to the deterministic case statistical ill-posed problems have finite complexity at every prescribed error level. Moreover, the ill-posedness of the problem raises the issue of adaptation to unknown solution smoothness, and we provide results in this direction.

1. Introduction

Information-based complexity is concerned with complexity issues of numerical problems given in terms of solution operators. We highlight this as

$$S \colon Y \longrightarrow X, \tag{1}$$

and we confine here to linear solution operators acting between Hilbert spaces Y and X. In order to tackle complexity, we need to specify some input class $\mathcal{G} \subset Y$, typically a unit ball of some (compactly) embedded space $Y_1 \hookrightarrow Y$. We recommend the seminal monograph [23] for details.

The theory is well developed for continuous linear solution operators. It was thus interesting to discuss whether the theory extends to unbounded linear operators. This problem was treated in the influential paper by A. Werschulz [25] in 1987, surveyed in [24]. In that study the problem

$$S \colon \mathcal{D}om(S) \subset Y \longrightarrow X, \tag{2}$$

Date: March 1, 2016.

for some unbounded linear mapping is considered, and the class $\mathcal{G}$ of instances is

$$\mathcal{G} = \{y \in \mathcal{D}om(S),\ \|y\|_Y \le 1\}.$$

Information is assumed to be any continuous linear functional. The major observation is then (Theorems 2.1, 2.2, ibid.) that the complexity is infinite, regardless whether the functionals are chosen nonadaptively or adaptively. This principal result received considerable attention.

The situation changes drastically if we impose additional knowledge on the solution element $x = S(y)$, for instance that it belongs to some compact subset $\mathcal{F} \subset X$. Actually, compactness is not necessary. This was also addressed in [24, p. 273] by arguing that "...it appears to be necessary to make additional a priori assumptions on the solution". For compact subsets $\mathcal{F}$ Tikhonov's Theorem [22] states that the ill-posedness of the original problem (2) will be *conditionally well-posed.* If this is the case then $S|_{S^{-1}(\mathcal{F})}$ is a continuous linear operator.

Remark 1. *Werschulz' results were also presented in the monograph* [23, Chapt. 5.7]. *In a discussion paper* [20] *Th. Seidman tried to explain how to remedy the infinite complexity of ill-posed equations, by turning to different concepts of solutions, topologies,... In another paper* [21, § 2] *the importance of* auxiliary information*, in particular in the light of Tikohonov's theorem is highlighted.*

For statistical inverse problems the present authors explicitly discussed complexity issues in [14].

If we have conditional well-posedness then it is natural to consider the *modulus of continuity*, say $\tilde{\omega}$ for a moment, of the mapping $S|_{S^{-1}(\mathcal{F})}$, given as a function of $\delta > 0$ as

$$\tilde{\omega}_{S^{-1}(\mathcal{F})}(\delta) := \sup\left\{\|Sy - Sy'\|_X,\ y, y' \in S^{-1}(\mathcal{F}),\ \|y - y'\|_Y \le \delta\right\}.$$

If we now introduce the operator $T = S^{-1}\colon X \to Y$, then the modulus rewrites as

$$\tilde{\omega}_{\mathcal{F}}(\delta) = \sup\left\{\|x - x'\|_X,\ x, x' \in \mathcal{F},\ \|Tx - Tx'\|_Y \le \delta\right\}.$$

The continuity of the restricted mapping $S|_{S^{-1}(\mathcal{F})}$ yields that the modulus tends to zero as $\delta \to 0$. For balanced convex sets $\mathcal{F}$ this modulus reduces to

$$\omega_{\mathcal{F}}(\delta) := \sup\left\{\|x\|_X,\ x \in \mathcal{F},\ \|Tx\|_Y \le \delta\right\}, \quad \delta > 0. \tag{3}$$

Here we shall consider sets $\mathcal{F} \subset X$ which are *ellipsoids*, i.e., the image of a unit ball under some self-adjoint positive operator $G\colon X \to X$, hence $\mathcal{F} = \{Gv,\ \|v\| \le 1\}$. Furthermore, we need to specify the error criterion for any reconstruction, say $\hat{S}\colon Y \to X$.

The fundamental importance of this function ω is given as the following

Fact (see [6, Thm. 4.4]). *Suppose that $\mathcal{F}$ is an ellipsoid. For each $\delta > 0$ there is a bounded linear reconstruction $\hat{S}\colon Y \to X$ with the following property. If we are given noisy data*

$$y^{\delta} = Tx + \delta\xi, \tag{4}$$

where δ denotes the noise level, $\|\xi\| \leq 1$ is arbitrary, then

$$\mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(\hat{S}) = \omega_{\mathcal{F}}(\delta),$$

where the error is the uniform error, given as

$$\mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(\hat{S}) := \sup_{x\in\mathcal{F}} \sup_{\left\|Tx-y^{\delta}\right\|_Y \leq \delta} \left\| x - \hat{S}(y^{\delta})\right\|_X . \tag{5}$$

This result is based on the seminal paper [17]. The reconstruction $\hat{S}$ is given explicitly in terms of some *Tikhonov regularization.* From this result we draw the conclusion that the natural setup of unbounded numerical problems is given as in (4), and hence this paradigm is taken here. In addition we specify the bounded linear operator $T\colon X \to Y$ to be injective, and compact.

Although the above fact is important it does not touch the question whether such optimal reconstruction can be attained by some finite amount of linear continuous data, thus we allow for *continuous linear information*

$$N_n(y^{\delta}) = \left(L_1(y^{\delta}), \dots, L_n(y^{\delta})\right) \in \mathbb{R}^n,$$

i.e., the functionals $L_1, \dots, L_n$ are continuous linear functionals on Y. Any $\psi\colon \mathbb{R}^n \to X$ may be used to design an *algorithm*

$$x_n^{\delta} := A_n(y^{\delta}) = \psi(N_n(y^{\delta})) \in X.$$

It is the objective of the present study to discuss the information complexity of such *linear ill-posed problems* as in (4) with noisy data. Extending the above derivation, we shall also discuss the case of *statistical inverse problems*, i.e., when ξ is Gaussian white noise. In the latter case the error criterion must be modified, and we use the root mean squared error given as

$$\mathrm{e}^{\mathrm{stat}}_{\mathcal{F},\delta}(A_n) := \sup_{x\in\mathcal{F}} \left(\mathbb{E}\left\|x - A_n(y^{\delta})\right\|_X^2\right)^{1/2}. \tag{6}$$

Above, the expectation is with respect to the sampling distribution P^x. Notice that the sampling distribution is improper, but by its very definition the application of a linear functional, say $b \in Y'$ to the noise ξ results in a univariate centered Gaussian random variable with variance $\|b\|_Y^2$. In order to have the expectation well defined we need to assume that the mapping ψ is Borel measurable. The latter is automatically fulfilled when considering projection schemes, as studied below.

Remark 2. *The error analysis for deterministic and statistical inverse problems is well established, starting from monograph* [9] *in 1962 for deterministic problems. Within a statistical setup we mention the studies* [2]*, and* [4] *from 1978. In a more general statistical framework the monograph* [7] *and the study* [19] *received considerable attention.*

The goal is to construct, given noisy data y^δ, a good approximation x^δ to x, based on a finite number of bounded linear functionals applied to y^δ. The fundamental question which we are going to address is: How much information from y^δ must be extracted in order to recover x with given accuracy, say $\varepsilon > 0$? We are going to discuss the latter in more detail by using *projection schemes*.

With the class of admissible information and algorithms specified, and with the above error criteria we can now introduce the following crucial function for *Information-based complexity*, see e.g. [23, Chapt. 4.4], where this was called ε-cardinality number.

Definition 1 (Information complexity)**.** *Suppose that we have fixed a class $\mathcal{F}$, and level $\delta > 0$. Given $\varepsilon > 0$ we assign*

$$\operatorname{comp}^{\circ}_{\mathcal{F},\delta}(\varepsilon) := \min\left\{n, \quad \exists A_n \text{ with } \mathrm{e}^{\circ}_{\mathcal{F},\delta}(A_n) \le \varepsilon\right\}, \quad \circ \in \{\det, \mathrm{stat}\}.$$

In addition we shall discuss the complexity when the class $\mathcal{F}$, sometimes called smoothness class, is unknown. This is the typical case for inverse problems. Here we ask whether it is possible to device an algorithm which yields a reconstruction for given error level ε, but without knowing the class $\mathcal{F}$. This is the subject of *a posteriori parameter choice* in inverse problems, and it is one of the most important tasks. Also, we shall address the problem, whether there is a minimal amount of information necessary to provide some a posteriori parameter choice (discretization level).

We are going to bound the information complexity, for both the deterministic and statistical inverse problems, starting with lower bounds in Section 2, and describing specific projection schemes which (up to some constants) attain the lower bounds, in section 3. It will be shown in Section 4 that in many cases the complexity results are sharp in order, see Theorem 2, Theorem 3, and the examples, given there.

In Section 5 we discuss the information complexity when smoothness is unknown.

2. Lower bounds

2.1. **Deterministic problems.**

In order to understand the *information complexity* we must first study the *uniform error*, and we start with the deterministic problem. Here the key quantity is the modulus of continuity from (3). We give the following result.

Fact (see [9], and [1, Prop. 2.6])**.**

(1) *For a* balanced convex *set* $\mathcal{F}$ *we have for every reconstruction* A_n *that* $\mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(A_n) \geq \omega_{\mathcal{F}}(\delta)$.
(2) *Suppose that* $\mathcal{F}$ *is balanced and convex. If for some* $\delta_0 > 0$ *we have that* $\omega_{\mathcal{F}}(\delta_0) = 0$ *then* $\omega_{\mathcal{F}}(\delta) = 0$ *for every* $\delta > 0$.

For a balanced convex set $\mathcal{F}$ the latter is the case if and only if $\mathcal{F} = \{0\}$, since the operator is assumed to be injective.

Proposition 1. *Suppose that* $\mathcal{F} \neq \{0\}$ *is balanced and convex. For each* $\delta > 0$ *there is* $\varepsilon_0 > 0$ *such that* $\mathrm{comp}^{\mathrm{det}}_{\mathcal{F},\delta}(\varepsilon) = \infty$ *for* $\varepsilon \leq \varepsilon_0$.

Summarizing, for ellipsoids in Hilbert space there will be some positive $\varepsilon_0 > 0$, and the information complexity will be infinite whenever $\varepsilon \leq \varepsilon_0$.

Next we provide another bound, which depends on the cardinality n of the information used in A_n. To this end we recall the notion of a *Gelfand width* of a set $\mathcal{F} \subset X$, see e.g. [18].

Definition 2 (nth Gelfand width)**.** *For a balanced convex subset* $\mathcal{F} \subset X$ *its nth Gelfand width is given as*

$$c_n(\mathcal{F}, X) := \inf_{\mathrm{codim}(M) \leq n} \sup_{x \in \mathcal{F} \cap M} \|x\|_X ,$$

where the infimum is over all n-codimensional subspaces $M \subset X$.

It is well known from Information-based complexity (for bounded solution operators) that the nth Gelfand width provides a lower bound for the error of any algorithm that uses n linear functionals as information, see [23, Chapt. 5.4]. Here we give a similar result for the unbounded solution operator, following the lines of [3, Thm. 10.4], where the statement is actually extended to adaptively chosen functionals $L_1, \dots, L_n$.

Theorem 1. *Fix a balanced convex set* $\mathcal{F}$. *Let* A_n *be any algorithm that uses some linear information* N_n. *Then*

$$\mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(A_n) \geq c_n(\mathcal{F}, X).$$

Proof. We shall actually prove (the stronger) version for noise level $\delta = 0$, i.e., we have exact data $y = Tx$. Fix any linear information N_n, and assign the (at most) n codimensional space $M_n := \ker(N_n \circ T)$. This gives

$$c_n(\mathcal{F}, X) \leq \sup_{x \in \mathcal{F} \cap M_n} \|x\|_X . \tag{7}$$

Then for any $x \in \mathcal{F} \cap M_n$ and any mapping $\psi \colon \mathbb{R}^n \to X$ we have that

$$\|x - \psi(0)\|_X = \|x - \psi(N_n(Tx))\|_X \leq \mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(A_n)$$

Since the set $\mathcal{F}$ is assumed to be balanced we also see that

$$\|-x - \psi(0)\|_X = \|-x - \psi(N_n(T(-x)))\|_X \leq \mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(A_n)$$

This yields for every such $x \in \mathcal{F} \cap M_n$ that

$$\begin{aligned} \|x\|_X &= \left\| \frac{1}{2}\left(x - \psi(0)\right) - \frac{1}{2}\left(-x - \psi(0)\right) \right\|_X \\ &\leq \frac{1}{2}\|x - \psi(0)\|_X + \frac{1}{2}\|-x - \psi(0)\|_X \leq \mathrm{e}^{\mathrm{det}}_{\mathcal{F},\delta}(A_n). \end{aligned}$$

From (7) we can complete the proof. □

Remark 3. *Notice that this bound is not sharp for large n. For large n the lower bound through the modulus of continuity from Proposition 1 may be larger.*

Next we specify the above lower bounds for specific ellipsoids, called source sets. These source sets are based on the concept of an *index function*, by which we mean a continuous non-decreasing function $\varphi\colon [0, \|T^*T\|_{\mathcal{L}(X)}] \to \mathbb{R}^+$ which satisfies $\varphi(0) = 0$.

Definition 3 (source set)**.** *For a given index function φ we assign*

$$T_\varphi := \left\{x, \quad x = \varphi(T^*T)v, \|v\|_X \leq 1\right\}.$$

Recall that the operator T admits a singular value decomposition.

Definition 4 (Singular value decomposition)**.** *There are two orthonormal bases $u_1, u_2, \dots \in X$ and $v_1, v_2, \dots \in Y$ and a non-increasing sequence $s_1 \geq s_2 \geq \dots \geq 0$ of singular numbers such that*

$$Tx = \sum_{j=1}^{\infty} s_j \langle x, u_j \rangle v_j, \quad x \in X.$$

For source sets in the Hilbert space X the following holds true.

Lemma 1. *We have that $c_n(T_\varphi, X) = \varphi(s_{n+1}^2)$.*

Proof. This can be seen from the following facts, cf. [18]. Since we consider ellipsoids in Hilbert space the Gelfand widths coincide with the linear widths. Furthermore, since T_φ is the image of the unit ball in X under the operator $\varphi(T^*T)$ the nth linear width coincides with the $(n+1)$st approximation number, which in turn equals the $(n+1)$st eigenvalue of the operator $\varphi(T^*T)$, which is $\varphi(s_{n+1}^2)$. □

We now turn to the lower bound on the deterministic errors given by the modulus of continuity, and as captured in ε_0 from Proposition 1. An explicit representation was given first in [8]. This was then reformulated within the context of general source sets in [5]. Here, we recall the formulation from [15]. Given an index function φ we shall introduce the companion function

$$\Theta(t) := \sqrt{t}\varphi(t), \quad t > 0. \tag{8}$$

Fact (see [15, Thm. 1]). *Let φ be any index function. Then*

$$\omega_{T_\varphi}(\Theta(s_j^2)) \geq \varphi(s_j^2),\ j = 1, 2, \dots.$$

Moreover, if the function $t \to \varphi^2((\Theta^2)^{-1}(t))$ is concave, then

$$\omega_{T_\varphi}^2(\delta) = s(\delta^2), \quad 0 < \delta \leq \|T\|_{\mathcal{L}(X,Y)}, \tag{9}$$

where s is a piece-wise linear spline, interpolating

$$s(\Theta^2(s_j^2)) = \varphi^2(s_j^2), \quad j = 1, 2, \dots. \tag{10}$$

As a consequence,

$$\omega_{T_\varphi}(\delta) \leq \varphi(\Theta^{-1}(\delta)), \quad 0 < \delta \leq \|T\|_{\mathcal{L}(X,Y)}. \tag{11}$$

Therefore a lower bound for the modulus is obtained by letting

$$\bar{n} = \bar{n}(\varphi, \delta) := \max\left\{j, \quad \Theta(s_j^2) > \delta\right\}, \tag{12}$$

since then we have that $\omega_{T_\varphi}(\delta) \geq \varphi(s_{\bar{n}+1}^2)$.

We turn to describe a cardinality, called n_ε which will become significant, both for deterministic and statistical problems.

To this end we require that $0 < \delta \leq \Theta(s_1^2)$. In this range we have the spline representation (10).

We fix the noise level δ. Then for $\bar{n}$ given by (12) we have

$$\Theta(s_{\bar{n}}^2) > \delta \geq \Theta(s_{\bar{n}+1}^2), \tag{13}$$

and the corresponding minimal modulus

$$\varepsilon_{\min} := \varphi(s_{\bar{n}+1}^2) \quad (\leq \varphi(\Theta^{-1}(\delta))). \tag{14}$$

Within the range $\varepsilon_{\min} < \varepsilon \leq \varphi(s_1^2)$ of error requirements we can give bounds for the information complexity. For ε in this range let

$$n_\varepsilon := \max\left\{j, \quad \varphi(s_j^2) \geq \varepsilon\right\}. \tag{15}$$

By Theorem 1 and Lemma 1 we see that $\mathrm{comp}^{\det}_{T_\varphi,\delta}(\varepsilon) \geq n_\varepsilon$.

Remark 4. *It is immediate from the definitions that $1 \leq n_\varepsilon \leq \bar{n}$. Indeed, we see that*

$$\varphi(s_{\bar{n}+1}^2) < \varepsilon \leq \varphi(s_{n_\varepsilon}^2),$$

from which we conclude that $n_\varepsilon < \bar{n} + 1$, and hence that $n_\varepsilon \leq \bar{n}$.

2.2. Statistical problems.

In the most general setup we choose an orthonormal system $b_1, b_2, \dots, b_k \in Y$ and consider the observation model

$$Y_i = L_i(y^\delta) = \langle Tx, b_i\rangle + \delta\xi_i \quad i = 1, 2, \dots, k, \tag{16}$$

where ξ_i are i.i.d. standard Gaussian random variables

At first we assume to have observations (16) without repetitions, which means that each of the functionals $L_i(y^\delta) = \left\langle y^\delta, b_i\right\rangle, i = 1, 2, \dots, k$ is observed only once. Turning from complete noisy data y^δ to (16) means that instead of (4) we turn to a projection scheme

$$Q_k y^\delta = Q_k Tx + \delta Q_k \xi, \tag{17}$$

where Q_k is the orthogonal projection onto $\operatorname{span}\{b_i, i = 1, 2, \dots, k\}$. If we denote by $a_1, a_2, \dots, a_k$ the orthonormal basis of $\operatorname{span}\{T^* b_i, i = 1, 2, \dots, k\}$ (with the orthogonal projection P_k), then general projection algorithm corresponding to (17) can be written as

$$x_k^\delta = A_k(y^\delta) = (Q_k T)^\dagger Q_k y^\delta = (Q_k T P_k)^\dagger Q_k y^\delta,$$

and

$$\mathbb{E}\left\| x - x_k^\delta \right\|_X^2 = \left\| (I - (Q_k T P_k)^\dagger Q_k T) x \right\|_X^2 + \delta^2 \mathbb{E}\left\| (Q_k T P_k)^\dagger \xi \right\|_X^2 . \tag{18}$$

It is clear that

$$\sup_{x \in T_\varphi} \left\| (I - (Q_k T P_k)^\dagger Q_k T) x \right\|_X^2 \geq \left\| (I - P_k)\varphi(T^*T) \right\|_X^2 \geq \varphi^2(s_{k+1}^2), \tag{19}$$

and

$$\mathbb{E}\left\| (Q_k T P_k)^\dagger \xi \right\|_X^2 = \sum_{j=1}^{k} s_j^{-2}(Q_k T P_k),$$

where $s_j(Q_k T P_k)$ denotes the j-th singular number of the operator $Q_k T P_k$. Moreover, we have that

$$s_j(Q_k T P_k) \leq \|Q_k\| \, s_j(T) \, \|P_k\| = s_j,$$

and therefore it holds

$$\left[\mathrm{e}^{\mathrm{stat}}_{T_\varphi,\delta}(A_k) \right]^2 \geq \varphi^2(s_{k+1}^2) + \delta^2 \sum_{j=1}^{k} s_j^{-2}. \tag{20}$$

It is easy to see that this lower bound is attained when Q_k, and consequently P_k, is onto initial segment of the singular elements of T (see Definition 4).

Therefore, if ε is smaller than the minimum of the right-hand side of (20) over k, then within the observation model (16) without repetitions in retrieving data, the complexity $\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon)$ is infinite. Otherwise it is finite and depends on the interplay between smoothness index function φ and the decay of singular values s_j that will be discussed in Examples 1 and 2 below.

Once we allow for repetition in retrieving data then there is no non-trivial lower bound for the reconstruction error $\mathrm{e}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon)$. Specifically, if we agree to observe $L_i(y^\delta)$ m_i times, with multiplicity m_i to be determined later, then our observation model reads as

$$Y_{i,j} := \langle Tx, b_i \rangle + \delta \xi_{i,j}, \quad j = 1, \dots, m_i, \; i = 1, \dots, k, \tag{21}$$

where the noise elements $\xi_{i,j}$ are i.i.d. standard Gaussian random variables. We shall first see that this new observation model is equivalent to the original model (without repetitions), but with heteroscedastic noise levels.

Lemma 2. *The projection scheme with repeated observations as in (21) provides the same least-squares solution as the one without repetitions*

$$\bar{Y}_i = \langle Tx, b_i\rangle + \delta\xi_{i,\bullet}, \quad i = 1,\dots,k, \tag{22}$$

where $\xi_{i,\bullet} := \frac{1}{m_i}\sum_{j=1}^{m_i}\xi_{i,j}$*. For i.i.d. Gaussian variables* $\xi_{i,j}$ *the related random variables* $\xi_{i,\bullet}$*,* $i = 1,\dots,k$ *are also independent Gaussian, but with varying variances* m_i^{-1}*.*

Proof. Expanding $x = \sum_{l=1}^k x_l a_l$ we arrive at a linear system of $n := \sum_{i=1}^k m_i$ equations for k variables $(x_1,\dots,x_k)$, with $n \geq k$, and we head for the *Moore–Penrose inverse* of the following matrix equation

$$\begin{pmatrix} \langle Ta_1, b_1\rangle, \dots, \langle Ta_k, b_1\rangle \\ \dots \\ \langle Ta_1, b_1\rangle, \dots, \langle Ta_k, b_1\rangle \\ \langle Ta_1, b_2\rangle, \dots, \langle Ta_k, b_2\rangle \\ \dots \\ \langle Ta_1, b_2\rangle, \dots, \langle Ta_k, b_2\rangle \\ \dots \\ \langle Ta_1, b_k\rangle, \dots, \langle Ta_k, b_k\rangle \\ \dots \\ \langle Ta_1, b_k\rangle, \dots, \langle Ta_k, b_k\rangle \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{pmatrix} = \begin{pmatrix} Y_{1,1} \\ \dots \\ Y_{1,m_1} \\ Y_{2,1} \\ \dots \\ Y_{2,m_2} \\ \dots \\ Y_{k,1} \\ \dots \\ Y_{k,m_k} \end{pmatrix}.$$

We apply Gauss elimination as follows: Suppose that $m_1 > 1$, then we add the identical rows one to $m_1 - 1$ to the m_1st row, and divide the latter by m_1 to end up in the original m_1st row. Then we subtract this m_1st row from the previous ones to obtain $m_1 - 1$ zero rows. The right-hand side changes as follows. We introduce the averages

$$Y_{i,\bullet} := \frac{1}{m_i}\sum_{j=1}^{m_i} Y_{i,j} = \langle Tx, b_i\rangle + \frac{\delta}{m_i}\sum_{j=1}^{m_i}\xi_{i,j}, \quad i = 1,\dots,k.$$

The corresponding (unknown) noise is thus obtained as $\xi_{i,\bullet}$. Then the first m_1 entries on the right are $Y_{1,1} - Y_{1,\bullet}, Y_{1,2} - Y_{1,\bullet}, \dots, Y_{1,\bullet}$. Thus we got

$$\begin{pmatrix} 0 \\ \dots \\ \langle Ta_1, b_1\rangle, \dots, \langle Ta_k, b_1\rangle \\ \langle Ta_1, b_2\rangle, \dots, \langle Ta_k, b_2\rangle \\ \dots \\ \langle Ta_1, b_2\rangle, \dots, \langle Ta_k, b_2\rangle \\ \dots \\ \langle Ta_1, b_k\rangle, \dots, \langle Ta_k, b_k\rangle \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{pmatrix} = \begin{pmatrix} Y_{1,1} - Y_{i,\bullet} \\ \dots \\ Y_{1,\bullet} \\ Y_{2,1} \\ \dots \\ Y_{2,m_2} \\ \dots \\ Y_{k,m_k} \end{pmatrix}.$$

Repeating this for the other multiple observations, and reordering respectively, we end up with a system

$$\begin{pmatrix} 0 \\ \dots \\ \langle Ta_1, b_1\rangle, \dots, \langle Ta_k, b_1\rangle \\ \dots \\ \langle Ta_1, b_k\rangle, \dots, \langle Ta_k, b_k\rangle \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{pmatrix} = \begin{pmatrix} Y_{1,1} - Y_{1,\bullet} \\ \dots \\ Y_{1,\bullet} \\ \dots \\ Y_{k,\bullet} \end{pmatrix}.$$

Applying the transpose from the left the zero rows yield that the resulting system is as if there were no repetitions

$$\begin{pmatrix} \langle Ta_1, b_1\rangle, \dots, \langle Ta_k, b_1\rangle \\ \langle Ta_1, b_2\rangle, \dots, \langle Ta_k, b_2\rangle \\ \dots \\ \langle Ta_1, b_k\rangle, \dots, \langle Ta_k, b_k\rangle \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{pmatrix} = \begin{pmatrix} Y_{1,\bullet} \\ \dots \\ Y_{k,\bullet} \end{pmatrix},$$

however with averaged observations $\bar{Y}_i = Y_{i,\bullet}$, $i = 1, \dots, k$, which completes the proof. □

Lemma 2 is important for controlling the statistical error within the observation model with repeatedly retrieved data (21). Recall that in the case of the model without repetitions (16) the lower bound of the reconstruction error (20) is attained when the algorithm A_k is formed by the projections P_k, Q_k onto the initial segments of singular elements. To obtain upper bounds for the model with repetitions, we restrict ourselves to such projections only, and pre-whiten the data (22) by introducing the operator $D_k\colon Y \to Y$

$$D_k y = \sum_{j=1}^{k} \sqrt{m_j} \langle y, v_j\rangle v_j, \quad y \in Y.$$

Then (22) is equivalent to the standard homoscedastic case, but with modified operator $\bar{T} := D_k T$, instead.

From Definition 4 it is clear that $s_j(\bar{T}) = \sqrt{m_j} s_j$. Then

$$V_k := \delta^2 \mathbb{E} \left\| \left(Q_k \bar{T} P_k\right)^{\dagger} \xi \right\|_X^2 = \delta^2 \sum_{j=1}^{k} \frac{1}{s_j^2(\bar{T})} = \delta^2 \sum_{j=1}^{k} \frac{1}{s_j^2 m_j}.$$

Also, we observe the obvious fact that by increasing the number of repetitions we can decrease the variance to arbitrarily low level.The same argument as for (20) shows that for the observation model (21) it holds

$$\left[e^{\text{stat}}_{T_\varphi,\delta}(A_k) \right]^2 = \varphi^2(s_{k+1}^2) + \delta^2 \sum_{j=1}^{k} \frac{1}{s_j^2 m_j}. \tag{23}$$

From the point of view of information-based complexity the following question is important: How to choose the numbers $m_1, \dots, m_k$ of repetitions in order to achieve a given level, say $2\varepsilon^2$ with minimal amount of information $n := \sum_{j=1}^{k} m_j$. Clearly, from (23) we see that $\varphi(s_{k+1}^2) \leq \varepsilon$, and hence that $k = n_\varepsilon$ from (15). Having fixed this number, we arrive at the following optimization problem.

$$\sum_{j=1}^{n_\varepsilon} m_j \longrightarrow \text{MIN!} \quad \text{subject to } \varepsilon^2 \geq V_{n_\varepsilon}. \tag{24}$$

This is an convex integer program with linear objective function on a convex domain

$$D_\varepsilon := \left\{ (m_1, \ldots, m_{n_\varepsilon}) \in \mathbb{N}^{n_\varepsilon}, \quad \sum_{j=1}^{n_\varepsilon} \frac{1}{s_j^2 m_j} \leq \frac{\varepsilon^2}{\delta^2} \right\}.$$

The following auxiliary result is useful.

Lemma 3. *Given non-decreasing* $0 < a_1 \leq a_2 \leq \cdots \leq a_k$, *we consider the integer program*

$$\sum_{j=1}^{k} m_j \longrightarrow \textit{MIN!} \quad \textit{subject to } \sum_{j=1}^{k} \frac{a_j^2}{m_j} \leq 1.$$

This problem has a solution $m^{\text{opt}} = \left(m_1^{\text{opt}}, \ldots, m_k^{\text{opt}}\right)$ *with non-decreasing integers* $m_1^{\text{opt}} \leq m_2^{\text{opt}} \leq \cdots \leq m_k^{\text{opt}}$, *and the objective function obeys*

$$\left(\sum_{j=1}^{k} a_j\right)^2 \leq \sum_{j=1}^{k} m_j^{\text{opt}} \leq k + \left(\sum_{j=1}^{k} a_j\right)^2. \tag{25}$$

Proof. The point $m_j := \lceil k a_j^2 \rceil$ is a feasible point, hence a solution exists. Also, without loss of generality we may assume this to be non-decreasing. Suppose that this is not the case, and that there is some index j with $m_j > m_{j+1}$. But then, since $a_j^2 \leq a_{j+1}^2$, we see that

$$\frac{a_j^2}{m_{j+1}} + \frac{a_{j+1}^2}{m_j} \leq \frac{a_j^2}{m_j} + \frac{a_{j+1}^2}{m_{j+1}},$$

such that we can interchange the integers without violating the constraints. By repeating this we can ensure the numbers $m_1, \ldots, m_k$ to be ordered.

We use *Lagrange relaxation* and solve the non-integer program by using Lagrange multipliers. This gives the (non-integer) solution

$$\tilde{m}_j := a_j \sum_{i=1}^{k} a_i, \quad j = 1, \ldots, k,$$

and its value gives a lower bound for the minimum. On the other hand, the integers $m_j := \lceil a_j \sum_{i=1}^{k} a_i \rceil$ provide us with an interior point, hence yield the upper bound as stated. □

We apply the above result with $k = n_\varepsilon$ and $a_j := \delta/(\varepsilon s_j)$, $j = 1, \dots, n_\varepsilon$, and corresponding repetition numbers $m_j = \lceil \frac{\delta^2}{\varepsilon^2} s_j^{-1} \sum_{i=1}^{n_\varepsilon} s_i^{-1} \rceil$, and we obtain a lower bound for the minimal number of required observations, say N_ε, given as

$$N_\varepsilon := \sum_{j=1}^{n_\varepsilon} m_j \geq \frac{\delta^2}{\varepsilon^2} \left(\sum_{j=1}^{n_\varepsilon} \frac{1}{s_j} \right)^2 . \tag{26}$$

3. Upper bounds with truncated SVD

Truncated SVD uses the orthogonal projections P_n, Q_n onto the first n components of the singular system of the operator T, cf. Definition 4. Recall that due to the special nature of these systems we have that $Q_n T = T P_n$.

Definition 5 (truncated singular value decomposition (TSVD))**.** *Given data y^δ we assign*

$$A_n(y^\delta) = \sum_{j=1}^{n} \frac{1}{s_j} \langle y^\delta, v_j \rangle u_j. \tag{27}$$

For each realization ξ of the noise this gives the following representation for the error

$$x - A_n(y^\delta) = (I - P_n)x + \sum_{j=1}^{n} \frac{1}{s_j} \langle \xi, v_j \rangle u_j. \tag{28}$$

Notice that the error in (28) is finite (almost surely) even for statistical noise, because only a finite amount of functionals is evaluated. The representation (28) yields that

$$\left\| x - A_n(y^\delta) \right\|_X^2 = \left\| (I - P_n)x \right\|_X^2 + \delta^2 \sum_{j=1}^{n} \frac{\left| \langle \xi, v_j \rangle \right|^2}{s_j^2}.$$

Since the projections P_n are spectral we obtain for bounded deterministic noise that

$$\sup_{x \in T_\varphi} \left\| x - A_n(y^\delta) \right\|_X^2 \leq \varphi^2(s_{n+1}^2) + \frac{\delta^2}{s_n^2}. \tag{29}$$

We apply this for n_ε from (15), where we also have

$$\varphi(s_{n_\varepsilon+1}^2) < \varepsilon \leq \varphi(s_{n_\varepsilon}^2)$$

such that

$$\left[\mathrm{e}_{T_\varphi,\delta}^{\mathrm{det}}(A_{n_\varepsilon}) \right]^2 \leq \varepsilon^2 + \frac{\delta^2}{\varphi^{-1}(\varepsilon)}.$$

On the other hand, the assumption that $\varepsilon \geq \varphi(\Theta^{-1}(\delta))$ can be written as $\delta \leq \Theta(\varphi^{-1}(\varepsilon))$. Then by definition of Θ we get

$$\left[\mathrm{e}_{T_\varphi,\delta}^{\mathrm{det}}(A_{n_\varepsilon}) \right]^2 \leq \varepsilon^2 + \frac{\Theta^2(\varphi^{-1}(\varepsilon))}{\varphi^{-1}(\varepsilon)} = \varepsilon^2 + \varphi^2(\varphi^{-1}(\varepsilon)) = 2\varepsilon^2. \tag{30}$$

We turn to statistical noise, where we have the error representation (23). This also gives, for $n := n_\varepsilon$ from (15), and by the construction of $m_1, \dots, m_{n_\varepsilon}$ as

$$m_j := \lceil \frac{\delta^2}{\varepsilon^2} s_j^{-1} \sum_{i=1}^{n_\varepsilon} s_i^{-1} \rceil \tag{31}$$

that

$$\sup_{x \in T_\varphi} \mathbb{E} \left\| x - A_{n_\varepsilon}(y^\delta) \right\|_X^2 = \varphi^2(s_{n_\varepsilon+1}^2) + \varepsilon^2 \le 2\varepsilon^2, \tag{32}$$

however with (up to) $N_\varepsilon \le n_\varepsilon + \frac{\delta^2}{\varepsilon^2} \left(\sum_{j=1}^{n_\varepsilon} s_j^{-1} \right)^2$ evaluations of functionals. There is a simple criterion to check whether repetitions are necessary or not. By looking at (23) we see that repetitions are necessary if $m_j = 1, j = 1, 2, \dots, n, \delta^2 \sum_{j=1}^{n} \frac{1}{m_j s_j^2} > \varepsilon^2$. Thus we get that

- either $\sum_{j=1}^{n_\varepsilon} s_j^{-2} \le \varepsilon^2/\delta^2$, and no repetitions are necessary (and $m_j = 1, j = 1, 2, \dots, n_\varepsilon, N_\varepsilon = n_\varepsilon$),
- or $\sum_{j=1}^{n_\varepsilon} s_j^{-2} > \varepsilon^2/\delta^2$, then the described choice of repetitions yields the best possible error, however, on expense of additional information ($N_\varepsilon > n_\varepsilon$).

Remark 5. *We make another interesting observation. An equal number of repetitions for each of the n_ε Fourier coefficients would require a number m of repetitions proportional to $m \propto \sum_{j=1}^{n_\varepsilon} 1/s_j^2$, resulting in an overall number $m \times n_\varepsilon$ observations. By virtue of the Cauchy–Schwarz Inequality this will be (typically much) larger than N_ε.*

4. Main results, examples and discussion

We summarize the above bounds, both for bounded deterministic and for statistical noise.

Theorem 2. *Let n_ε be as in (15). Then for $\varepsilon \ge \varphi(\Theta^{-1}(\delta))$ the information complexity under bounded deterministic noise obeys*

$$\mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta}(\sqrt{2}\varepsilon) \le n_\varepsilon \le \mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta}(\varepsilon).$$

Proof. The right inequality follows from Theorem 1 and Lemma 1, and the left on is a consequence of (30). □

Remark 6. *In view of (14) the assumption that $\varepsilon > \varphi(\Theta^{-1}(\delta))$ is not a real restriction in the context of deterministic problems.*

Theorem 3. *The information complexity under Gaussian white noise obyes*

$$[n_\varepsilon] \le \mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\sqrt{2}\varepsilon) \le n_\varepsilon + \frac{\delta^2}{\varepsilon^2} \left(\sum_{j=1}^{n_\varepsilon} s_j^{-1} \right)^2.$$

Proof. By using A_{n_ε}, with repetitions if necessary, from (23), (31)–(32) we obtain the upper bound. The lower bound follows from (19). $\square$

We shall now discuss the main results on the information complexity at some examples.

A solution smoothness, as described by Definition 3, allows a classification of ill-posed problems into several categories based upon the type of growth of the index function φ (see, e.g. [10]).

A problem (4) with $x \in T_\varphi$ is called moderately ill-posed problems if the function φ grows like a power, as for example, $\varphi(t) = t^q, q > 0$.

It is called severely ill-posed if φ growths logarithmically, as for example, $\varphi(t) = \ln^{-q}\frac{1}{t}, t \in (0, s_1^2], q > 0$, where we assume that $s_1 < 1$.

Usually, severely ill-posed problems are associated with an exponential decay of singular numbers s_j, while a power decay of s_j leads to a moderate ill-posedness.

The examples below demonstrate that for both cases of ill-posedness the bounds given by Theorems 2, and 3 are rather tight.

Example 1 (Severely ill-posed problems)**.** *Consider the case when for some positive constants* c_1, c_2 *and* p *we have* $c_1 e^{-pj} \leq s_j \leq c_2 e^{-pj}, j = 1, 2, \dots,$*, abbreviated as*

$$s_j \asymp e^{-pj}. \tag{33}$$

The classical (ill-posed) Cauchy problem for the Laplace equation in a rectangular domain can be reduced to (4) *with* T *satisfying* (33)*, see, e.g.* [10].

From (15), by the very definition, we have

$$n_\varepsilon = \frac{1}{2p} \ln \frac{1}{\varphi^{-1}(\varepsilon)} + O(1). \tag{34}$$

Assume that φ *is such that the function* $\ln \frac{1}{\varphi^{-1}(t)}$ *obeys a* Δ_2*-type condition, i.e for any* $c > 0$

$$\ln \frac{1}{\varphi^{-1}(ct)} \asymp \ln \frac{1}{\varphi^{-1}(t)}, \tag{35}$$

which is valid for the function $\varphi(t) = \ln^{-q}\frac{1}{t}$ *mentioned above.*

Then Theorem 2 together with (34), (35) *implies*

$$\mathrm{comp}^{\det}_{T_\varphi,\delta}(\varepsilon) \asymp \ln \frac{1}{\varphi^{-1}(\varepsilon)},$$

where $\varepsilon \geq \varphi(\Theta^{-1}(\delta))$*.*

Let us now consider the same problem under white noise. From (33) *and* (34) *we have*

$$\sum_{j=1}^{n_\varepsilon} s_j^{-1} \asymp e^{pn_\varepsilon} \asymp \frac{1}{\sqrt{\varphi^{-1}(\varepsilon)}}.$$

Then Theorem 3 implies that for some positive ε, δ-independent numbers d_1, d_2

$$d_1 \ln \frac{1}{\varphi^{-1}(\varepsilon)} \leq \mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) \leq d_2 \left(\ln \frac{1}{\varphi^{-1}(\varepsilon)} + \frac{\delta^2}{\Theta^2(\varphi^{-1}(\varepsilon))} \right).$$

It is interesting to observe that for $\varepsilon \geq \varphi(\Theta^{-1}(\delta))$ the latter bound gives us

$$\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) \asymp \mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta_0}(\varepsilon) \asymp \ln \frac{1}{\varphi^{-1}(\varepsilon)},$$

such that the deterministic and statistical complexities are of the same order for severely ill-posed problems.

Example 2 (Moderately ill-posed problems)**.** *Consider the case when*

$$s_j \asymp j^{-r}, \quad \textit{for some } r > 0. \tag{36}$$

The range of ill-posed problems described by the operators T with such singular numbers rises from numerical differentiation, where $r = 1$, to satellite gravity gradiometry, where one can take $r = 5.5$, cf. [11].

From (36) *and the definition of n_ε we get*

$$n_\varepsilon \asymp [\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}}. \tag{37}$$

If $\varphi(t)$ increases as a power of t, then its inverse obeys a Δ_2-type condition, such that for any $c > 0$ it holds

$$\varphi^{-1}(c\varepsilon) \asymp \varphi^{-1}(\varepsilon), \tag{38}$$

and from Theorem 2 we have

$$\mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta}(\varepsilon) \asymp [\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}}. \tag{39}$$

In case of white noise, (36)–(38) *give us*

$$\sum_{j=1}^{n_\varepsilon} s_j^{-1} \asymp n_\varepsilon^{r+1} \asymp [\varphi^{-1}(\varepsilon)]^{-\frac{r+1}{2r}},$$

and Theorem 3 implies that for some positive ε, δ-independent numbers d_1, d_2 we have

$$d_1[\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}} \leq \mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) \leq d_2 \left([\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}} + \frac{\delta^2}{\varepsilon^2}[\varphi^{-1}(\varepsilon)]^{-\frac{r+1}{r}} \right). \tag{40}$$

From (39) and (40) one can see that, in contrast to the case of severely ill-posed problems discussed in Example 1, for $\varepsilon \geq \varphi(\Theta^{-1}(\delta))$ the complexity of moderately ill-posed problems may have different order in the deterministic and stochastic settings. This is in agreement with the previous results [16], where the focus was on achieving the best possible accuracy under a given noise level δ, and no repetitions, as in (21) were allowed.

To compare (39), (40) with the results of that study we consider the function $\Theta_r(t) = \varphi(t)t^{1/2+1/4r}$, $t > 0$, and assume that φ belongs to the class of operator monotone functions, which includes, as a special case, the index functions $\varphi(t) = t^q, q \in (0,1]$. Then from [16] we find that under Assumption (36) it holds

$$\inf_n \mathrm{e}^{\mathrm{stat}}_{T_\varphi,\delta}(A_n) \asymp \varphi(\Theta_r^{-1}(\delta)), \tag{41}$$

which means that if no repetitions are allowed then in statistical setting no algorithm can achieve an accuracy of order better than $\varphi(\Theta_r^{-1}(\delta))$.

However, for $\varepsilon \geq \varphi(\Theta_r^{-1}(\delta))$ it holds

$$[\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}} \geq \frac{\delta^2}{\varepsilon^2}[\varphi^{-1}(\varepsilon)]^{-\frac{r+1}{r}} = \frac{\delta^2}{\Theta_r^2(\varphi^{-1}(\varepsilon))}[\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}},$$

such that the bounds (39) and (40) tell us that for $\varepsilon \geq \varphi(\Theta_r^{-1}(\delta)) > \varphi(\Theta^{-1}(\delta))$

$$\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) \asymp \mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta}(\varepsilon) \asymp [\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}}.$$

Put in words, this means that if in the statistical setting no repetitions (21) are necessary for achieving ε-accuracy then the ε-complexity has the same order as the one in the deterministic setting. On the other hand, if in the statistical setting the desired accuracy ε cannot be achieved without repetitions then deterministic and statistical ε-complexities differ by order.

In particular, if ε is such that $\varepsilon = o(\varphi(\Theta_r^{-1}(\delta)))$, but $\varepsilon \geq \varphi(\Theta^{-1}(\delta))$ then

$$\mathrm{comp}^{\mathrm{det}}_{T_\varphi,\delta_0}(\varepsilon) \asymp [\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}},$$

while within the observation model without repetitions

$$\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) = \infty,$$

showing a significantly different behavior for the information complexities.

At the same time, for the observation models with repetitions

$$\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) = \mathcal{O}\left(\frac{\delta^2}{\Theta_r^2(\varphi^{-1}(\varepsilon))}[\varphi^{-1}(\varepsilon)]^{-\frac{1}{2r}}\right).$$

5. Adaptation to unknown smoothness

To explain the problem of adaptation to unknown smoothness let us first consider the case of a well-posed solution operator S. Suppose for simplicity that $S\colon Y \to X$ is a positive compact diagonal operator with non-increasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \cdots \geq 0$. Suppose further that we do not know the smoothness of the problem instances, and still we want to achieve a given accuracy $\varepsilon > \delta$, where, again, δ is the noise level. In this case we may achieve the desired accuracy as follows. We let

$$N_\varepsilon := \min\{j,\ \lambda_j \le \varepsilon - \delta\}.$$

As before, information is assumed to be noisy coefficients with respect to the eigenbasis. We let A_{N_ε} be the truncated SVD from (27), but now with $s_j := \lambda_j^{-1}$. Then it is easy to check that this algorithm achieves the prescribed accuracy ε when measured in the space X, without any additional assumption on the problem instance $y \in Y$.

For ill-posed solution operators the situation is different, because the eigenvalues will be increasing, and the previous arguments do no longer apply. Therefore, in order to achieve a given accuracy we need to prescribe a certain rate of decay of the sequence of coefficients $\langle y, v_j\rangle$, as e.g. done in terms of source conditions, cf. Definition 3. Such smoothness typically will be unknown to us, and therefore adaptation to unknown smoothness in general cannot be avoided, once we want to have a desired accuracy.

5.1. **Lower bounds.** We first turn to some lower bound, and we therefore restrict the class, say Φ of admissible index function to those, which obey $\varphi(s_1^2) = 1$, for normalization. To specify the setting we agree upon the following

Definition 6 (minimal adpative discretization)**.** *Fix $\varepsilon_{\min} < \varepsilon < 1$. Let $n_{\text{ad}} = n_{\text{ad}}(\varepsilon)$ be the minimal cardinality such that for each $\varphi \in \Phi$ there is some reconstruction A_n with $\mathrm{e}^{\text{det}}_{T_\varphi,\delta}(A_n) \le \varepsilon$.*

We provide the following auxiliary result.

Lemma 4. *For each $0 < \varepsilon < 1$ and $0 < \delta < s_1^2$ there is an index function $\varphi \in \Phi$ with $\varphi(s_1^2) = 1$ and $\varphi(\delta^2) > \varepsilon$.*

Proof. Let us fix some $\bar\varepsilon \in (\varepsilon, 1)$ and consider the index function which is piece-wise linear from $\varphi(0) = 0$, $\varphi(\delta^2) = \bar\varepsilon$, and $\varphi(s_1^2) = 1$. It is readily seen that this function does the job. □

This result is used to prove the following lower bound. We shall use the number

$$n_\delta := \max\left\{j,\ \frac{\delta}{s_j} \le 1\right\} + 1. \tag{42}$$

Corollary 1. *The minimal adaptive discretization level n_{ad} must obey*

$$n_{\text{ad}} \ge n_\delta.$$

Proof. Suppose that $n_{\text{ad}} < n_\delta$. Then we find that $\delta/s_{n_{\text{ad}}+1} \le 1$, and hence that $s_{n_{\text{ad}}+1} \ge \delta$. But then we have for the index function constructed in the proof of Lemma 4 that

$$\varphi(s^2_{n_{\text{ad}}+1}) \ge \varphi(\delta^2) > \varepsilon.$$

By virtue of Theorem 1 and Lemma 1 we see that there is an index function $\varphi \in \Phi$ for which $\mathrm{e}^{\mathrm{det}}_{T_\varphi,\delta}(A_{n_{\mathrm{ad}}}) > \varepsilon$, regardless how $A_{n_{\mathrm{ad}}}$ was chosen. Thus, the minimal level n_{ad} must be larger than n_δ, which completes the proof. □

5.2. **Discrepancy principle with finitely many data.** We saw that the minimal amount of information must obey $s_{n_{\mathrm{ad}}-1} \leq \delta$. This result will be complemented by some explicit construction, which guarantees an error less than or equal to ε, provided that $n \geq n_{\mathrm{ad}}$. Actually, the following scheme (TSVD with discretized discrepancy principle) yields, up to some constant, the best possible accuracy, and hence is certainly below the required threshold.

We start discussing a version of the *discretized discrepancy principle* applied to the truncated singular value decomposition. For some fixed m, to be determined later, we introduce the following principle.

Definition 7 (discretized discrepancy principle (DDP)). *Given $\tau > 1$ we let n_{DDP} be the smallest n such that*

$$\left\| Q_m(Tx_n^\delta - y^\delta) \right\|_Y \leq \tau\delta.$$

The analysis of the DDP is similar to the one for the classical discrepancy principle, studied in regularization theory, and we follow the reasoning in [12].

Before doing so we observe that $n_{\mathrm{DDP}} \leq m$, since for $n = m$ we have that $Q_m(Tx_n^\delta - y^\delta) = Q_n(Q_n - I)y^\delta = 0$, such that stop cannot occur later.

Theorem 4. *Suppose that $x \in T_\varphi$ for some $\varphi \in \Phi$, and that we choose the discretization level according to the DDP with m satisfying $s_{m+1} \leq \delta$, and with $\tau = 2$. Then we have that*

$$\|x - x_{n_{\mathrm{DDP}}}\|_X \leq 4\varphi(\Theta^{-1}(\delta)), \quad 0 < \delta \leq 1.$$

Proof. The proof consists of several steps. First we discuss the case that we stop immediately, i.e., at $x_0^\delta = 0$. Then we must have that $\left\| Q_m y^\delta \right\|_Y \leq \tau\delta$. But this also shows that zero is an order optimal reconstruction. Indeed, we see that

$$\|x\|_X \leq \omega_{T_\varphi}((\tau + 2)\delta) \leq (\tau + 2)\omega_{T_\varphi}(\delta),$$

since

$$\begin{aligned} \|Tx\|_Y &\leq \|Q_m Tx\|_Y + \|(I - Q_m)Tx\|_Y \\ &\leq \left\| Q_m(Tx - y^\delta) \right\|_Y + \left\| Q_m y^\delta \right\|_Y + \|(I - Q_m)Tx\|_Y \\ &\leq (1 + \tau)\delta + \|x\|_X s_{m+1}. \end{aligned}$$

Since for $\varphi \in \Phi$ we have that $\|x\|_X \leq 1$ we conclude that $\|x\|_X s_{m+1} \leq \delta$, which show the optimality of the DDP at immediate stop, and for $\tau = 2$ with constant 4.

If $m \leq n \leq 1$ (no immediate stop) then we argue as follows. We start with the standard error decomposition for truncated SVD as

$$\left\|x_n^\delta - x\right\|_X \leq \|x_n - x\|_X + \delta/s_n, \tag{43}$$

where $x_n = P_n x$ denotes the noiseless reconstruction. We bound the bias $\|x_{n_{\mathrm{DDP}}} - x\|_X$ in terms of the modulus of continuity. Clearly, both x and $x_{n_{\mathrm{DDP}}}$ belong to T_φ. Moreover, we have

$$\begin{aligned} \|T(x_{n_{\mathrm{DDP}}} - x)\|_Y &= \|T(P_{n_{\mathrm{DDP}}} - I)x\|_Y \\ &\leq \|Q_m T(P_{n_{\mathrm{DDP}}} - I)x\|_Y + \|(I - Q_m)T(P_{n_{\mathrm{DDP}}} - I)x\|_Y \\ &= \|Q_m T(P_{n_{\mathrm{DDP}}} - I)x\|_Y + \|T(I - P_m)(P_{n_{\mathrm{DDP}}} - I)x\|_Y . \end{aligned} \tag{44}$$

The first summand can be bounded by using the DDP. We recall that in this case $\left\|Q_m(Tx_{n_{\mathrm{DDP}}} - y^\delta)\right\|_Y = \left\|Q_m(I - Q_n)y^\delta\right\|_Y \leq \tau\delta$, and consequently, that

$$\|Q_m T(P_{n_{\mathrm{DDP}}} - I)x\|_Y = \|Q_m(I - Q_n)Tx\|_Y \leq (\tau + 1)\delta.$$

The second summand in (44) will be bounded by observing that $(I - P_m)(P_{n_{\mathrm{DDP}}} - I) = P_m - I$. This gives, similarly to the immediate stop case, that

$$\|T(I - P_m)(P_{n_{\mathrm{DDP}}} - I)x\|_Y = \|T(I - P_m)x\|_Y \leq s_{m+1}\|x\|_X \leq \delta,$$

the latter by the choice of m and since $x \in T_\varphi$ for $\varphi \in \Phi$. Both the above bounds finally give that

$$\|T(x_{n_{\mathrm{DDP}}} - x)\|_Y \leq (\tau + 2)\delta,$$

which in turn yields

$$\|x_{n_{\mathrm{DDP}}} - x\|_X \leq \omega_{T_\varphi}((\tau + 2)\delta) \leq (\tau + 2)\omega_{T_\varphi}(\delta) \leq (\tau + 2)\varphi(\Theta^{-1}(\delta)).$$

It remains to bound $\delta/s_{n_{\mathrm{DDP}}}$. For $n < n_{\mathrm{DDP}}$ (before stopping) we have that $\tau\delta < \left\|Q_m(I - Q_n)y^\delta\right\|$. We use this information to derive

$$\begin{aligned} \tau\delta &< \left\|Q_m(I - Q_n)y^\delta\right\|_Y \leq \left\|Q_m(I - Q_n)(y^\delta - y)\right\|_Y + \|Q_m(I - Q_n)y\|_Y \\ &\leq \delta + \|(I - Q_n)Tx\|_Y \leq \delta + \Theta(s_{n+1}^2), \end{aligned}$$

since the element Tx obeys a source condition with the index function Θ. Thus we get that

$$(\tau - 1)\delta \leq \Theta(s_{n+1}^2).$$

In particular for $n := n_{\mathrm{DDP}} - 1$ we get $(\tau - 1)\delta \leq \Theta(s_{n_{\mathrm{DDP}}}^2)$, which in turn yields $s_{n_{\mathrm{DDP}}} \geq \Theta^{-1}((\tau - 1)\delta)$. Therefore,

$$\frac{\delta}{s_{n_{\mathrm{DDP}}}} \leq \frac{1}{\tau - 1}\frac{(\tau - 1)\delta}{\sqrt{\Theta^{-1}((\tau - 1)\delta)}} = \frac{1}{\tau - 1}\varphi(\Theta^{-1}((\tau - 1)\delta))$$

If we now specify $\tau := 2$ then we obtain the overall bound

$$\left\|x_{n_{\mathrm{DDP}}}^\delta - x\right\|_X \leq 3\varphi(\Theta^{-1}(\delta)) + \varphi(\Theta^{-1}(\delta)) = 4\varphi(\Theta^{-1}(\delta)),$$

which completes the proof. □

We now apply this result with $m := n_\delta$ from (42), for which we see that $s_{m+1} < \delta$.

Theorem 5. *The following assertions hold true.*

1. *The minimal discretization level n_{ad} must obey $n_{\mathrm{ad}} \geq n_\delta$, no matter which accuracy is required.*
2. *For the value $m := n_\delta$ the DDP (with $\tau = 2$) applied to truncated SVD provides us with a reconstruction of the error $4\varphi(\Theta^{-1}(\delta))$, thus for any error level $\varepsilon \geq 4\varphi(\Theta^{-1}(\delta))$.*

We add the following interesting result, which shows that for discretization levels $m < n_{\mathrm{ad}}$ the reconstruction with error less than ε is possible whenever the smoothness is given through some index function φ in a restricted subset.

Theorem 6. *Let $\varepsilon < 1$ be an error level, and let the discretization level m_ε be given as*

$$m_\varepsilon := \max\left\{ j, \quad \frac{\delta}{s_j} \leq \frac{\varepsilon}{\sqrt{2}} \right\}. \tag{45}$$

If the solution x obeys $x \in T_\varphi$ for some index function φ with

$$\varphi\left(\Theta^{-1}(\delta)\right) \leq \frac{\varepsilon}{\sqrt{2}}, \tag{46}$$

then we have for $x^\delta_{m_\varepsilon}$, the result of TSVD at level m_ε, the error bound

$$\left\| x - x^\delta_{m_\varepsilon} \right\|_X \leq \varepsilon. \tag{47}$$

Proof. Consider the following index $m_\varepsilon + 1$. This obeys, for an index function with (46) that

$$\frac{\delta}{s_{m_\varepsilon+1}} > \frac{\varepsilon}{\sqrt{2}} \geq \varphi\left(\Theta^{-1}(\delta)\right).$$

Now we use the following identity

$$\delta = \Theta\left(\Theta^{-1}(\delta)\right) = \varphi\left(\Theta^{-1}(\delta)\right)\sqrt{\Theta^{-1}(\delta)}$$

to derive that

$$\frac{\sqrt{\Theta^{-1}(\delta)}}{s_{m_\varepsilon+1}} > 1,$$

which in turn gives $s^2_{m_\varepsilon+1} < \Theta^{-1}(\delta)$, and hence that $\varphi\left(s^2_{m_\varepsilon+1}\right) \leq \varphi\left(\Theta^{-1}(\delta)\right) \leq \varepsilon/\sqrt{2}$. This bound and the choice of m_ε from (45) together with the error bound from (29) allow us to complete the proof. $\square$

5.3. **Discussion.** It turns out that in the deterministic worst case setting there are three levels of ε-complexity of the ill-posed problem (4) associated with a given operator T. These levels depend on our a priori knowledge about this problem. Namely, if we know only the noise level δ, then according to Corollary 1, the ε-complexity is measured by n_δ. If we know that for the problem (4) the level of accuracy ε is attainable, then according to Theorem 6, the ε-complexity is measured by m_ε. Finally, if we know that the solution of (4) belongs to T_φ, then according to Theorem 2, *the* ε-complexity is measured by n_ε.

In the statistical setting, and this can be seen from Theorem 3, the possibility of using repeated observations, and the knowledge of the solution smoothness, say φ yields that arbitrary level of accuracy can be attained, i.e., for every $\varepsilon > 0$, we have that $\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) < \infty$. On the other hand, from [16] we know that the amount n_δ of discrete information is enough to construct an algorithm, say A_n, $n \leq n_\delta$, which achieves the accuracy of order given in (41) without knowledge of the function φ.

Also, it is known from [13] that for any given solution element x there is neither minimal nor maximal smoothness when measured by some index function φ in terms of a source condition $x \in T_\varphi$. Thus, for any $\varepsilon \in (0,1)$, and arbitrarily large cardinality n of information one can construct some index function such that $\varphi(s^2_{n+1}) > \varepsilon$. In particular, cf. (32), this implies that the possibility to repeatedly retrieve information is not enough to guarantee a prescribed level ε without knowing the function φ. Precisely we claim that for every n there is some index function φ such that for any $\varepsilon = o\left(\varphi\left(\Theta_r^{-1}(\delta)\right)\right)$ we have that $\mathrm{comp}^{\mathrm{stat}}_{T_\varphi,\delta}(\varepsilon) > n$. Summarizing, without knowledge of the index function φ we cannot achieve arbitrary accuracy in the statistical setting by any means of adaptation, even if we allow for repetitions.

6. Acknowledgment

Part of this work was done while PM was visiting Fudan University, Shanghai. PM gratefully acknowledges the hospitality at Fudan, ans the support by the Humboldt foundation. SVP is partially supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF), grants P25424 and I1669.

References

1. Radu Boţ, Bernd Hofmann, and Peter Mathé, *Regularizability of ill-posed problems and the modulus of continuity*, Zeitschrift für Analysis und ihre Anwendungen. Journal of Analysis and its Applications **32** (2013), no. 3, 299–312.
2. A. M. Fedotov, *An estimate of the accuracy of approximate solutions of operator equations of the first kind with random errors in the original data*, Dokl. Akad. Nauk SSSR **240** (1978), no. 3, 545–548. MR 0501874

3. Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2013. MR 3100033
4. L. A. Halfin, *The statistical approach to the solution of ill-posed problems of geophysics*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **79** (1978), 67–81, 99, 103, Studies in statistical estimation theory, II. MR 545123
5. Markus Hegland, *Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization*, Appl. Anal. **59** (1995), no. 1-4, 207–223. MR 97a:65060
6. B. Hofmann, P. Mathé, and M. Schieck, *Modulus of continuity for conditionally stable ill-posed problems in Hilbert space*, J. Inverse Ill-Posed Probl. **16** (2008), no. 6, 567–585. MR MR2458286
7. I. A. Ibragimov and R. Z. Has′minskiĭ, *Statistical estimation*, Applications of Mathematics, vol. 16, Springer-Verlag, New York, 1981, Asymptotic theory, Translated from the Russian by Samuel Kotz. MR MR620321 (82g:62006)
8. V. Ivanov and T. Korolyuk, *Error estimates for solutions of incorrectly posed linear problems*, Zh. Vychisl. Mat. i Mat. Fiz. **9** (1969), 30–41 (Russian).
9. M. M. Lavrent′ev, *O nekotorykh nekorrektnykh zadachakh matematicheskoi fiziki*, Izdat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, 1962. MR MR0224319 (36 #7363)
10. Shuai Lu and Sergei V. Pereverzev, *Regularization theory for ill-posed problems*, Inverse and Ill-posed Problems Series, vol. 58, De Gruyter, Berlin, 2013, Selected topics. MR 3114700
11. ———, *Multiparameter regularization in downward continuation of satellite data*, Handbook of Geomathematics (Willi Freeden, M. Zuhair Nashed, and Thomas Sonar, eds.), Springer Berlin Heidelberg, 2014, pp. 1–20 (English).
12. Peter Mathé, *What do we learn from the discrepancy principle?*, Z. Anal. Anwend. **25** (2006), no. 4, 411–420. MR MR2285092
13. Peter Mathé and Bernd Hofmann, *How general are general source conditions?*, Inverse Problems **24** (2008), no. 1, 015009, 5. MR 2384768 (2009b:65144)
14. Peter Mathé and Sergei V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal. **38** (2001), no. 6, 1999–2021 (electronic). MR 1856240
15. ———, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems **19** (2003), no. 3, 789–803. MR MR1984890 (2004i:47021)
16. ———, *Regularization of some linear ill-posed problems with discretized random noisy data*, Math. Comp. **75** (2006), no. 256, 1913–1929. MR MR2240642
17. Avraham A. Melkman and Charles A. Micchelli, *Optimal estimation of linear operators in Hilbert spaces from inaccurate data*, SIAM J. Numer. Anal. **16** (1979), no. 1, 87–105. MR 80c:65119
18. Allan Pinkus, *n-widths in approximation theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 7, Springer-Verlag, Berlin, 1985. MR MR774404 (86k:41001)
19. M. S. Pinsker, *Optimal filtration of square-integrable signals in Gaussian noise.* MR 624591
20. Thomas I. Seidman, *Some "complexity" issues for ill-posed problems*, Inverse problems in partial differential equations (Arcata, CA, 1989), SIAM, Philadelphia, PA, 1990, pp. 17–35. MR 1046427 (91m:65313)
21. ———, *Interpretations and estimates for ill-posed problems*, Inverse problems in diffusion processes (Lake St. Wolfgang, 1994), SIAM, Philadelphia, PA, 1995, pp. 108–119. MR 1338625 (96c:65107)

22. A. N. Tikhonov, *On the stability of inverse problems*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **39** (1943), 176–179. MR 5,184e
23. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Computer Science and Scientific Computing, Academic Press Inc., Boston, MA, 1988, With contributions by A. G. Werschulz and T. Boult. MR MR958691 (90f:68085)
24. Arthur G. Werschulz, *An information-based approach to ill-posed problems*, J. Complexity **3** (1987), no. 3, 270–301. MR 919677 (89b:65141)
25. ———, *What is the complexity of ill-posed problems?*, Numerical Functional Analysis and Optimization **9** (1987), no. 9-10, 945–967.

Weierstrass Institute, 10117 Berlin, Mohrenstrasse 39, Germany
E-mail address: peter.mathe@wias-berlin.de

Radon Institute for Computational and Applied Mathematics
E-mail address: sergei.pereverzyev@oeaw.ac.at