Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Optimal discretization of inverse problems in Hilbert scales. Regularization and self–regularization of projection methods

Peter Mathé[1]

Sergei V. Pereverzev[2]

submitted: 28 Jan 2000

[1] Weierstraß Institute
for Applied Analysis and Stochastics
Mohrenstraße 39
D–10117 Berlin
Germany
E-Mail: mathe@wias-berlin.de

[2] Ukrainian Academy of Sciences
Inst. of Mathematics
Tereshenkivska Str. 3
Kiev 4
Ukraine
E-mail: serg-p@mail.kar.net

Preprint No. 548
Berlin 2000

1991 Mathematics Subject Classification. 62G05, 65J10.

Key words and phrases. Ill-posed problems, inverse estimation, operator equations, Gaussian white noise, information complexity.

Research supported by the Volkswagenstiftung (RiP–program at Oberwolfach).

Edited by
Weierstraß–Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
D — 10117 Berlin
Germany

Fax:	+ 49 30 2044975
E-Mail (X.400):	c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
E-Mail (Internet):	preprint@wias-berlin.de
World Wide Web:	http://www.wias-berlin.de/

ABSTRACT. We study the efficiency of the approximate solution of ill–posed problems, based on discretized observations, which we assume to be given afore–hand. We restrict ourselves to problems which can be formulated in Hilbert scales. Within this framework we shall quantify the degree of ill–posedness, provide general conditions on projection schemes to achieve the best possible order of accuracy. We pay particular attention on the problem of self–regularization vs. Tikhonov regularization.

Moreover, we study the information complexity. Asymptotically, any method, which achieves the best possible order of accuracy must use at least such amount of noisy observations.

We accomplish our study with two specific problems, Abel's integral equation and the recovery of continuous functions from noisy coefficients with respect to a given orthonormal system, both classical ill–posed problems.

1. INTRODUCTION AND STATEMENT OF THE MAIN PROBLEM

We study optimal discretizations of ill–posed problems in Hilbert scales. On the class of problems, which will be introduced in Section 2, the best order of accuracy for a given noise level is well known. But this quantity does not take into account any discretization. So we address two issues.

First, can this best possible order of accuracy be achieved by fully discretized regularization methods? More precisely, we aim at presenting general conditions, which allow to achieve this best possible order. This is made explicit for convergence analysis of corresponding schemes of Tikhonov regularization as well as for regularization by projection methods, self–regularization. In particular we pay attention to the limitations of self–regularization and indicate, how these naturally occur, when the design is given afore–hand.

A second issue to be addressed concerns the size, say $N = N(\delta)$ of the design, necessary for a given noise level $\delta > 0$, to enable best order of accuracy. This may be understood as the information complexity of the problem, since, in the asymptotic setting, no numerical method can achieve the best order of accuracy using less amount of noisy observations. We establish the asymptotic behavior $N(\delta),\ \delta \to 0$.

We shall study ill–posed problems, where we wish to recover some element x from some Hilbert space from indirectly observed data near $y = Ax$, where A is some injective compact linear operator acting from X to X. Such linear inverse problems often arise in scientific context, ranging from stereological microscopy (Abel's integral equation), physical chemistry (Fujita's equation) to satellite geodesy (gravity gradiometry equation). In practice indirect observations cannot be observed exactly but only in discretized and noisy form. To be more precise, we have only a vector $\varphi(y_\delta) = \{y_{\delta,i}\}_{i=1}^n \in \mathbb{R}^n$ defined by

$$y_{\delta,i} = \langle y_\delta, \varphi_i\rangle = \langle Ax, \varphi_i\rangle + \varepsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $\langle , \rangle$ denotes the inner product in X, $\varphi_i, i = 1, \dots, n$ is some orthonormal system, usually called design, and $\varepsilon_i, i = 1, \dots, n$ is the noise, which is assumed to be at some level δ and we will write $\varepsilon_i = \delta\xi_i$ with ξ_i denoting the normalized noise. The unknown solution is supposed to belong to some set $\mathcal{M} \subset Y$.

At this point it is important to note, that we assume to have observations without repetitions, which means, that each of the functionals $\langle y_\delta, \varphi_i \rangle$ is observed only once. The problem becomes different, if we allow repetitions. In that case the behavior depends very much on the kind of noise.

For deterministic noise we do not gain from repetitions; only the space spanned by the design elements is important. If we know, that the noise is random and independent, then we may use this to decrease the noise level at each functional by suitable repetitions. Under these circumstances one can asymptotically even achieve arbitrary accuracy. We postpone further discussion to Section 9.

Form our discussion above we can summarize, that given an equation

$$y_\delta = Ax + \delta\xi, \tag{1.2}$$

and assuming that the noise is either deterministic or random, the ill–posed problem is completely characterized by the triple $(A\colon Y \to X, \mathcal{M}, \delta)$, where X indicates the space where noisy observations are given, Y denotes the space in which we agree to measure the accuracy. The operator A in equation (1.2) determines the way the observations are indirect. The set $\mathcal{M} \subset Y$ describes our a priory knowledge on the exact solution and $\delta > 0$ indicates the noise level. We shall think of $(A\colon Y \to X, \mathcal{M}, \delta)$ as being the mathematical problem under consideration.

This mathematical problem is accompanied with a numerical one. When seeking approximate solutions to a given mathematical problem $(A\colon Y \to X, \mathcal{M}, \delta)$, we have to specify the class of admissible numerical methods. This class will generically be denoted by $\mathcal{U}$. Each method $u \in \mathcal{U}$ must be based on some design, say $\{\varphi_1, \dots, \varphi_n\}$, which describes the way we obtain noisy observations, see (1.1). The resulting approximation based on such design may be obtained by any (measurable) mapping $S : \mathbb{R}^n \to Y$, hence

$$u = S(\langle y_\delta, \varphi_1 \rangle, \langle y_\delta, \varphi_2 \rangle, \dots, \langle y_\delta, \varphi_n \rangle).$$

Let $\mathcal{U}_n$ denote the class of all methods based on design of at most n elements. We assume, that $\mathcal{U} = \bigcup_{n=1}^{\infty} \mathcal{U}_n$ and find it convenient to denote $(A\colon Y \to X, \mathcal{U}, \mathcal{M}, \delta)$, the numerical problem, corresponding to the mathematical problem $(A\colon Y \to X, \mathcal{M}, \delta)$.

We will make this clear in an example which is very simple.

Example 1. Let us fix the Hilbert space l_2 of square summable sequences and suppose we observe a sequence $y_\delta = x + \delta\xi,\ y_\delta \in l_2$, and we want to recover $x \in l_2$ under a priory knowledge, that for a certain $\mu > 0$, $\sum_{k=1}^{\infty} k^{2\mu} |x_k|^2 \le R^2$. Then, letting

$$\mathcal{M} := W_R^\mu := \left\{ (x_k)_{k=1}^{\infty}, \quad \sum_{k=1}^{\infty} k^{2\mu} |x_k|^2 \le R^2 \right\},$$

we can rephrase this as studying the mathematical problem $(I : l_2 \to l_2, W_R^\mu, \delta)$.

If observations in this setup are given by

$$y_{\delta,k} = x_k + \delta\xi_k, k = 1, \ldots, n, \tag{1.3}$$

then we restrict ourselves to method $u \in \mathcal{U}$, based on such observations. We thus aim at recovering a sequence, based on noisy coordinates, when some additional summability is known.

There is a considerable literature concerned with deterministic noise, the classical approach to inverse problems. We mention Tikhonov and Arsenin [41], Morozov [26], Vainikko and Veretennikov [46], Traub, Wasilkowski and Wozniakowski [43], Louis [21], Werschulz [48], Engl, Hanke and Neubauer [12]. On the other hand for stochastic noise, in which case we deal with a statistical problem, we refer to Wahba [47], Nychka and Cox [32], Johnstone and Silverman [17], Nussbaum [30], Donoho [10], Mair and Ruymgaart [24], Golubev and Khasminskii [14], Lukas [22], Cavalier and Tsybakov [6], Chow, Ibragimov and Khasminskii [7].

As far as the design is concerned it is sometimes possible to chose it. In this case one might prefer to chose elements from the singular value decomposition of the operator A or the wavelet–vaguelette decomposition. For such designs the recovery of x from noisy data as in (1.1) was studied by [38] and more recently studied by Johnstone and Silverman [17], Donoho [10] and Golubev and Khasminskii [14]. Often the design is given and independent of the operator. An example for this is the estimation of a probability density function x from discretized (binned) data or histogram as y_δ. In this case we assume that the operator A is of the form

$$Ax(t) = \int_0^1 a(t,\tau)x(\tau)d\tau, \tag{1.4}$$

acting in $X = L_2(0,1)$. A particular example for this is Abel's equation, see example 4 below. The data are given as averages of histogram bins $[u_{i-1,n}, u_{i,n})$ with bin limits $0 = u_{0,n} < u_{1,n} < \cdots < u_{n,n} = 1$, i.e.,

$$y_{\delta,i} = \frac{1}{u_{i,n} - u_{i-1,n}} \int_{u_{i-1,n}}^{u_{i,n}} Ax(t)dt + \varepsilon_i, \quad i = 1, \ldots, n. \tag{1.5}$$

This corresponds to equation (1.1) with design $\varphi_i = \varphi_i(t) = (u_{i,n} - u_{i-1,n})^{-1}\chi_{i,n}(t)$, where $\chi_{i,n}$ denotes the characteristic function of the interval $[u_{i-1,n}, u_{i,n})$. The approximate solution of Abel's integral equation based on histograms was studies by Nychka and Cox [32]. It is easy to verify that such histogram design does not correspond to neither the singular value decomposition nor the wavelet–vaguelette decomposition of the Abel integral operator, for example.

It is the aim of this paper to study efficiency issues for recovering the unknown element x from indirect and noisy discrete observations, as described above in terms of the numerical problem $(A, \mathcal{U}, \mathcal{M}, \delta)$.

For any given method $u \in \mathcal{U}$, its error at the exact solution x will be measured as

$$e^{det}(A, u, x, \delta) = \sup\{\|x - u(y_\delta)\|_Y, \quad \|\xi\|_X \le 1\}$$

for deterministic noise, while for stochastic noise we put

$$e^{ran}(A,u,x,\delta) = \left(\mathbf{E}\|x - u(y_\delta)\|_Y^2\right)^{1/2}, \tag{1.6}$$

where $\mathbf{E}$ denotes the expectation with respect to the random noise.

As usual, the uniform error over $x \in \mathcal{M}$ is defined to be the supremum over pointwise errors with respect to $x \in \mathcal{M}$, e.g. for deterministic noise we let $e^{det}(A,u,\mathcal{M},\delta) = \sup_{x\in\mathcal{M}} e^{det}(A,u,x,\delta)$. These quantities measure the quality of some specific method u. If we let u vary within class $\mathcal{U}$ of methods, then we may consider the minimal error within this class, $e^{det}(A,\mathcal{U},\mathcal{M},\delta) = \inf_{u\in\mathcal{U}} e^{det}(A,u,\mathcal{M},\delta)$, and the corresponding version $e^{ran}(A,\mathcal{U},\mathcal{M},\delta)$ for stochastic noise. In accordance with usual terminology in Information-based complexity, see [43] we denote

$$r_n^{det}(A,\mathcal{M},\delta) := e^{det}(A,\mathcal{U}_n,\mathcal{M},\delta) \tag{1.7}$$

denote the nth minimal radius of information within class $\mathcal{U}$, and $r_n^{ran}(A,\mathcal{M},\delta)$ its stochastic counterpart.

For any class $\mathcal{U}$ of methods a lower bound for $e^{det}(A,\mathcal{U},\mathcal{M},\delta)$ or $e^{ran}(A,\mathcal{U},\mathcal{M},\delta)$ is certainly given by

$$\mathcal{E}^{det}(A,\mathcal{M},\delta) := \inf_{u:X\to X^\nu} \sup_{x\in\mathcal{M}} \sup_{\|\xi\|_X\le 1} \|x - u(Ax+\delta\xi)\|_Y,$$

with the corresponding version for stochastic noise. Above the inf is taken over all possible (measurable) mappings. Most investigations on ill–posed problems have centered around this quantity.

Within the classical framework of optimal recovery (formally we may let $\delta = 0$) we have $e^{det}(A,\mathcal{U}_n,\mathcal{M},\delta) \to 0$, as n increases,which means that we can recover x with arbitrary accuracy by enlarging the design properly. In contrast, under the presence of noise $\delta > 0$ the sequence $r_n^{det}(A,\mathcal{M},\delta)$ will decrease, but there will be a positive limit

$$\lim_{n\to\infty} r_n^{det}(A,\mathcal{M},\delta) = e^{det}(X,\mathcal{U},\mathcal{M},\delta) \ge \mathcal{E}^{det}(A,\mathcal{M},\delta) > 0,$$

depending on the noise level. This limit cannot be beaten by any numerical method within class $\mathcal{U}$.

Thus we measure the quality of any numerical method u against this lower bound. For this purpose we fix $C > 1$ and seek for n with $r_n^{det}(A,\mathcal{M},\delta) \le C e^{det}(A,\mathcal{U},\mathcal{M},\delta)$. This results in a number $n = n(\delta)$ which measures the minimal amount of information necessary to recover x by methods from class $\mathcal{U}$ with best possible accuracy up to constant C,

$$N^{det}(A,\mathcal{U},\mathcal{M},\delta) := \inf\left\{n, r_n^{det}(A,\mathcal{M},\delta) \le C e^{det}(A,\mathcal{U},\mathcal{M},\delta)\right\}$$

and the respective version within the framework of stochastic noise. We agree to call these numbers the information complexity of the numerical problem $(A,\mathcal{U},\mathcal{M},\delta)$ under deterministic or random noise, respectively.

Thus our focus will be on two problems. Given a numerical problem, can we describe the asymptotic behavior of the information complexity? We shall establish lower bounds and then indicate, how these bounds can be achieved by regularization methods based on fully discrete projection schemes. Moreover, we will discuss, whether self–regularization of projection methods, based on given design is possible, or some other regularization is required.

2. Degree of ill-posedness in Hilbert scales

We recall that we are given the mathematical problem $(A : Y \to X, \mathcal{M}, \delta)$, where δ describes the level of noise in the data. We wish to ensure that any possible solution x_δ also varies within a ball, proportional to δ. It is easy to see, that this is exactly the case if $\|A^{-1} : X \to Y\| < \infty$. Otherwise the problem is ill–posed and we lose accuracy when recovering x from noisy data. Thus we want to assign each mathematical problem its degree of ill–posedness, a notion which was first be coined by Wahba [47]. In general it is hard to quantify this degree of ill–posedness. If we formulate the problem in a parametric scale of spaces, then we may express the degree in terms of these parameters. Therefore we shall study problems in Hilbert scales.

A Hilbert scale $\{X^\lambda\}_{\lambda\in\mathbb{R}}$ is a family of Hilbert spaces X^λ with inner product $\langle u, v\rangle_\lambda := \langle L^\lambda u, L^\lambda v\rangle$, where L is a given unbounded strictly positive self–adjoint operator in a dense domain of some initial Hilbert space, say X. To be more precise, X^λ is defined as the completion of the intersection of domains of the operators $L^\nu, \nu \geq 0$, accomplished with norm $\|\cdot\|_\lambda$ defined as $\|x\|_\lambda := \langle x, x\rangle_\lambda^{1/2} = \|L^\lambda x\|_0$, where $\|\cdot\|_0 = \|\cdot\|_X$. Usually $\{X^\lambda\}_{\lambda\in\mathbb{R}}$ are specific Sobolev spaces, say $H^\lambda(0,1)$; for this reason and for definiteness of scaling we assume λ be chosen to fit the usual smoothness as e.g., in $H^\lambda(0,1)$. This goal is achieved by assuming that the canonical embedding $J_\lambda : X^\lambda \to X, \lambda > 0$, obeys

$$a_n(J_\lambda) := \inf\left\{\|J_\lambda - U\|_{X^\lambda\to X}, \operatorname{rank} U < n\right\} \asymp n^{-\lambda}. \tag{2.1}$$

Thus $a_n(J_\lambda)$ denotes the nth approximation number (see [34]) and $\asymp$ means equivalent in order.

Example 2. Consider the diagonal mapping $L\,(x_k)_{k=1}^\infty \in l_2 \to (kx_k)_{k=1}^\infty \in l_2$. This operator generates the scale W^λ, as described above. It is immediate, that the set W_R^μ introduced in Example 1 is just the ball in W^μ of radius R.

We will, as this is often done for problems formulated in Hilbert scales assume, that the operator A is adapted to the scale in the following sense: For some parameter $a > 0$ there exist constants $D, d > 0$ such that

$$d\|x\|_{\lambda-a} \leq \|Ax\|_\lambda \leq D\|x\|_{\lambda-a} \tag{2.2}$$

for all $\lambda \in \mathbb{R}$ and $x \in X^{\lambda-a}$. In other words, the operator A acts along the Hilbert scale with step a as isomorphism between pairs $X^{\lambda-a}$ and X^λ.

Within the scale $W^\lambda, \lambda \in \mathbb{R}$, any diagonal mapping $D_a\,(x_k)_{k=1}^\infty \to (k^{-a}x_k)_{k=1}^\infty$ is easily seen to establish isometries between W^λ and $W^{\lambda+a}$.

There is substantial literature devoted to inverse problems with operators acting along Hilbert scales. We mention Natterer [28], Neubauer [29], Mair [23], Hegland

[15], Tautenhahn [40] and Dicken and Maass [9]. These papers studied only problems with deterministic noise. Moreover, the recovery problem from finitely many observations, i.e., a finite design with elements not necessarily depending on the operator A was not considered.

We illustrate this setup by two examples.

Example 3. We introduce Symm's equation

$$\int_\Gamma \log(|u-v|)z(v)dS_v = g(u), \quad u \in \Gamma, \tag{2.3}$$

arising from the Dirichlet boundary value problem for the Laplace equation in a region with boundary curve Γ. We assume that Γ admits a C^∞–smooth 1–periodic parametrization $\gamma : [0,1] \to \Gamma$. Then we can rewrite equation (2.3) as

$$Ax(t) := \int_0^1 \log(|\gamma(t) - \gamma(\tau)|)x(\tau) = y(t), \quad t \in [0,1],$$

where $x(t) := z(\gamma(t))\,|\gamma'(t)|$ and $y(t) := g(\gamma(t))$. It can be seen that the operator A obeys condition (2.2) with $a = 1$ within the scale $X^\lambda := H^\lambda(0,1), \lambda \in \mathbb{R}$ of Sobolev spaces of 1–periodic functions (distributions), see e.g., Bruckner et al. [5] for details.

Further examples of integral operators can be found in Neubauer [29], Mair and Ruymgaart [24].

Often a given operator does not fit any of the known scales. However in many cases one can construct a scale adapted to the operator. This is the case, when A acts compactly and injectively in some Hilbert space X. Then A meets condition (2.2) with $a = 1/2$ in the scale generated by the operator $L := (A^*A)^{-1}$, see Natterer [28] and Hegland [15]. It should however be noted that further verification is required to see that scaling is according to (2.1). We illustrate this by introducing the following important example.

Example 4. Let $X = L_2(0,1)$ and the operator A be Abel's integral operator

$$Ax(t) := \frac{1}{\sqrt{\pi}} \int_t^1 \frac{x(\tau)}{\sqrt{\tau - t}} d\tau, \quad t \in (0,1). \tag{2.4}$$

If now X^λ is generated by $L = (A^*A)^{-1}$, where

$$A^*x(t) = \frac{1}{\sqrt{\pi}} \int_0^t \frac{x(\tau)}{\sqrt{t-\tau}} d\tau, \tag{2.5}$$

then we indeed have

Proposition 1. *The Hilbert scale generated by $L = (A^*A)^{-1}$ from Abel's integral operator satisfies*

$$a_n(J_\lambda : X^\lambda \to X) = a_n(L^{-\lambda} : X \to X) \asymp n^{-\lambda}, \quad \lambda > 0.$$

Proof. We first observe

$$a_n(J_\lambda : X^\lambda \to X^0) = a_n((A^*A)^\lambda) = s_n((A^*A)^{1/2})^{2\lambda} = s_n(A^*)^{2\lambda}, \tag{2.6}$$

above s_n denotes the n–th singular number, see [34]. We shall provide a simple argument yielding

$$s_n(A^*) \asymp n^{-1/2} \tag{2.7}$$

The exact asymptotics was derived in [11].

To this end we recall the integral modulus of continuity

$$\omega_2(f,h) := \sup_{0<t\leq h} \left(\int_0^{1-t} |f(t+\tau) - f(t)|^2 \, d\tau \right)^{1/2}$$

and introduce the corresponding space $H_2^{1/2} \subset L_2(0,1)$ of square integrable functions for which

$$\|f\|_{H_2^{1/2}} := \|f\|_{L_2} + \sup_{0<h<1} \frac{\omega_2(f,h)}{\sqrt{h}} < \infty.$$

It is well known, that the embedding $H_2^{1/2} \to L_2(0,1)$ is compact. More specifically, if $0 = u_{0,n} < u_{1,n} < \ldots < u_{n,n} = 1, n \in \mathbb{N}$ is a sequence of partitions of $[0,1]$, and Q_n denote the corresponding projections in $L_2(0,1)$ onto the subspaces of piecewise constant functions on these subintervals, then

$$\|f - Q_n f\| \leq \max_{j=1,\ldots,n} \sqrt{(u_{j,n} - u_{j-1,n})} \|f\|_{H_2^{1/2}}. \tag{2.8}$$

In particular, if

$$\max_{j=1,\ldots,n} |u_{j,n} - u_{j-1,n}| \asymp n^{-1}, \tag{2.9}$$

then $\|f - Q_n f\| \leq C n^{-1/2} \|f\|_{H_2^{1/2}}$.

This is important in connection with the operators A and A^*, since both act continuously from $L_2(0,1) \to H_2^{1/2}$, we refer to [36, Thm. 14.2]. By factorization through $H_2^{1/2}$ we derive

$$\begin{aligned} &\|(I - Q_n)A\| \\ &\leq \|A : L_2 \to H_2^{1/2}\| \|f - Q_n f : H_2^{1/2} \to L_2(0,1)\| \leq C n^{-1/2} \end{aligned} \tag{2.10}$$

and respectively

$$\begin{aligned} &\|(I - Q_n)A^*\| \\ &\leq \|A^* : L_2 \to H_2^{1/2}\| \|f - Q_n f : H_2^{1/2} \to L_2(0,1)\| \leq C n^{-1/2} \end{aligned} \tag{2.11}$$

whenever the partitions obey (2.9). Therefore $s_n(A^*) \leq C n^{-1/2}$.

On the other hand, using the semigroup properties of fractional integration, see e.g. [36], we see that

$$(A^*)^2 x(t) = \int_0^t x(\tau) d\tau.$$

For this operator the singular numbers are well known, see [21, Beisp. 2.1.5], $s_n((A^*)^2) = ([n + 1/2]\pi)^{-1}$. This leads to

$$s_n(A^*)^2 \geq s_{2n}((A^*)^2) = ([2n + 1/2]\pi)^{-1},$$

which establishes a lower bound for $s_n(A^*)$. Both (2.6) and (2.7) provide the desired asymptotics. □

Returning to our problem $(A, \mathcal{M}, \delta)$ we assume that the noisy data are given in $X = X^0$ and we measure accuracy in $Y := X^\nu$. Suppose now, that $A : X^\nu \to X^0$ has a bounded inverse. Then we may formally invert it and rewrite

$$y_\delta = Ax + \delta\xi, \tag{2.12}$$

as

$$x = A^{-1}y_\delta - \delta A^{-1}\xi. \tag{2.13}$$

So, if the noise is deterministic with $\|\xi\|_0 \le 1$, and A has a bounded inverse, then in principle we can recover the unknown solution x up to order δ; Otherwise it might be ill–posed and the quantity

$$\alpha := \sup\left\{\lambda, \quad \|A^{-1} : X^0 \to X^\lambda\| < \infty\right\} \tag{2.14}$$

is less than or equal to ν. So it is natural to define $(\nu - \alpha)_+$ to be the degree of ill–posedness.

In case the operator A obeys property (2.2) for some a the supremum in condition (2.14) is attained. and $(\nu - \alpha)_+ = (\nu + a)_+$ as the degree of ill–posedness of the operator A. We mention, that for $\nu = 0$ the notion of degree of ill–posedness coincides with the one introduced by Wahba [47].

For stochastic noise the situation is a little more difficult. Even if we arrive at equation (2.13), and assume that the operator A has a bounded inverse, then, since the noise is stochastic, we cannot guarantee that the accuracy is kept, since $\mathbf{E}\|A^{-1}\xi\|_\nu^2$ may be unbounded. So, in analogy to the deterministic setting we will say, that the problem is well–posed, if the above expectation is bounded and ill–posed otherwise. If the latter is the case, then we let

$$\alpha := \sup\left\{\lambda, \quad \mathbf{E}\|A^{-1}\xi\|_\lambda < \infty\right\}.$$

Again we agree do denote the degree of ill–posedness by $(\nu - \alpha)_+$, now for random noise.

We make this precise for the white noise model, where ξ is a weak or generalized random element, such that for any $f \in X$ the inner product $\langle f, \xi\rangle$ is a measurable function, mapping a probability space $(\Omega, \Sigma, \mathbb{P})$ into $\mathbb{R}$ equipped with its Borel σ-field. Moreover, for any $f, g \in X$ the functions $\langle f, \xi\rangle, \langle g, \xi\rangle$ are square-summable with respect to the probability measure $\mathbb{P}$ and

$$\mathbf{E}\langle f, \xi\rangle = \mathbf{E}\langle g, \xi\rangle = 0, \quad \mathbf{E}\langle f, \xi\rangle\langle g, \xi\rangle = \langle f, g\rangle. \tag{2.15}$$

For such noise we can compute the degree of ill–posedness.

Proposition 2. *An operator $A : X^\nu \to X^0$, which obeys assumption* (2.2) *has degree of ill–posedness $(\nu + a + 1/2)_+$, when the noise in equation* (2.12) *is assumed to be white noise.*

Proof. Suppose, we have fixed some λ, for which $\mathbf{E}\|A^{-1}\xi\|_\lambda^2 < \infty$. Using the properties of white noise and the generator L of the Hilbert scale, we may rewrite this expectation as

$$\mathbf{E}\|A^{-1}\xi\|_\lambda^2 = \mathbf{E}\|L^\lambda A^{-1}\xi\|_0^2 = \sum_{k=1}^\infty \sigma_k^2(L^\lambda A^{-1}), \tag{2.16}$$

where $\sigma_k(L^\lambda A^{-1})$ denotes the kth singular number of the operator $L^\lambda A^{-1} : X^0 \to X^0$. In particular, this operator needs to be compact, which in turn implies, that $\lambda < -a$. Since $L^\lambda A^{-1}$ acts in a Hilbert space, the singular numbers coincide with the approximation numbers, see e.g., [34, Prop. 11.3.3]. But, since $L^\lambda : X^\lambda \to X^0$ is an isometry, we derive at

$$\sigma_k(L^\lambda A^{-1}) = a_k(L^\lambda A^{-1}) = a_k(A^{-1} : X^0 \to X^\lambda).$$

Now we study the imbedding $J_{-a}^\lambda : X^{-a} \to X^\lambda$, which is compact, since $\lambda < -a$. Since by assumption (2.2) the operator acts along the Hilbert scale, we use the factorization $J_{-a}^\lambda = A^{-1}A : X^{-a} \to X^0 \to X^\lambda$ to derive $a_k(J_{-a}^\lambda) \le \|A\| a_k(A^{-1})$. By our scaling assumption (2.1) and using property (2.2) this yields a lower bound

$$a_k(A^{-1}) \ge D^{-1} a_k(J_{-a}^\lambda) \asymp k^{a+\lambda}$$

. By (2.16) we arrive at

$$\sum_{k=1}^\infty k^{2(a+\lambda)} \le C \sum_{k=1}^\infty \sigma_k^2(L^\lambda A^{-1}) < \infty,$$

which is true whenever $\lambda < -a - 1/2$. Thus $\alpha = -a - 1/2$, which corresponds to a degree of ill–posedness of $(\nu + a + 1/2)_+$. □

We end our short digression with the remark, that for deterministic noise, the bound α was attained, assuming property (2.2). For white noise this is not the case; but as we will see below, the best possible accuracy will nevertheless be reflected by the degree of ill–posedness as just defined.

In addition, we assumed that the actual solution belongs to $\mathcal{M} \subset X^\nu$. If $\mathcal{M}$ is a ball in some X^μ, then this requires $\mu \ge \nu$ and we let $\mu - \nu$ be the effective smoothness of $\mathcal{M}$. Thus our problem is characterized by $(\nu + a)_+$ and $(\mu - \nu)$ in the deterministic case, whereas for white noise these parameters are $(\nu + a + 1/2)_+$ and $(\mu - \nu)$.

3. Minimal errors: Lower bounds

In this section we shall provide lower bounds for minimal errors for the numerical problem $(A, \mathcal{U}, \mathcal{M}, \delta)$, using the asymptotics of the corresponding best possible accuracies $\mathcal{E}^{det}(A, \mathcal{M}, \delta)$ and $\mathcal{E}^{ran}(A, \mathcal{M}, \delta)$. For this purpose we fix $\mathcal{M}$ being the ball X_R^μ in X^μ of radius $R > 0$.

We first note, that in the deterministic setting, as in our discussion above, we have well–posedness for $\nu \le -a$. This yields $\mathcal{E}^{det}(A, \mathcal{M}, \delta) \asymp \delta$. For $\nu > -a$, the asymptotics of $\mathcal{E}^{det}(A, \mathcal{M}, \delta)$ is well known. We recall the following result, see e.g. Natterer [28] and Tautenhahn [40], but also [12].

Proposition 3. *Under assumption* (2.2) *we have for* $\mu, a > 0$ *and* $\nu \in [-a, \mu]$ *the following asymptotics.*

$$\mathcal{E}^{det}(A, X^{\mu}_{R}, \delta) \asymp \delta^{\frac{\mu-\nu}{(\mu-\nu)+(\nu+a)_+}}. \tag{3.1}$$

Below,we shall obtain the stochastic analog of estimate (3.1) under the additional assumption, that the operator A and the generator L of the Hilbert scale are properly related. Following Mair and Ruymgaart (1996) we assume that the eigenvectors of the operator L coincide with the eigenvectors of A^*A. This means that both the operator L^{-1} and the operator A can be represented in the form

$$L^{-1}g = \sum_{k=1}^{\infty} l_k(g, f_k)f_k, \quad Ag = \sum_{k=1}^{\infty} \gamma_k(g, f_k)u_k, \tag{3.2}$$

where $\{f_k\}$, $\{u_k\}$ are some orthonormal bases of X. Taking into account assumptions (2.1) (2.2) (3.2), we see, that

$$l_k \asymp k^{-1}, \quad \gamma_k \asymp k^{-a}. \tag{3.3}$$

We have

Proposition 4. *Let the assumptions* (2.1), (2.2) *and* (3.2) *be fulfilled. Then for* $\mu, a > 0$ *and* $\nu \in [-a, \mu]$ *we have*

$$\mathcal{E}^{ran}(A, X^{\mu}_{R}, \delta) \asymp \delta^{\frac{\mu-\nu}{(\mu-\nu)+(\nu+a+1/2)}}.$$

The proof relies on the following Lemma, originally proven in Korostelev and Tsybakov [18, Chapt. 9], see also [3].

Lemma 1. *Suppose we are given*

$$v_k = \theta_k + \delta\sigma_k\xi_k, \quad k = 1, 2, \dots, \tag{3.4}$$

where ξ_k *are i.i.d* $N(0,1)$, $\sigma_k \asymp k^b$ *and* $\theta = (\theta_1, \theta_2, \dots)$ *is unknown, but belongs to*

$$B^s_R := \{\theta : \sum_k \lambda_k^2 \theta_k^2 \le R^2, \ \lambda_k \asymp k^s\}.$$

Then for $b, s > 0$ *the following asymptotics holds,*

$$\inf_{\hat\theta(v)} \sup_{\theta \in B^s_R} \mathbf{E}\|\theta - \hat\theta(v)\|^2_{l_2} \asymp \delta^{\frac{2s}{s+b+1/2}},$$

where the inf *is taken over all estimators* $\hat\theta(v)$ *based on observations* (3.4).

Let us mention, that this Lemma may also be viewed in connection with Example 1. When putting this carefully into that framework, we see, that the corresponding ill–posed problem has degree of ill–posedness $b + 1/2$.

We note also, that Donoho [10] obtained an analog of Lemma 1 in the more general situation with Besov bodies and l_p–norm. We turn to the

Proof of Proposition 4. Using (3.2) and (3.3) we can represent the observations (1.2) in the equivalent form (3.4) with $v_k = \gamma_k^{-1} l_k^{-\nu}(y_\delta, u_k)$ $\sigma_k = \gamma_k^{-1} l_k^{-\nu} \asymp k^{\nu+a}$ and $\theta_k = (L^\nu x, f_k)$, $k = 1, 2, \dots$ Here it as important that no repetitions are allowed. Since we assumed $x \in X_R^\mu$, the ball in X^μ of radius R, we conclude that $\theta := (\theta_1, \theta_2, \dots) \in B_R^s$ for $s := \mu - \nu$. Indeed, since $\lambda_k = l_k^{-(\mu-\nu)}$ we derive

$$\sum_{k=1}^\infty \lambda_k^2 \theta_k^2 = \sum_{k=1}^\infty l_k^{-2(\mu-\nu)} (L^\nu x, f_k)^2 = \sum_{k=1}^\infty l_k^{-2\mu} (x, f_k)^2 = \|L^\mu x\|^2 \le R^2.$$

Note, that any estimator $\hat\theta(v)$ of θ based on observations (3.4) yields an approximation $u := \sum_{k=1}^\infty \hat\theta_k(v) f_k$ for $L^\nu x$ and vice versa. Therefore, applying Lemma 1 with $b := \nu + a$ and $s := \mu - \nu$ we obtain the desired asymptotics

$$\mathcal{E}^{ran}(A, X_R^\mu, \delta) \asymp \delta^{\frac{\mu-\nu}{\mu+a+1/2}},$$

completing the proof. □

Comparing the asymptotics from Proposition 4 with the one in (3.1) we conclude that indeed white noise introduces an additional degree 1/2 of ill–posedness, this time caused by the stochastic nature. This influence of white noise to the degree of ill–posedness was observed by Nussbaum [30] for the special case, when the operator A denotes a–fold integration considered in Sobolev spaces. He also gave a heuristic explanation, see [31].

The above bounds on the optimal accuracy yield respective lower bounds on the minimal error. This will be discussed below. We turn to the main result of this section, a lower bound for the minimal errors in the deterministic as well as for the stochastic setting for the class $\mathcal{U}$ of methods based on designs consisting of linear functionals, as in equation (1.1). Recall that, given n we denote by $\mathcal{U}_n$ the class of all methods based on design of at most n elements.

Corollary 1. *We assume that* (2.1) *and* (2.2) *hold.*

In the deterministic setting, we have for $-a \le \nu \le \mu$ *the lower bound*

$$r_n^{det}(A, X_R^\mu, \delta) \ge c\left\{n^{-(\mu-\nu)} + \delta^{\frac{\mu-\nu}{\mu+a}}\right\}. \tag{3.5}$$

In the white noise model, and under the additional assumption (3.2) *we have for* $\nu \in [-a, \mu]$ *the lower bound*

$$r_n^{ran}(A, X_R^\mu, \delta) \ge c\left\{n^{-(\mu-\nu)} + \delta^{\frac{\mu-\nu}{\mu+a+1/2}}\right\} \tag{3.6}$$

Proof. We shall indicate the proof only for stochastic noise. We omit the proof for deterministic noise, since it is very similar. To make arguments easier, we consider x as well as $u(y_\delta)$ as elements in $L_2(X^\nu)$, thus we rewrite (1.6) as

$$e^{ran}(A, u, x, \delta) = \left(\mathbf{E}\|x - u(y_\delta)\|_\nu^2\right)^{1/2} = \|x - u(y_\delta)\|_{L_2(X^\nu)}.$$

To shorten notation we assume that each method $u \in \mathcal{U}_n$ uses its own design $\varphi_1, \dots, \varphi_n$. We now estimate

$$
\begin{aligned}
r_n^{ran}(A, X_R^\mu, \delta) &\geq \inf_{u \in \mathcal{U}_n} \sup_{\substack{x \in X_R^\mu \\ \langle Ax, \varphi_i \rangle = 0 \\ i=1,\dots,n}} \|x - u(\delta\xi)\|_{L_2(X^\nu)} \\
&\geq \inf_{u \in \mathcal{U}_n} \sup_{\substack{x \in X_R^\mu \\ \langle Ax, \varphi_i \rangle = 0 \\ i=1,\dots,n}} \max\left\{\|x - u(\delta\xi)\|_{L_2(X^\nu)}, \|x + u(\delta\xi)\|_{L_2(X^\nu)}\right\} \\
&\geq \inf_{u \in \mathcal{U}_n} \sup_{\substack{x \in X_R^\mu \\ \langle Ax, \varphi_i \rangle = 0 \\ i=1,\dots,n}} \|x\|_{L_2(X^\nu)} \\
&= \inf_{u \in \mathcal{U}_n} \sup_{\substack{x \in X_R^\mu \\ \langle Ax, \varphi_i \rangle = 0 \\ i=1,\dots,n}} \|x\|_\nu \geq \inf_{\varphi_1,\dots,\varphi_n} \sup_{\substack{x \in X_R^\mu \\ \langle x, \varphi_i \rangle = 0 \\ i=1,\dots,n}} \|x\|_\nu, \qquad (3.7)
\end{aligned}
$$

since x was deterministic. The last expression in (3.7) is just the nth Gelfand number of the embedding $J_\mu^\nu : X^\mu \to X^\nu$, see [25] for details on optimal approximative methods and s–numbers. Since for operators acting between Hilbert spaces all s–numbers coincide, see ibid., we conclude

$$
r_n^{ran}(A, X_R^\mu, \delta) \geq a_n(J_\mu^\nu) \asymp n^{-(\mu-\nu)}. \qquad (3.8)
$$

Since on the other hand, by the very definition

$$
r_n^{ran}(A, X_R^\mu, \delta) \geq \mathcal{E}^{ran}(A, X_R^\mu, \delta) \asymp \delta^{\frac{\mu-\nu}{\mu+a+1/2}} \qquad (3.9)
$$

we arrive together with estimate (3.8) at

$$
r_n^{ran}(A, X_R^\mu, \delta) \geq c\left\{n^{-(\mu-\nu)} + \delta^{\frac{\mu-\nu}{\mu+a+1/2}}\right\}. \qquad (3.10)
$$

This completes the proof for white noise. □

The above lower bounds are important as they connect the level of accuracy, expressed in terms of δ with $n^{-(\mu-\nu)}$, which depends only on the size of the design. As the proof indicated, this quantity actually is determined by the Gelfand numbers of the embedding J_μ^ν. Therefore, in more general classes of Banach spaces we will meet situations where nonlinear approximation may be superior to linear, we refer to [10] for approximation in Besov scales. In Hilbert scales this cannot be observed, since all s-numbers coincide.

4. Discretized noisy observations as data for projection schemes. Regularization and self–regularization of projection methods

In Sections 5 and 6 it will be shown, that the minimal error as indicated in Corollary 1 can be attained for a variety of designs when applying some regularization method to some projection schemes. We recall that the noisy equation is given by

$$
y_\delta = Ax + \delta\xi,
$$

where at the moment noise may be deterministic or random. When fixing a design $\{\varphi_1, \ldots, \varphi_n\}$ we may rewrite (1.1) as

$$Q_n(y_\delta) = Q_n(Ax + \delta\xi), \tag{4.1}$$

where Q_n denotes the orthogonal projection onto $\mathsf{span}\,\{\varphi_1, \ldots, \varphi_n\}$. Note that (4.1) is the standard form of a projection scheme for the approximate solution of the noisy operator equation (1.2). But if this is ill–posed, then some regularization may be required for solving (4.1).

The classical approach to regularization was proposed by Tikhonov; its application to non–discretized equations (1.2) in Hilbert scales was analyzed by Natterer [28].

For the particular case of deterministic noise and a Hilbert scale, generated by the operator $L = (A^*A)^{-1}$, Tikhonov regularization was extended to the situation, when the operator A in equation (1.2) is only known up to some deterministic noise, i.e., instead of A we have A_h, with $\|A - A_h\|_{X \to X} \le h$. We refer to Vaĭnikko and Veretennikov [46, Chapt. 4]. If we understand A_h as an approximation to the original operator after some discretization, then the essence of their considerations is the following. If $h \asymp \delta$, then applying appropriate regularization to the noisy equation $A_h x = y_\delta$, we can obtain the best possible accuracy, as indicated in Proposition 3 with $\nu = 0, a = 1/2$. For the projection scheme (4.1) this has the following interpretation. Suppose that we choose the number n of observations in such a way, that

$$\|A - Q_n A\|_{X \to X} \le \delta, \tag{4.2}$$

and we apply to (4.1) some regularization method. Then the result just mentioned guarantees the best order of accuracy. But, as can be drawn from Plato and Vainikko [35], we refer also to the example studied in Section 7, the number of observations chosen from (4.2) is far from being optimal. Therefore, in general the discretization effect of projection schemes cannot be reduced to this general setup.

Another approach to the approximate solution of equation (1.2) was undertaken by Engl and Neubauer [13] and Neubauer [29], who studied semi–discrete methods. Within this framework of semi–discrete methods the operator equation is discretized only from one side, passing from (1.2) to

$$AP_{m,s}x = y_\delta, \tag{4.3}$$

where $P_{m,s}$ denotes the orthogonal projection (now in some X^s) onto some finite dimensional subspace $V_m \subset X^s$. An application of Tikhonov regularization to semi–discrete equation was studied by Engl and Neubauer [13] and Neubauer [29]. If we apply the scheme as proposed by these authors, then this will result in a discretization of the data y_δ with respect to a basis which is determined by the operator A, as will be discussed in Remark 1 in Section 5. But, as the example of binned data (1.5) shows, the discretization with respect to the noisy data y_δ is often given afore-hand.

For this reason it is important to study the "two–sided" discretization of (1.2), namely

$$Q_n A P_{m,s} x = Q_n y_\delta. \tag{4.4}$$

Regularization of this fully discretized equation was studied by Plato and Vainikko [35], when the scale was generated by $L := (A^*A)^{-1}$. We delay further discussion to Remark 2 in Section 5, where consequences of the discretization of the operator A will be addressed.

The fully discrete equation (4.4) may not require further regularization by some Tikhonov method. Instead it may happen, that regularization may be achieved by just choosing the discretization parameters properly. This is called self–regularization, see Natterer [27], which is based on a stability property, see Lemma 4 below. In fact, assumption(2.2), made on the operator A has an important feature, since for suitably chosen sequences Q_n, P_m , the operators Q_nAP_m also enjoy similar properties, uniformly in m, n, when large enough.

Self–regularization of specific ill–posed problems in Hilbert scales through projection methods based on special orthonormal design elements has been analyzed in [5] and [9]. In a more general framework, but still not in Hilbert scales, self–regularization has been studied in [45]. To the best of our knowledge, all projection schemes which were applied and enjoy the property of self–regularization may in principle be derived from the standard form of the least–square method, as introduced in equation (6.1). They necessary lead to a one–sided discretization as in (4.3), hence require specific projections Q_n.

Therefore we will study a modification of least square methods for two–sided discretization, applicable for deterministic as well as for stochastic noise. We add that in all previous study of self–regularization only deterministic noise was considered.

We close this section mentioning that assumptions to be made on the design which enable self–regularization are more restrictive than for Tikhonov regularization. This issue will be addressed in Section 6 in more detail. Therefore, if these assumptions are not met, than some different regularization is unavoidable. We shall provide an example in Section 7.

5. Tikhonov regularization of projection schemes in Hilbert scales

The classical variant of Tikhonov regularization based on observations (1.1) or in equivalent form (4.1) is obtained by minimizing the functional

$$\|Q_nAx - Q_ny_\delta\|_0^2 + \alpha\|x\|_s^2 \tag{5.1}$$

over some finite dimensional subspace $V_m \subset X^s$, where we assume that the true solution $x_0 = A^{-1}y \in X_R^\mu \subset X^s$. In statistics this kind of regularization is called regularization estimator and its behavior was studied by Li [19] and Speckman [39]. It can be seen from Neubauer [29], that the unique minimizer, say $x^\delta_{\alpha,n,m}$ of problem (5.1) has the form

$$x^\delta_{\alpha,n,m} = x^\delta_{\alpha,n,m}(\xi) = G_{\alpha,n,m}(y_\delta), \tag{5.2}$$

where

$$G_{\alpha,n,m} = (T^\#_{n,m}T_{n,m} + \alpha I)^{-1}T^\#_{n,m} = L^{-s}(B^*_{n,m}B_{n,m} + \alpha I)^{-1}B^*_{n,m},$$

$$T_{n,m} = Q_n A P_{m,s}, \quad B_{n,m} = Q_n A P_{m,s} L^{-s},$$

$P_{m,s}$ is the orthogonal projector form X_s onto V_m and $B^*, T^\#$ denote the adjoint operators of $B : X \to X$ and $T : X^s \to X$, respectively. In particular

$$T^\#_{n,m} = P_{m,s} L^{-2s} A^* Q_n, \quad B^*_{n,m} = L^s P_{m,s} L^{-2s} A^* Q_n.$$

Let us return to Example 1 introduced in Section 1. For fixed s and $x \in l_2$ we have Q_n being the projection onto the first n coordinates, and the projections $P_{m,s}$ are the same mappings, but considered as acting in W^s. Then it is easy to see, that for any choice of regularization parameter $\alpha > 0$ the corresponding solution to the problem in Example 1 based on observations (1.3) can be written as

$$G_{\alpha,n,m}(x) = \begin{cases} \frac{x_k}{1+\alpha k^{2s}}, & \text{if } k \le \min\{m,n\}, \\ 0, & \text{else} \end{cases}. \tag{5.3}$$

To estimate the performance of the approximating $x^\delta_{\alpha,n,m}$, additional properties of the design $\{\varphi_1, \dots, \varphi_n\}$ as well as of the choice of spaces V_m are required.

To be precise, we assume that

$$\|I - Q_n\|_{X_{a+t} \to X} \le c\, n^{-(a+t)}, \quad t \le s. \tag{5.4}$$

Note, that by assumption (2.1) the best possible order of approximation of elements from X^{a+t} in X^0 using designs of at most n elements is $n^{-(a+t)}$. We thus assume that this is achieved by our chosen design. For the projections $P_{m,s}$, projecting onto $V_m \subset X^s$ we require, as e.g. in Neubauer [29] that

$$\|I - P_{m,s}\|_{X_{a+2s} \to X_s} \le c\, m^{-(a+s)}. \tag{5.5}$$

If $s \ge (\mu - a)/2$ and (5.5) is fulfilled, then standard interpolation techniques, we refer to Babuška and Aziz [2] for details, yield

$$\|(I - P_{m,s}) x_0\|_s \le c\, m^{-(\mu - s)}, \tag{5.6}$$

whenever $x_0 = A^{-1} y \in X^\mu_R$, which will be useful below.

Since we have

$$\begin{aligned} x_0 - x^\delta_{\alpha,n,m}(\xi) &= x_0 - G_{\alpha,n,m} A x_0 - \delta G_{\alpha,n,m} \xi \\ &= (I - G_{\alpha,n,m} A) x_0 - \delta G_{\alpha,n,m} Q_n \xi, \end{aligned} \tag{5.7}$$

we will estimate both contributions separately. Precisely, we shall estimate

$$b_{\alpha,n,m}(x_0) := \|(G_{\alpha,n,m} A - I) x_0\|, \quad v_{\alpha,n,m}(\xi) := \delta \|G_{\alpha,n,m} \xi\|,$$

separately.

Lemma 2. *Let the assumptions* (2.2), (5.4), (5.5) *be fulfilled. Assume that for some* $\varkappa < 1$

$$\|I - Q_n\|_{X_a \to X} \le \varkappa\, D^{-1} (2^a d^s)^{\frac{1}{a+s}} \alpha^{\frac{a}{2(a+s)}}, \tag{5.8}$$

where D, d are the constants from (2.2). *Then*

$$b_{\alpha,n,m}(x_0) \le c \Big[\alpha^{\frac{\mu}{2(a+s)}} + m^{-s} \Big(1 + m^{-a} \alpha^{-\frac{a}{2(a+s)}}\Big) \Big(\alpha^{\frac{\mu - s}{2(a+s)}} + \|(I - P_{m,s}) x_0\|_s\Big) \Big].$$

Proof. Let $y = Ax_0$ be the true free term of our equation. Consider the elements $x^0_{\alpha,n,m} = G_{\alpha,n,m}y$ and

$$x^0_{\alpha,m} = (T^\#_m T_m + \alpha I)^{-1} T^\#_m y,$$

where $T_m = AP_{m,s}$, $T^\#_m = P_{m,s}L^{-2s}A^*$. It follows from Lemma 2.2, Lemma 3.2 by Neubauer (1988) that

$$\|(T^\#_m T_m + \alpha I)^{-1} T^\#_m\|_{X\to X} \le (2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}}, \tag{5.9}$$

and

$$\begin{aligned} &\|x_0 - x^0_{\alpha,m}\| \\ &\le c\Big[\alpha^{\frac{\mu}{2(a+s)}} + m^{-s}\Big(1 + m^{-a}\alpha^{-\frac{a}{2(a+s)}}\Big)\Big(\alpha^{\frac{\mu-s}{2(a+s)}} + \|(I - P_{m,s})x_0\|_s\Big)\Big]. \end{aligned} \tag{5.10}$$

Note that

$$b_{\alpha,n,m}(x_0) = \|x_0 - x^0_{\alpha,n,m}\| \le \|x_0 - x^0_{\alpha,m}\| + \|x^0_{\alpha,m} - x^0_{\alpha,n,m}\|. \tag{5.11}$$

Moreover, from (2.2), (5.8) and (5.9) it follows that

$$\begin{aligned} &\|x^0_{\alpha,m} - x^0_{\alpha,n,m}\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1}[(T^\#_m - T^\#_{n,m})y - (T^\#_m T_m - T^\#_{n,m}T_{n,m})x^0_{\alpha,n,m}]\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1} T^\#_m[(I - Q_n)y - (I - Q_n)AP_{m,s}x^0_{\alpha,n,m}]\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1} T^\#_m (I - Q_n)A(x_0 - x^0_{\alpha,n,m})\| \\ &\quad \le (2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}} \|I - Q_n\|_{X_a\to X_0} \|A(x_0 - x^0_{\alpha,n,m})\|_a \\ &\quad \le D(2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}} \|I - Q_n\|_{X_a\to X_0} b_{\alpha,n,m}(x_0) \\ &\quad \le \varkappa b_{\alpha,n,m}(x_0). \end{aligned} \tag{5.12}$$

Combining (5.11) and (5.12) we obtain

$$b_{\alpha,n,m}(x_0) \le (1 - \varkappa)^{-1} \|x_0 - x^0_{\alpha,m}\|. \tag{5.13}$$

The assertion of the lemma follows from (5.10) and (5.13). □

Lemma 3. *Let the assumptions* (2.2), (5.4), (5.5) *be fulfilled. If* $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$ *then*

$$\|G_{\alpha,n,m}\|_{X\to X} \le c\alpha^{-\frac{a}{2(a+s)}},$$

where c *does not depend on* α, n, m.

Proof. It is well known that for an arbitrary compact operator B from X to X

$$\|(B^*B + \alpha I)^{-1}B^*\|_{X\to X} \le \frac{1}{2\sqrt{\alpha}}, \quad \|B(B^*B + \alpha I)^{-1}B^*\|_{X\to X} \le 1. \tag{5.14}$$

In particular we have for any $f \in X$ the bound

$$\begin{aligned} &\|G_{\alpha,n,m}f\|_s = \|L^s G_{\alpha,n,m}f\| \\ &\qquad = \|(B^*_{n,m}B_{n,m} + \alpha I)^{-1}B^*_{n,m}f\| \le \frac{1}{2\sqrt{\alpha}}\|f\|. \end{aligned} \tag{5.15}$$

Moreover, from (2.2) and (5.14) it follows that

$$\begin{aligned}
\|G_{\alpha,n,m}f\|_{-a} &\leq d^{-1}\|AG_{\alpha,n,m}f\| \\
&= d^{-1}\|AL^{-s}(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\
&\leq d^{-1}\|B_{n,m}(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\
&+ d^{-1}\|(AL^{-s}-B_{n,m})(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\
&\leq d^{-1}\Big(1+\frac{1}{2\sqrt{\alpha}}\|AL^{-s}-B_{n,m}\|_{X\to X}\Big)\|f\|.
\end{aligned} \tag{5.16}$$

Now we derive an estimate for $\|AL^{-s}-B_{n,m}\|$:

$$\begin{aligned}
\|AL^{-s}-B_{n,m}\|_{X\to X} &\leq \|AL^{-s}-AP_{m,s}L^{-s}\|_{X\to X} \\
&+ \|AP_{m,s}L^{-s}-Q_nAP_{m,s}L^{-s}\|_{X\to X}.
\end{aligned} \tag{5.17}$$

By (2.2) and (5.5) we can continue

$$\begin{aligned}
\|AL^{-s}-AP_{m,s}L^{-s}\|_{X\to X} &\asymp \|(I-P_{m,s})L^{-s}\|_{X\to X_{-a}} \\
&= \|L^{-a}(I-P_{m,s})L^{-s}\|_{X\to X} \\
&= \|L^{s}L^{-s-a}(I-P_{m,s})L^{-s}\|_{X\to X} \\
&= \|L^{-s-a}(I-P_{m,s})\|_{X_s\to X_s} \\
&= \|(I-P_{m,s})L^{-s-a}\|_{X_s\to X_s} \\
&= \|(I-P_{m,s})\|_{X_{2s+a}\to X_s} \\
&\leq c\, m^{-(s+a)}.
\end{aligned} \tag{5.18}$$

(Note that $L^{-\nu}: X_t \to X_t$ is self-adjoint for $\nu \geq t$). Further, using (2.2) and (5.4) we find

$$\begin{aligned}
\|AP_{m,s}L^{-s}-Q_nAP_{m,s}L^{-s}\|_{X\to X} &\leq \|I-Q_n\|_{X_{a+s}\to X}\|AP_{m,s}L^{-s}\|_{X\to X_{a+s}} \\
&\leq c\, n^{-(a+s)}\|P_{m,s}L^{-s}\|_{X\to X_s} \\
&\asymp n^{-(a+s)}\|P_{m,s}\|_{X_s\to X_s} = n^{-(a+s)}.
\end{aligned} \tag{5.19}$$

If $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$ then (5.16)–(5.19) imply

$$\|G_{\alpha,n,m}f\|_{-a} \leq c\,\|f\|.$$

Interpolation together with (5.15) yields

$$\|G_{\alpha,n,m}f\|_0 \leq \|G_{\alpha,n,m}f\|_s^{\frac{a}{a+s}}\|G_{\alpha,n,m}f\|_{-a}^{\frac{s}{a+s}} \leq c\,\alpha^{-\frac{a}{2(a+s)}}\|f\|.$$

The lemma is proved. □

In order to optimize the rate of convergence in the above Tikhonov regularization we will determine $\alpha = \alpha(\delta)$ in such a way, that the contributions in both terms in the decomposition (5.7) are of the same order as $\delta \to 0$. This is accomplished in

Theorem 1. *Let the assumptions* (2.2), (5.4) *and* (5.5) *be fulfilled and suppose that the unknown solution belongs to* X_R^μ. *We further assume, that* $s \geq \max\{\nu, (\mu - a)/2\}$.

If, for deterministic noise, we chose

$$\alpha \asymp \delta^{\frac{2(a+s)}{\mu+a}}, \quad n \asymp m \asymp \delta^{-\frac{1}{\mu+a}},$$

then

$$\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|_\nu \leq c\delta^{\frac{\mu-\nu}{\mu+a}}, \quad \nu \in [-a, \mu]. \tag{5.20}$$

For Gaussian white noise satisfying (2.15) *a choice of*

$$\alpha \asymp \delta^{\frac{2(a+s)}{\mu+a+1/2}}, \quad n \asymp m \asymp \delta^{-\frac{1}{\mu+a+1/2}}$$

leads for $\nu \in [-a, \mu]$ *to a bound*

$$\left(\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|_\nu^2\right)^{1/2} \leq c\delta^{\frac{\mu-\nu}{\mu+a+1/2}}. \tag{5.21}$$

(The constants above in estimates (5.20) and (5.21) do not depend on α, δ, m and n.)

Proof. The proof is carried out in two steps. We first provide the required estimates for $\nu = 0$.

It follows from assumption (5.4) that for $t = 0$ and $n \asymp \alpha^{-\frac{1}{2(a+s)}}$ condition (5.8) is fulfilled. Thus we may apply Lemmas 2 and 3. Together with the bound (5.6) we obtain with $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$ for any realization of ξ the estimate

$$\begin{aligned} \|x_0 - x_{\alpha,n,m}(\xi)\|_0 &\leq b_{\alpha,n,m}(x_0) + v_{\alpha,n,m}(\xi) \\ &\leq c\alpha^{\frac{\mu}{2(a+s)}} + \delta\|G_{\alpha,n,m}\|_{X\to X}\|Q_n\xi\|_0 \\ &\leq c\left[\alpha^{\frac{\mu}{2(a+s)}} + \delta\alpha^{-\frac{a}{2(a+s)}}\|Q_n\xi\|_0\right]. \end{aligned} \tag{5.22}$$

For deterministic noise $\|Q_n\xi\|_0 \leq \|\xi\|_0 \leq 1$. Therefore, letting $\alpha \asymp \delta^{\frac{2(a+s)}{\mu+a}}$, we arrive at the desired estimate (5.20) in case $\nu = 0$.

On the other hand, for white noise, property (2.15) implies

$$\mathbf{E}\|Q_n\xi\|_0^2 = \sum_{i=1}^n \mathbf{E}\langle\varphi_i, \xi\rangle^2 = \sum_{i=1}^n \|\varphi_i\|_0^2 = n,$$

such that, letting $n \asymp \alpha^{-\frac{1}{2(a+s)}} \asymp \delta^{-\frac{1}{\mu+a+1/2}}$, we can continue from (5.22) as follows.

$$\begin{aligned} \mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|_0^2 &\leq c\left[\alpha^{\frac{\mu}{a+s}} + \delta^2\alpha^{-\frac{a}{a+s}}n\right] \\ &\leq c\left[\alpha^{\frac{\mu}{(a+s)}} + \delta^2\alpha^{-\frac{2a+1}{2(a+s)}}\right] \leq c\delta^{\frac{2\mu}{\mu+a+1/2}}. \end{aligned}$$

This is estimate (5.21) for $\nu = 0$.

We now turn to the general case $\nu \neq 0$. We note that, if we let $x^{\delta,\nu}_{\alpha,n,m}(\xi) := L^\nu x^\delta_{\alpha,n,m}(\xi)$, then this is the unique minimizer of the functional

$$\|Q_nAL^{-\nu}x - Q_ny_\delta\|_0^2 + \|x\|_{s_\nu}^2,$$

where we let $s_\nu := s - \nu$, on the space $V_{m,\nu} := L^\nu V_m \subset X^{s-\nu}$. Moreover, the orthogonal projector $P_{m,s-\nu} := L^\nu P_{m,s} L^{-\nu}$ in $X^{s-\nu}$ onto this subspace obeys the corresponding variant of (5.5)

$$\|I - P_{m,s-\nu}\|_{X_{a_\nu+2s_\nu}\to X_{s_\nu}} \le c\, m^{-(a_\nu+s_\nu)},$$

with respective $a_\nu := a + \nu$, since

$$\begin{aligned}\|I - P_{m,s-\nu}\|_{X_{a_\nu+2s_\nu}\to X_{s_\nu}} &= \|I - L^\nu P_{m,s} L^{-\nu}\|_{X_{a_\nu+2s_\nu}\to X_{s_\nu}}\\ &= \|I - P_{m,s}\|_{X_{a+2s}\to X_s}\\ &\le cm^{-(a+s)} = cm^{-(a_\nu+s_\nu)}.\end{aligned}$$

The same is true for (5.4). Moreover, the operator $A_\nu := AL^{-\nu}$ obeys condition (2.2) with a_ν, and finally $x_0^\nu := L^\nu x_0 \subset X_R^{\mu-\nu}$. Thus, using the estimates of the first part of the proof for $A_\nu, a_\nu, \mu_\nu := \mu - \nu, s_\nu$ and projection P_{m,s_ν} we obtain, for example in the white noise setting for $\alpha \asymp \delta^{\frac{2(a_\nu+s_\nu)}{a+\mu+1/2}}$ that

$$\begin{aligned}\left(\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|_\nu^2\right)^{1/2} &= \left(\mathbf{E}\|L^\nu x_0 - L^\nu x^\delta_{\alpha,n,m}(\xi)\|_0^2\right)^{1/2}\\ &= \left(\mathbf{E}\|x_0^\nu - x^{\delta,\nu}_{\alpha,n,m}(\xi)\|_0^2\right)^{1/2}\\ &\le c\delta^{\frac{\mu_\nu}{\mu_\nu+a_\nu+1/2}} = c\delta^{\frac{\mu-\nu}{\mu+a+1/2}}.\end{aligned}$$

This concludes the proof for $\nu \neq 0$ and of the Theorem. ☐

We mention explicitly, that the same method was chosen to yield the optimal approximation in all spaces X^ν. This will be optimal as long as the parameter s is chosen large enough.

Theorem 1 asserts among others that the best possible accuracy $\mathcal{E}^{det}(A, X_R^\mu, \delta)$ and respectively $\mathcal{E}^{ran}(A, X_R^\mu, \delta)$ of problems (A, X_R^μ, δ), indicated in Proposition 3 and 4, can be achieved using an appropriate Tikhonov regularization. In conjunction with the lower bounds for the respective radii of information $r_n^{det}(A, X_R^\mu, \delta)$ and $r_n^{ran}(A, X_R^\mu, \delta)$, as provided in Corollary 1 we arrive at the following asymptotics for the information complexities, as introduced at the end of Section 1.

Theorem 2. *In the deterministic case, for $\nu \in [-a, \mu]$, and under assumptions* (2.2) *and* (2.1) *the following asymptotics holds true.*

$$N^{det}(A, \mathcal{U}, X_R^\mu, \delta) \asymp \delta^{-\frac{1}{\mu+a}} \tag{5.23}$$

If, for white noise in addition condition (3.2) *is fulfilled, then for $\nu \in [-a, \mu]$*

$$N^{ran}(A, \mathcal{U}, X_R^\mu, \delta) \asymp \delta^{-\frac{1}{\mu+a+1/2}}. \tag{5.24}$$

In both cases, the optimal order of information complexity is achieved by Tikhonov regularization (5.1)*, if only the design obeys* (5.4).

Proof. We shall indicate the proof only for estimate (5.24). The respective arguments for (5.23) are similar. From Theorem 1 we draw that by choosing $n \ge n(\delta) \asymp m(\delta) \asymp$

$\alpha^{-\frac{1}{2(a+s)}} \asymp \delta^{-\frac{1}{\mu+a+1/2}}$ the following estimate holds true.

$$
\begin{aligned}
e^{ran}(A,\mathcal{U},X^\mu_R,\delta) &\le r^{ran}_n(A,X^\mu_R,\delta) \\
&\le r^{ran}_{n(\delta)}(A,X^\mu_R,\delta) \le c\delta^{\frac{\mu-\nu}{\mu+a+1/2}}.
\end{aligned}
\tag{5.25}
$$

Since by Proposition 4

$$
\delta^{\frac{\mu-\nu}{\mu+a+1/2}} \asymp \mathcal{E}^{ran}(A,X^\mu_R,\delta) \le e^{ran}(A,\mathcal{U},X^\mu_R,\delta),
$$

we conclude, that actually $e^{ran}(A,\mathcal{U},X^\mu_R,\delta) \asymp \delta^{\frac{\mu-\nu}{\mu+a+1/2}}$, such that for some sufficiently large C we have

$$
N^{ran}(A,\mathcal{U},X^\mu_R,\delta) \le n(\delta) \asymp \delta^{-\frac{1}{\mu+a+1/2}}.
$$

On the other hand, using Corollary 1 and estimate (5.25) we deduce

$$
\begin{aligned}
&\{n, \quad r^{ran}_n(A,X^\mu_R,\delta) \le Ce^{ran}(A,\mathcal{U},X^\mu_R,\delta)\} \\
&\subset \left\{n, \quad r^{ran}_n(A,X^\mu_R,\delta) \le C\delta^{\frac{\mu-\nu}{\mu+a+1/2}}\right\} \\
&\subset \left\{n, \quad n^{-(\mu-\nu)} \le \tilde{C}\delta^{\frac{\mu-\nu}{\mu+a+1/2}}\right\}.
\end{aligned}
$$

Hence

$$
N^{ran}(A,\mathcal{U},X^\mu_R,\delta) \ge \inf\left\{n, \quad n^{-(\mu-\nu)} \le \tilde{C}\delta^{\frac{\mu-\nu}{\mu+a+1/2}}\right\} \asymp \delta^{-\frac{1}{\mu+a+1/2}},
\tag{5.26}
$$

which concludes the proof of the Theorem. □

Remark 1. As already mentioned in Section 4, previous study of Tikhonov regularization was restricted to the semi–discrete setup, see [13]. There the approximate solution of equation (1.2) is determined by

$$
x^\delta_{\alpha,m} := \operatorname{argmin}\left\{\|Ax-y_\delta\|_0^2 + \alpha\|x\|_s^2, \quad x \in V_m \subset X^s\right\},
$$

which can be represented also as $x^\delta_{\alpha,m} = \left(\alpha I + T^\#_m T_m\right)^{-1} T^\#_m y_\delta$, where $T_m := AP_{m,s}$ and $T^\#_m := P_{m,s}L^{-2s}A^*$, cf. (5.2). From this representation it follows, that in order to construct $x^\delta_{\alpha,m}$ one needs the following discrete information of the data, namely

$$
y_{\delta,i} = \langle\psi_i, A^*y_\delta\rangle = \langle y_\delta, A\psi_i\rangle, i = 1,2,\dots,m,
$$

where $\{\psi_i\}_{i=1}^m$ is some orthonormal basis of $V_m \subset X^s$. Now this immediately implies, that in the version of Tikhonov regularization just described, the design cannot be chosen independently of the operator A, which is a drawback, since it is often given, e.g., as histogram (1.5).

Remark 2. Note that representation (5.2) is needed only for the analysis of the rate of convergence. The construction of $x^\delta_{\alpha,n,m}$ actually reduces to solving a system of linear algebraic equations and is completely determined by the choice of design $\{\varphi_1,\dots,\varphi_n\}$, parameter s and $\{\psi_i\}_{i=1}^m$, a basis for the finite dimensional subspace $V_m \subset X_s$ and finally by choosing the regularization parameter α. Then we can construct the regularized solution in the form $x_{\alpha,n,m}^\delta = \sum_{i=1}^m x_i\psi_i$, with unknown coefficients obtained from the system

$$\alpha x_i + \sum_{k=1}^m a_{ik}x_k = b_i, \quad i = 1,\dots,m \tag{5.27}$$

with coefficients

$$a_{ik} := \sum_{j=1}^n \langle A\psi_i, \varphi_j\rangle\langle A\psi_k, \varphi_j\rangle, \quad b_i := \sum_{j=1}^n \langle A\psi_i, \varphi_j\rangle\langle \varphi_j, y_\delta\rangle.$$

We stress, that to obtain such a system it is crucial, that the $\{\psi_i\}_{i=1}^m$ were orthogonal in X^s, which may hardly be constructed unless in special cases, e.g., if we let $s = 0$. This simple Tikhonov regularization, combined with projection scheme

$$Q_nAP_mx = Q_ny_\delta, \tag{5.28}$$

here P_m is the orthogonal projection onto $V_m \subset X^0$ was studied before; in particular by Plato and Vainikko [35]. But it is well known, that Tikhonov regularization for $s = 0$ has a saturation and cannot yield better asymptotic rates than $\delta^{2/3}$. Hence, if we want to use a simple projection scheme, as (5.28) and we want to achieve the optimal order of accuracy for $x \in X_R^\mu$, then there are only two possibilities. One either applies so–called higher skilled methods, as for example the iterated Tikhonov method or Landweber iteration to (5.28), which significantly complicates the numerical procedure, this was considered by Plato and Vainikko [35] and more recently by Pereverzev and Solodky [33] in the particular case, when the Hilbert scale is generated by $L = (A^*A)^{-1}$ and the noise was assumed to be deterministic, or we try to use the phenomenon of self–regularization, the topic we will study in the next section.

6. Self–regularization properties of projection methods. Two–sided discretization of the least–square method

As we have already discussed in Section 4, the most prominent projection scheme with self–regularization properties is the least–square method, where we agree to consider as regularized solution of equation (1.2) any element x_m^δ, which minimizes

$$x_m^\delta := \operatorname{argmin}\{\|Ax - y_\delta\|_0, \quad x \in V_m\}. \tag{6.1}$$

The least–square method in the form as given above has extensively been studied by Natterer [27], Vaĭnikko and Khyamarik [45], Louis [21] and Dicken and Maass [9].

As it turns out, this leads again to a one–sided discretization, because (6.1) is equivalent to Gauss symmetrization of the one–sided discretized equation

$$AP_my = y_\delta. \tag{6.2}$$

More precisely, x_m^δ solves the equation

$$P_mA^*AP_mx = P_mA^*y_\delta. \tag{6.3}$$

This leads to discretized information of the form

$$y_{\delta,i} = \langle \psi_i, A^* y_\delta \rangle = \langle A\psi_i, y_\delta \rangle, \quad i = 1, \dots, m,$$

where again $\{\psi_i\}_{i=1}^m$ is an orthonormal basis of $V_m \subset X$. Thus in its standard form (6.1), (6.3), the least–square method probably cannot use given observations, as e.g. in (1.1), independently of the operator A. Of course, this can be overcome when using instead of (6.3) the equation

$$AP_m x = Q_n y_\delta,$$

and obtain the regularized solution $\hat{x}^\delta_{n,m}$ from the resulting symmetrization

$$P_m A^* A P_m x = P_m A^* Q_n y_\delta,$$

being equivalent to

$$\hat{x}^\delta_{n,m} := \operatorname{argmin}\{\|Ax - Q_n y_\delta\|_0, \quad x \in V_m\}. \tag{6.4}$$

But this is inconvenient, since it requires two different discretizations of the operator A, instead of the simple projection scheme (5.28).

For the fully discrete projection scheme (5.28) a little more efforts have to be made, in order to even assume solvability. In fact, even if the original operator A was invertible, it may occur, that the sections $P_n A P_n$ are not. Böttcher [4] gave a simple example for an operator in Hilbert space l_2 and P_n the finite sections, where such effect occurs for all odd n. We refer to the paper for more details. But after Gauss symmetrization of (5.28), by which we mean turning to

$$P_m A^* Q_n A P_m x = P_m A^* Q_n y_\delta, \tag{6.5}$$

there is always a solution, say $x^\delta_{n,m}$. This however need not be unique, since the set of possible solutions of (6.5) is determined by

$$Q_n A P_m x = \Pi_{n,m} Q_n y_\delta, \tag{6.6}$$

where $\Pi_{n,m}$ is the orthogonal projection onto the image $\operatorname{Im}(Q_n A P_m)$. So, uniqueness can be achieved only if $\ker(Q_n A P_m) \cap \operatorname{Im}(P_m) = \{0\}$, which was observed by Vaĭnikko and Khyamarik [45]. This in turn can be achieved when the operators $Q_n A P_m$ obey the following stability property: there is a constant $c_s > 0$, such that for all $u \in V_m$ we have

$$c_s \|P_m u\|_\nu \le \|Q_n A P_m u\|_{\nu+a},$$

see Lemma 4 below.

We now turn to the assumptions, both of Jackson and Bernstein type, made on the projections P_m onto the subspace V_m and Q_n corresponding to the design $\varphi = \{\varphi_1, \dots, \varphi_n\}$, in order to make the projection scheme efficient. Precisely we assume that there is a positive $s > 0$, and there are constants $b, q > 0$, such that P_m obeys for $\nu \in [-a, \mu]$ the estimate

$$\|I - P_m\|_{X^\mu \to X^\nu} \le q\, m^{-(\mu-\nu)}, \tag{6.7}$$

and

$$\|P_m u\|_{s+\nu} \le b\, m^s \|P_m u\|_\nu, \quad \nu \in [-a, \mu]. \tag{6.8}$$

The projections Q_n satisfy

$$\|I - Q_n\|_{X^{s+\nu+a} \to X^{\nu+a}} \le q\, n^{-s}, \tag{6.9}$$

and also

$$\|Q_n u\|_{\nu+a} \le b\, n^{\nu+a} \|Q_n u\|_0. \tag{6.10}$$

For the white noise case we additionally assume

$$\|I - Q_n\|_{X^{\mu+a} \to X^0} \le q\, n^{-(\mu+a)}. \tag{6.11}$$

Assumptions of this kind were also made in [8]. Under the above assumptions the following bounds for the approximate solutions x_m^δ and $\hat{x}_{n,m}^\delta$, as defined through (6.1) and (6.4), respectively, have been proven, see e.g. [21, Satz 4.5.6]

Proposition 5. *Assume that assumptions* (6.7) *and* (6.8) *hold. Then for operators A along the Hilbert scale satisfying* (2.2) *and for $x_0 := A^{-1}y \in X^\mu$ we have*

$$\|x_0 - x_m^\delta\|_\nu \le c \left(\|(I - P_m)x_0\|_\nu + \delta m^{\nu+a} \|\xi\|_0 \right),$$

and respectively

$$\|x_0 - \hat{x}_{n,m}^\delta\|_\nu \le c \left(\|(I - P_m)x_0\|_\nu + \delta m^{\nu+a} \|y - Q_n y_\delta\|_0 \right).$$

Usually this kind of stability is assumed to be fulfilled for a given projection scheme, see e.g. [5], or it can be deduced from properties of the operator A, here we refer to [37], who studied Symm's equation. If we allow $m \ne n$, then we may deduce stability from our assumptions made above.

Lemma 4. *Assume* (2.2) *and* (6.8), (6.9) *to hold. Then there is $0 < c_0 < 1$, such that for $m = c_0 n$ stability*

$$c_s \|P_m u\|_\nu \le \|Q_n A P_m u\|_{\nu+a} \tag{6.12}$$

holds.

In particular, for $c_0 := (d/(2Dqb))^{1/s}$ we have (6.12) *with $c_s = d/2$.*

Proof. We derive from Assumption (2.2), that

$$d\|P_m u\|_\nu \le \|AP_m u\|_{\nu+a} \le \|(I - Q_n)AP_m u\|_{\nu+a} + \|Q_n A P_m u\|_{\nu+a}$$

for every $u \in V_m$. Using now the assumption made above, this yields

$$\begin{aligned} \|(I - Q_n)AP_m u\|_{\nu+a} &\le q n^{-s} \|AP_m u\|_{s+\nu+a} \\ &\le Dqn^{-s} \|P_m u\|_{s+\nu} \le Dqb(m/n)^s \|P_m u\|_\nu. \end{aligned}$$

This implies

$$[d - Dqb(m/n)^s] \|P_m u\|_\nu \le \|Q_n A P_m u\|_{\nu+a}.$$

It is now easy to derive the remaining assertion for $c_0 := (d/(2Dqb))^{1/s}$. □

We turn to the main result of this section.

Theorem 3. *Assume that the operator A obeys (2.2) and that the assumptions (6.7)–(6.10) are fulfilled. In the deterministic noise setting, for $x_0 = a^{-1}y \in X_R^\mu$ and $m = c_0 n \asymp \delta^{-1/(\mu+a)}$, where c_0 from Lemma 4, there is a constant c, for which we can bound the error of the approximate solution obtained from (6.5) by*

$$\|x_0 - x_{n,m}^\delta\|_\nu \le c\delta^{\frac{\mu-\nu}{\mu+a}}. \tag{6.13}$$

If, for white noise, we additionally assume that (6.11) holds, then for $m = c_0 n \asymp \delta^{-1/(\mu+a+1/2)}$ we have

$$\left(\mathbf{E}\|x_0 - x_{n,m}^\delta\|_\nu^2\right)^{1/2} \le c\delta^{\frac{\mu-\nu}{\mu+a+1/2}}. \tag{6.14}$$

Proof. We shall prove only estimate (6.14) and outline the one for (6.13) roughly. Using Proposition 5 and assumptions (6.7) as well as (6.11) we derive for $m = c_0 n$

$$\|x_0 - x_{n,m}^\delta\|_\nu \le \|x_0 - \hat{x}_{n,m}^\delta\|_\nu + \|\hat{x}_{n,m}^\delta - x_{n,m}^\delta\|_\nu. \tag{6.15}$$

The first norm can further be estimated as

$$\begin{aligned}\|x_0 - \hat{x}_{n,m}^\delta\|_\nu &\le c[\|(I-P_m)x_0\|_\nu + m^{\nu+a}\|(I-Q_n)y\|_0 + m^{\nu+a}\delta\|Q_n\xi\|_0] \\ &\le c[m^{-(\mu-\nu)}\|x_0\|_\mu + m^{\nu+a}m^{-(\mu+a)}\|y\|_{\mu+a} + m^{\nu+a}\delta\|Q_n\xi\|_0] \\ &\le c[m^{-(\mu-\nu)}\|x_0\|_\mu + m^{\nu+a}\delta\|Q_n\xi\|_0].\end{aligned} \tag{6.16}$$

If we now use $\hat{x}_{n,m}^\delta, x_{n,m}^\delta \in V_m$, then we can derive from (6.10) and (6.12) the following chain of estimates for the second summand in (6.15).

$$\begin{aligned}\|\hat{x}_{n,m}^\delta - x_{n,m}^\delta\|_\nu &\le c_s^{-1}\|Q_nAP_m(\hat{x}_{n,m}^\delta - x_{n,m}^\delta)\|_{\nu+a} \\ &\le c_s^{-1}bn^{\nu+a}\|Q_nAP_m(\hat{x}_{n,m}^\delta - x_{n,m}^\delta)\|_0 \\ &\le cn^{\nu+a}[\|Q_nAP_m\hat{x}_{n,m}^\delta - Q_ny_\delta\|_0 + \|Q_nAP_mx_{n,m}^\delta - Q_ny_\delta\|_0].\end{aligned} \tag{6.17}$$

Now $x_{n,m}^\delta$ from equation (6.5) has a representation as

$$x_{n,m}^\delta = \operatorname{argmin}\left\{\|Q_nAP_mx - Q_ny_\delta\|_0, \quad x \in V_m\right\}.$$

Using (6.7) we can continue

$$\begin{aligned}\|Q_nAP_mx_{n,m}^\delta - Q_ny_\delta\|_0 &\le \|Q_nAP_mx_0 - Q_ny_\delta\|_0 \\ &\le \|Q_nAx_0 - Q_ny_\delta\|_0 + \|Q_nA(I-P_m)x_0\|_0 \\ &\le \delta\|Q_n\xi\|_0 + \|(I-P_m)x_0\|_{-a} \\ &\le cm^{-(\mu+a)}\|x_0\|_\mu + \delta\|Q_n\xi\|_0.\end{aligned}$$

Moreover, from the definition of $\hat{x}_{n,m}^\delta$ in (6.4) and assumption (6.11) we conclude

$$\begin{aligned}\|Q_nAP_m\hat{x}_{n,m}^\delta - Q_ny_\delta\|_0 &\le \|AP_mx_{n,m}^\delta - Q_ny_\delta\|_0 \le \|AP_mx_0 - Q_ny_\delta\|_0 \\ &\le \|A(I-P_m)x_0\|_0 + \|(I-Q_n)y\|_0 + \delta\|Q_n\xi\|_0 \\ &\le c[m^{-(\mu+a)}\|x_0\|_\mu + \delta\|Q_n\xi\|_0].\end{aligned} \tag{6.18}$$

Thus, for $m = c_0 n$ the previous estimates (6.16) and (6.17), after being inserted into (6.15) yield

$$\|x_0 - x_{n,m}^\delta\|_\nu \le cm^{-(\mu-\nu)}\|x_0\|_\mu + \delta m^{\nu+a}\|Q_n\xi\|_0.$$

Since, as in the proof of Theorem 1 $\mathbf{E}\|Q_n\xi\|_0^2 = n$, we deduce for $n \asymp m \asymp \delta^{-\frac{1}{\mu+a+1/2}}$ finally

$$\left(\mathbf{E}\|x_0 - x_{n,m}^\delta\|_\nu^2\right)^{1/2} \leq c[m^{-(\mu-\nu)}\|x_0\|_\mu + \delta m^{\nu+a+1/2}] \leq c\delta^{\frac{\mu-\nu}{\mu+a+1/2}},$$

which proves (6.14). The corresponding estimate (6.13) for deterministic noise is obtained using x_m^δ instead of $\hat{x}_{n,m}^\delta$. We then conclude, using Proposition 5

$$\begin{aligned}\|x_0 - x_{n,m}^\delta\|_\nu &\leq \|x_0 - x_m^\delta\|_\nu + \|x_m^\delta - x_{n,m}^\delta\|_\nu \\ &\leq c[m^{-(\mu-\nu)}\|x_0\|_\nu + \delta m^{\nu+a}] + \|x_m^\delta - x_{n,m}^\delta\|_\nu.\end{aligned}$$

Moreover, from the definition (6.1) of x_m^δ and assumption (6.7) we obtain

$$\begin{aligned}\|Q_nAP_mx_m^\delta - Q_ny_\delta\|_0 \leq \|AP_mx_m^\delta - y_\delta\|_0 &\leq \|AP_mx_0 - y_\delta\|_0 \\ &\leq \|y - y_\delta\|_0 + \|A(I - P_m)x_0\|_0 \\ &\leq \delta + cm^{-(\mu+a)}\|x_0\|_\mu.\end{aligned}$$

So, using these estimates instead of (6.16) and (6.18) we arrive at the desired bound (6.13). Note, that in this case condition (6.11) was not used. □

Theorems 2 and 3 show, that the simple projection scheme (5.28) has after Gauss symmetrization (6.5) the self–regularization property. It is also order optimal in the sense of Information complexity. This means, that if the involved projections obey assumptions (6.7)-(6.10) (in the white noise setting additionally (6.11)), and the discretization parameters are properly chosen, then no further regularization is required, to approximately solve the problem (1.1), (1.2). The conditions (6.9) and (6.10) in particular require, that the design elements are smooth enough, i.e., $\varphi_i \in X^{\nu+a}, i = 1, \ldots, n$. This smoothness ensures the stability property (6.12), which is the key for self–regularization. On the other hand, if the design is given afore hand, then this smoothness might not be fulfilled. In this situation one can try to use Tikhonov regularization instead, because there the requirements, in particular (5.4) are less restrictive.

7. Application to Abel's equation

We will apply the results of the previous sections to Abel's equation as introduced in Example 4 in Section 2. It will be important to assume, that the design is given afore-hand in form of histograms (1.5). Noisy Abel's equation (2.12), with operator (2.4) arises from a diverse range of applications in the physical sciences and in stereological microscopy. Some pertinent references are Nychka and Cox [32], Johnstone and Silverman [17] and Donoho [10]. As already mentioned in Example 4, we study the problem in the Hilbert scale generated by $L := (A^*A)^{-1}$, such that assumption (2.2) is fulfilled with $a = 1/2$. We shall restrict to the case, where the exact solution belongs to X_R^1, thus $\mu = 1$. In this case, the exact solution admits a representation

$$x_0 = A^*Ag_0, \tag{7.1}$$

for some $g_0 \in L_2(0,1)$, $\|g\|_0 = R$. Using Corollary 1 in Samko (1968) we obtain the following representations

$$Af(t) = A^*Vf(t), \tag{7.2}$$

where A^* is given by (2.5). and the operator

$$Vf(t) := \frac{1}{\pi\sqrt{t}} \int_0^1 \frac{\sqrt{\tau} f(\tau) d\tau}{\tau - t}$$

acts boundedly from $L_2(0,1)$ into the space $L_{2,2\varepsilon}(0,1)$ of functions that are square-summable on (0,1) with weight $t^{2\varepsilon}$, where $\varepsilon > 0$ is arbitrarily small. This means that for any $f \in L_2(0,1)$ there exists $f_\varepsilon \in L_2(0,1)$ such that

$$Vf(t) = t^{-\varepsilon} f_\varepsilon(t) \quad \text{and} \quad \|f_\varepsilon\| \le c\|f\|. \tag{7.3}$$

Then (7.1) and (7.2) together with the semigroup property of fractional integration imply that

$$x_0(t) = A^*Ag_0(t) = A^*A^*Vg_0(t) = \int_0^t \tau^{-\varepsilon} g_{0,\varepsilon}(\tau) d\tau. \tag{7.4}$$

Thus, x_0 has derivative $x_0' \in L_{2,2\varepsilon}(0,1)$ for any small $\varepsilon > 0$. In terms of the modulus of continuity we can estimate the smoothness of the solution x_0 by

$$\omega_2(x_0, h) = O\big(h^{1-\varepsilon}\big) \tag{7.5}$$

for any small $\varepsilon > 0$.

As in Nychka and Cox [32] we will assume that bin limits of histograms (1.5) obey (2.9). As before we represent (1.5) in the form (4.1), where Q_n is the orthogonal projector on the subspace of piecewise constant functions having discontinuities at the points $\{u_{i,n}\}$.

Finally we agree to measure the error in $X = L_2(0,1)$. Note, that under these assumptions the histogram design does not meet conditions (6.9) and (6.10)., which are required for self–regularization. Indeed, assumption (6.9) would require, that $\mathrm{Im}(Q_n) \subset X^{1/2} = \mathrm{Im}(A^*A)^{-1/2} = \mathrm{Im}(A^*)$. But for the constant function $\varphi_0(t) \equiv 1$, which is certainly in the range of Q_n, this would imply the existence of a function $v_0 \in L_2(0,1)$, for which

$$A^*v_0(t) = \frac{1}{\sqrt{\pi}} \int_0^t \frac{v_0(\tau)}{\sqrt{t-\tau}} d\tau = \varphi_0(t) = 1. \tag{7.6}$$

But we can use the inversion formula to see

$$v_0(t) = \frac{1}{\sqrt{\pi}} \frac{d}{dt} \int_0^t \frac{d\tau}{\sqrt{t-\tau}} = \frac{1}{\sqrt{\pi t}}.$$

But it is immediate, that this function cannot belong to $L_2(0,1)$, which in turn implies $\mathrm{Im}(Q_n) \not\subset X^{1/2}$. Therefore the histogram design does not give rise to self–regularization and we are lead to apply Tikhonov regularization based on the design (1.5).

We first observe, that the bound (2.10) derived in Example (4) is just condition (5.4) for $t = 0$. A straightforward application of Theorem 1 in the present case requires to take $s \geq \frac{\mu - a}{2} = \frac{1}{4}$. On the other hand, it is inconvenient to use Tikhonov functional (5.1), (5.2) when s is a fraction. But for $s = 1$ the condition (5.4) breaks down for our case because using Q_n we cannot obtain an accuracy being superior to $O(n^{-1})$. Therefore we let $s = 0$. Then condition $s \geq \frac{\mu - a}{2}$ is violated, but if we only slightly change the value m ($m = \alpha^{-\frac{1}{1-\varepsilon}}$ instead of $m = \alpha^{-1}$) then estimate (7.5) allows to obtain the same order of accuracy as in Theorem 1.

Since $s = 0$, we let $P_{m,0} = Q_m$, the orthogonal projector like Q_n but corresponding to m bins. Estimate (7.5) implies

$$(7.7) \qquad \|(I - P_{m,0})x_0\|_0 = \|(I - Q_m)x_0\|_{L_2} \leq c\omega_2(x_0, m^{-1}) \leq cm^{-1+\varepsilon},$$

for any small $\varepsilon > 0$. Using (7.7) instead of (5.6) we arrive at

Theorem 4. *Suppose, that the exact solution of Abel's integral equation satisfies* (7.1)*. Let $x^\delta_{n,m}$ be the solution obtained from noisy data* (1.5) *when applying Tikhonov regularization* (5.1)*,* (5.2) *for $s = 0$, $P_{m,0} = Q_m$.*

For uniformly bounded deterministic noise we have the error bound

$$\|x_0 - x^\delta_{\alpha,n,m}\|_0 \leq c\delta^{2/3},$$

which is achieved by letting $\alpha \asymp \delta^{2/3}, n \asymp \delta^{-2/3}$ and $m \asymp \delta^{-2/(3-3\varepsilon)}$ for any fixed $1 > \varepsilon > 0$.

In case of white noise (2.15) *we obtain*

$$\left(\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}\|^2_0\right)^{1/2} \leq c\delta^{1/2},$$

which is achieved for $\alpha \asymp \delta^{1/2}$, $n \asymp \delta^{-1/2}$ and $m \asymp \delta^{-1/(2-2\varepsilon)}$ for any fixed $1 > \varepsilon > 0$.

We note that the information complexity $N^{det}(A, \mathcal{U}, X^1_R, \delta)$ and correspondingly $N^{ran}(A, \mathcal{U}, X^1_R, \delta)$ can be achieved by the given histogram bins (1.5).

Remark 3. We mention, that the assumptions (2.10), (2.11) and (7.1), are exactly those used by Plato and Vainikko [35] to apply Theorem 3 in that paper. Within the deterministic noise model this theorem guarantees the optimal accuracy as presented in Theorem 4, but with a value of $m \asymp \delta^{-4/3}$, which is worse than the one presented above.

Remark 4. We mentioned in the discussion in Section 2, that a sufficient discretization amount could be derived from (4.2). Taking into account the behavior of (2.7), in the case under consideration this would yield $n \asymp \delta^{-2}$, which is far from being optimal, as can be seen from Theorem 4 above.

8. Application to stable summation of Fourier series with noisy coefficients

The problem of stable summation of Fourier series with respect to a given orthonormal system of function $\{\varphi_k, k = 1, 2, \dots\}$ under small changes in the coefficients, measured in l_2 is a classical example of an ill–posed problem, in particular if we want to measure the error in the sup–norm. To be more specific we study functions belonging to $C(0,1)$, the space of continuous functions on $[0,1]$. The problem may now be formulated as follows. We want to recover a continuous function y from noisy (Fourier) coefficients with respect to a given system $\{\varphi_k, k = 1,2,\dots\}$, but instead of $y_k = \langle y, \varphi_k\rangle, k = 1,2,\dots$ we are given only a noisy sequence of numbers

$$(y_{\delta,k})_{k=1}^{\infty}\,, \quad \text{satisfying} \quad \sum_{k=1}^{\infty}(y_k - y_{\delta,k})^2 \le \delta^2. \tag{8.1}$$

This classical ill–posed problem was studied in [41, Chapt. 6], in an appendix to the textbook [16], in papers by Aliev [1] and many others. In all cases the application of Tikhonov regularization was considered. Standard assumptions on the smoothness of the true solution were expressed in terms of spaces W_2^{μ}, associated with the given system $\{\varphi_k, k = 1,2,\dots\}$, i.e.,

$$W_2^{\mu} := \left\{ y \in L_2(0,1), \quad \|y\|_{\mu}^2 := \sum_{k=1}^{\infty} k^{2\mu}\left|\langle y, \varphi_k\rangle\right|^2 < \infty \right\}.$$

Under these assumptions, Il'in and Pozniak [16] for the trigonometrical system and Aliev [1] for the more general case of any system with uniformly bounded norm $\|\varphi_k\|_{\infty} \le C,\ k = 1,2,\dots$, proved that for $p \in (1/2, 2\mu - 1)$ Tikhonov regularization $y_\alpha := \sum_{k=1}^{\infty} \frac{y_{\delta,k}}{1+\alpha k^p}\varphi_k$ yields $\|y - y_\alpha\|_\infty \le C(\sqrt{\alpha} + \delta/\alpha)$. So, the optimal choice for α is $\alpha_0 = \delta^{2/3}$ for which

$$\|y - y_\alpha\|_\infty \le C\delta^{1/3}. \tag{8.2}$$

To the best of our knowledge, we refer also to the survey by Liskovets [20], this is the culmination of all previous work on this particular problem. Still some questions remain open.

First, the estimate (8.2) does not take into account the given smoothness. Moreover, it does not indicate the actual degree of ill–posedness of the problem. It is common belief, that this degree depends on the growth of $\|\varphi_k\|_\infty \to \infty$. We will study this problem under two different assumptions on the growth, a summability one and an element wise one. More precisely, we assume

$$\sum_{k=1}^{\infty} \frac{\|\varphi_k\|_\infty^2}{k^{2\nu}} < \infty \tag{M_ν}$$

for some $\nu > 0$. On the other hand we will assume

$$\|\varphi_k\|_\infty \asymp k^\beta, k = 1,2,\dots, \tag{K^β}$$

now for some $\beta \ge 0$. It is immediate, that systems $\{\varphi_1, \varphi_2, \dots\}$, which obey (K^β), will satisfy (M_ν), with $\nu = \beta + 1/2 + \varepsilon$ for any small $\varepsilon > 0$. The trigonometric system has property (K^β) for $\beta = 0$, whereas the system of Legendre polynomials requires to take $\beta = 1$. One way to obtain systems from M_ν is to consider lacunary sequences $\{\varphi_k = \psi_{n_k}, n_k = \theta(k)\}$, where $\theta(k)$ increases. If $\{\psi_k, k = 1,2,\dots\}$ obeys (K^β), then $\{\psi_{n_k}, k = 1,2,\dots\}$ satisfies (M_ν), if $\theta(k)^\beta k^{-\nu}$ is square summable.

We now return to the problem of stable summation. Below it will turn out, that the solution to the problem of stable summation heavily depends as well on the classes of systems of functionals as on the kind of noise. Within the deterministic noise framework, functionals which enjoy property (M_ν) best fit to our previous discussion, and the solution can easily be derived from previous results. The pointwise growth does not exactly fit, but still a slight modification allows to complete the proof.

For white noise, the situation is completely different, since the behavior of the noise is very much dependent on geometric properties of the underlying space. This effect is especially seen, when measuring the noise in $C(0,1)$. It does not allow to work in Hilbert scales. We have to argue directly in the space of continuous functions. We start with systems which obey property (M_ν). We will show, that Theorems 1 and 3 apply to this situation. To achieve this goal we will need to find a formulation of a related problem in a suitable Hilbert scale. For this purpose we observe, that under (M_ν) we can conclude

$$\| \sum x_k \varphi_k \|_\infty \leq \left(\sum |x_k|^2 k^{2\nu} \right)^{1/2} \left(\sum \frac{\|\varphi_k\|_\infty^2}{k^{2\nu}} \right)^{1/2} < \infty, \tag{8.3}$$

if only $\sum |x_k|^2 k^{2\nu} < \infty$. Thus we may switch to a different point of view. We first try to recover the sequence $y = (y_k)_{k=1}^\infty$ of coefficients from the noisy data (8.1), considered as data (1.3) to obtain $\{\bar{y}_k\}$, and then use $\sum \bar{y}_k \varphi_k$ as approximation of the unknown function $y \in C(0,1)$.

Because of (8.3) this permits us to study the recovery problem in the scale W^λ, $\lambda \in \mathbb{R}$ for the identity mapping, thus example 1. We need to find an approximate solution in W^ν to deduce $\sum \bar{y}_k \varphi_k \in C(0,1)$. As has already been discussed in Section 5, the approximation to $x_0 \in W^\nu$ based on observations (1.3), obtained from Tikhonov regularization has the form presented in (5.3). Thus in this particular case $(a = 0)$ Theorem 1 translates into

Theorem 5. *Let $\{\varphi_k, k = 1, 2, \dots\}$ obey assumption (M_ν). For a function $y \in W_2^\mu$, corresponding to this system with $\mu > \nu$ the following result is true. If we are given noisy observation* (8.1)*, Tikhonov regularization, applied with $s \geq \max\{\mu/2, \nu\}$ and parameter $\alpha \asymp \delta^{2s/\mu}$, using only $n = m \asymp \delta^{-1/\mu}$ noisy data allows the following error estimate*

$$\|y - \sum_{k=1}^{n} \frac{y_{\delta,k}}{1 + \alpha k^{2s}} \varphi_k \|_\infty \leq \delta^{(\mu-\nu)/\mu} \|y\|_\mu. \tag{8.4}$$

We note that in the present context self–regularization simply means to take the truncated series $\sum_{k=1}^n y_{\delta,k} \varphi_k$ as approximation.

We turn to the corresponding result for systems, which obey (K^β).

Theorem 6. *Let $\{\varphi_k, k = 1, 2, \dots\}$ obey assumption (K^β). For a function $y \in W_2^\mu$, corresponding to this system with $\mu > \beta + 1/2$ the following result is true. If we* *choose the number of observations* $n \asymp \delta^{-1/\mu}$, *then the error can be bounded as*

$$\|y - \sum_{k=1}^{n} y_{\delta,k}\varphi_k\|_\infty \le c\delta^{(\mu-\beta-1/2)/\mu}\|y\|_\mu.$$

We note, that the bound given above is stronger, than the one we would obtain, simply using that (K^β) implies (M_ν) for $\nu = \beta + 1/2 + \varepsilon$ and applying Theorem 5.

Proof. Direct calculations show, that for $\mu > \beta + 1/2$ and $y \in W_2^\mu$ we obtain

$$\begin{aligned}\|y - \sum_{k=1}^{n} y_k\varphi_k\|_\infty &\le \sum_{k=n+1}^{\infty} |\langle y, \varphi_k\rangle| \, \|\varphi_k\|_\infty \\ &\le c\|y\|_\mu \left(\sum_{k=n+1}^{\infty} k^{2(\beta-\mu)}\right)^{1/2} \\ &\le cn^{-\mu+\beta+1/2}\|y\|_\mu.\end{aligned} \tag{8.5}$$

Moreover, if Q_n denotes the projection onto the first n functions $\varphi_1, \dots, \varphi_n$, then for any u we have the following Nikolsky type inequality.

$$\begin{aligned}\|Q_n u\|_\infty &\le \|\sum_{k=1}^{n} u_k\varphi_k\|_\infty \le c\sum_{k=1}^{n} |u_k| \, k^\beta \\ &\le c\left(\sum_{k=1}^{n} |u_k|^2\right)^{1/2} \left(\sum_{k=1}^{n} k^{2\beta}\right)^{1/2} \le cn^{\beta+1/2}\|Q_n u\|_0.\end{aligned}$$

Thus for $n \asymp \delta^{-1/\mu}$ this gives for $\|y\|_\mu \le R$ a bound

$$\begin{aligned}\|y - \sum_{k=1}^{n} y_{\delta,k}\varphi_k\|_\infty &\le \|y - Q_n y\|_\infty + \|Q_n(\sum_{k=1}^{n}(y_k - y_{\delta,k})\varphi_k)\|_\infty \\ &\le c[n^{-\mu+\beta+1/2} + n^{\beta+1/2}\delta] \le c\delta^{\frac{\mu-\beta-1/2}{\mu}},\end{aligned}$$

which completes the proof of the theorem. □

We mention, that this theorem evolved during discussions with V. Temlyakov, Univ. of South Carolina.

We also note, that a similar statement can be proven for systems, satisfying (K^β), when we use the method obtained by Tikhonov regularization, applied with $\alpha \asymp \delta^{2s/\mu}$, $2s > \mu$ and $n \asymp \delta^{-1/\mu}$.

The problem of stable summation of Fourier series with random noise

$$y_{\delta,k} = \langle y, \varphi_k\rangle + \delta\langle \xi, \varphi_k\rangle, \quad k = 1, \dots, n,$$

has also been studied in [41], and more recently by Tsybakov [44]. For trigonometric systems $\{\varphi_k, k = 1, 2, \dots\}$ and $y \in W_2^\mu$ Tsybakov indicated the best possible order for the expected value of of the error, measured in $C(0,1)$. This turns out to be of the order $(\delta \log^{1/2}(1/\delta))^{\frac{\mu-1/2}{\mu}}$. He proved that this order cannot be improved even if the trigonometric system is replaced by any other orthonormal system, still assuming $y \in W_2^\mu$, associated to the trigonometric system. It is worth mentioning, that in this paper, discretized Tikhonov regularization was used and that the regularization parameter α was chosen adapting to the unknown smoothness. Therefore he required, that the number n of observations was at least

$$n \geq \delta^{-2/(\min\{1,\mu_0-1/2\})}, \tag{8.6}$$

where μ_0, the minimal smoothness, was supposed to be known.

Using some geometric property of $C(0,1)$ we will show, that for given μ the necessary number of observations can be reduced to $n \asymp (\delta\sqrt{\log(1/\delta)})^{-1/\mu}$, still retaining the best order of accuracy.

We shall however again turn to the more general setup of systems obeying property (M_ν) or (K^β), respectively.

Theorem 7. *Suppose we are given data*

$$y_{\delta,k} = \langle y, \varphi_k\rangle + \delta\langle \xi, \varphi_k\rangle, \quad k = 1, \ldots, n,$$

where ξ is white noise. We assume to know, that the unknown function y belongs to W_2^μ.

For a system $\{\varphi_k, k = 1, 2, \ldots\}$ satisfying (M_ν), $\mu > \nu$, Tikhonov regularization with $2s > \max\{\mu - \nu, \nu\}$, and

$$n \asymp (\delta\sqrt{\log(1/\delta)})^{-1/\mu} \quad \textit{and} \quad \alpha \asymp (\delta\sqrt{\log(1/\delta)})^{2s/\mu}, \tag{8.7}$$

the following bound holds true

$$\left(\mathbf{E}\|y - \sum_{k=1}^n \frac{y_{\delta,k}}{1+\alpha k^{2s}}\varphi_k\|_\infty^2\right)^{1/2} \leq c(\delta\sqrt{\log(1/\delta)})^{\frac{\mu-\nu}{\mu}}. \tag{8.8}$$

For a system $\{\varphi_k, k = 1, 2, \ldots\}$ satisfying (K^β), $\mu > \beta + 1/2$, Tikhonov regularization with $2s > \max\{\mu - \beta, \beta\}$ and α and n chosen as in (8.7), we have the following bound.

$$\left(\mathbf{E}\|y - \sum_{k=1}^n \frac{y_{\delta,k}}{1+\alpha k^{2s}}\varphi_k\|_\infty^2\right)^{1/2} \leq c(\delta\sqrt{\log(1/\delta)})^{\frac{\mu-\beta-1/2}{\mu}}. \tag{8.9}$$

Proof. We shall carry out the proof only for estimate (8.9). The proof for systems satisfying (M_ν) is similar. First we decompose the error into the bias and pure stochastic term as usual,

$$b_{\alpha,n} := \|y - \sum_{k=1}^n \frac{y_k}{1+\alpha k^{2s}}\varphi_k\|_\infty \text{ and } v_{\alpha,n} := \|\sum_{k=1}^n \frac{\langle \xi, \varphi_k\rangle}{1+\alpha k^{2s}}\varphi_k\|_\infty. \tag{8.10}$$

The stochastic term $v_{\alpha,n}$ can now be estimated using the following bound on the type–2–constant of finite dimensional subspaces of $C(0,1)$, see [42, pp. 14-16],

$T_2(l_\infty^n) \le c\sqrt{\log(n)}$. This means in the present context

$$(8.11) \qquad \left(\mathbf{E}v_{\alpha,n}^2\right)^{1/2} \le c\sqrt{\log(n)}\left(\sum_{k=1}^n \frac{\|\varphi\|_\infty^2}{(1+\alpha k^{2s})^2}\right)^{1/2}.$$

The following estimates are based on elementary calculus, precisely

$$(8.12) \qquad \max_{z\in[1,\infty)} \frac{z^p}{1+\alpha z^q} = \alpha^{-p/q} q^{-1} p^{p/q}(q-p)^{(q-p)/q}, \quad p < q.$$

Inserting this into (8.11) we obtain

$$(8.13) \qquad \left(\mathbf{E}v_{\alpha,n}^2\right)^{1/2} \le c\alpha^{-\beta/(2s)}\sqrt{n\log(n)}, \text{ if } 2s > \beta.$$

The bias term also admits a similar representation

$$(8.14) \qquad b_{\alpha,n}(y) \le \|y - Q_n(y)\|_\infty + \alpha \sum_{k=1}^n \frac{k^{2s}}{1+\alpha k^{2s}} |y_k| \, \|\varphi_k\|_\infty.$$

The first term can be estimated similar to (8.5) above, which yields

$$(8.15) \qquad \|y - \sum_{k=1}^n y_k\varphi_k\|_\infty \le cn^{-\mu+\beta+1/2}\|y\|_\mu.$$

Moreover, using (8.12) again we obtain

$$(8.16) \qquad \sum_{k=1}^n \frac{k^{2s}}{1+\alpha k^{2s}} |y_k| \, \|\varphi_k\|_\infty \le c\sqrt{n}\alpha^{-\frac{2s+\beta-\mu}{2s}}\|y\|_\mu, \text{ if } 2s > \mu - \beta.$$

Gathering the estimates (8.13)–(8.16) and inserting this into the decomposition (8.10) we obtain under assumptions (8.7)

$$\begin{aligned} &(\mathbf{E}\|y - \sum_{k=1}^n \frac{y_{\delta,k}}{1+\alpha k^{2s}}\varphi_k\|_\infty^2)^{1/2} \\ &\le c\|y\|_\mu (n^{\beta-\mu+1/2} + \sqrt{n}\alpha^{(\mu-\beta)/(2s)}) + c\delta\alpha^{-\beta/(2s)}\sqrt{n\log(n)} \\ &\le c(\delta\sqrt{\log(1/\delta)})^{\frac{\mu-\beta-1/2}{\mu}}, \end{aligned}$$

which completes the proof of (8.9). □

We note, that the number n of observations and the regularization parameter α, which provide the optimal order do not depend on properties of the underlying system $\{\varphi_k, k = 1, 2, \dots\}$. Therefore, since the trigonometric system belongs to K_0, we obtain the best order of accuracy for this system with a number of observations much less than the one in (8.6).

9. Concluding discussion

We want to conclude this study with a discussion of some topics, which seem to be important.

One of the aims in the present study was to determine the information complexity, the minimal amount of information required to achieve the best possible accuracy. This asymptotics was indicated in Theorem 2 in Section 5 and appeared as a consequence of Tikhonov regularization. On the other hand, self–regularization achieves the same optimal order with asymptotically the same amount of information, see Theorem 3. Both results were based on assumptions on the design as well as on the finite dimensional space, the approximate solution was looked for, see (5.4) and (5.5) for Tikhonov regularization and (6.9)–(6.11) for self–regularization. As we saw in Section 7, design based on histogram bins did not allow to use self–regularization, whereas for stable summation of Fourier series, Section 8 either method worked.

At a first glance it is not easy to compare these conditions. We confine ourselves on the condition imposed on the design. If the design elements $\{\varphi_1, \dots, \varphi_n\}$ are smooth enough, then this usually implies good approximative power. For example, polynomial splines, which obey (6.10), will automatically satisfy (5.4) in the Sobolev scale, if only their degree is sufficiently high. But, at the same time, splines, which have an appropriate degree, but maximal defect, will not satisfy (6.10). So, loosely speeking, conditions (6.9)–(6.11), which were assumed for self–regularization, are more restrictive than (5.4). But, from a practical point of view, self–regularization is preferable, since it leads to a simpler numerical scheme, than Tikhonov regularization. So, if we can choose a design, or if it is smooth enough to enable self–regularization, then this is in favour.

As already mentioned in the discussion at the end of Section 6, the conditions (6.9) and (6.10) seem to be necessary for order optimal self–regularization, since the were the key to establish the stability property, see Lemma 4. As far as Tikhonov regularization is concerned, its structure, cf. (5.2) presumes that elements $P_{m,s}L^{-2s}A^*y_\delta$ are well approximated by $P_{m,s}L^{-2s}A^*Q_n y_\delta$. Now observe, that

$$\|P_{m,s}L^{-2s}A^*(I - Q_n)\|_{X\to X} = \|(I - Q_n)AP_{m,s}L^{-2s}\|_{X\to X}.$$

Since $AP_{m,s}L^{-2s}u \in X^{a+s}$ if $u \in X$, then it is apparent, that Q_n must approximate well on elements from X^{a+s}, which is the essence of assumption (5.4).

Another topic shall be mentioned. We already touched the problem, whether observations as in (1.1) might be replicated to improve performance in Section 1. In fact, let us look at non–parametric regression, where we observe function values

$$y_{\delta,k} = f(t_k) + \delta\xi_k, \quad k = 1, \dots, n,$$

$0 \le t_1 < \dots . t_n \le 1$, say, and we want to recover the unknown function f, which is assumed to be smooth. We also refer to a more recent example, studied in [18, Chapt. 9], an example where the observations were indirect. In this kind of regression problems we may use the closeness of the observation points to decrease the noise level. This may typically be done with kernel estimators by choosing an appropriate bandwidth.

In the present case, where the design is supposed to consist of orthogonal functionals in some Hilbert space, it is hard to imagine an analogous estimator. Instead we may think of direct repetitions of observations for each of the functionals φ_k. Within this framework the methods of Sections 5 and 6 apply and we shall indicate their order of approximation, expressed in the overall number of observations. To be more precise, let us consider the mathematical problem (A, X_R^μ, δ) and let n be the size of a design. If we allow k replications at each element, then the overall number of observations is $N = k \times n$. If we now average the observations at each element

$$\bar{y}_{\delta,i} := \frac{1}{k}\sum_{l=1}^{k} y_{\delta,i} = \langle Ax, \varphi_i\rangle + \delta\frac{1}{k}\sum_{l=1}^{k}\langle \xi_i^l, \varphi_i\rangle,$$

then we can see, that the data $\{\bar{y}_{\delta,i}\}_{i=1}^n$ are exactly those, which were considered in our study. For deterministic noise, we will not gain anything from these replications. But for white noise, the new data have reduced noise level $\delta_k := \delta/\sqrt{k}$. If we now apply one of the order optimal methods, say u, based on the design of size n, then we obtain an accuracy

$$e(A, u, X_R^\mu, \delta_k) \le c\left(\delta/\sqrt{k}\right)^{\frac{\mu-\nu}{\mu+a+1/2}}, \tag{9.1}$$

if only $n \asymp \left(\delta/\sqrt{k}\right)^{-\frac{1}{\mu+a+1/2}}$. This translates to an estimate

$$e(A, \bar{u}, X_R^\mu, \delta_k) \le cN^{-\frac{\mu-\nu}{2(\mu+a+1)}},$$

where $\bar{u}$ denotes the compound method, based on the original $N = k \times n$ data. Thus for stochastic white noise, we can even achieve arbitrary accuracy. With a different attitude, Mair and Ruymgaart [24] ended at a similar method, see Theorem 4.4 there, although these authors used more restrictive assumptions, in particular (3.2) and considered estimates based on the singular value decomposition, only. The convergence analysis in [24], which for the present case results in Theorem 4.2, there, is expressed in terms of the number of repetitions, which is k above. Then it is easily seen, that the bound given in [24, Thm. 4.2] is exactly the one from (9.1).

Moreover, it can easily be seen, that this method is order optimal in certain cases. In particular, if we assume (3.2) and use design based on the singular value decomposition of the operator A.

References

[1] B. Aliev. Estimation of the regularization method for the problem of the summability of the Fourier series. *Dokl. Akad. Nauk Tadzhik. SSR*, 21(4):3–6, 1978.

[2] I. Babuška and A. K. Aziz. Survey lectures on the mathematical foundation of the finite element method. In Aziz. A. K., editor, *The Mathematical Foundation of the Finite Element Method with Appl. to Part. Diff. Equations*, New York, 1972. Acad. Press.

[3] E. N. Belitser and B. Y. Levit. On minimax filtering over ellipsoids. *Math. Methods Statist.*, 4(3):259–273, 1995.

[4] Albrecht Böttcher, Aad Dijksma, Heinz Langer, Michael A. Dritschel, James Rovnyak, and M. A. Kaashoek. *Lectures on operator theory and its applications.* American Mathematical Society, Providence, RI, 1996. Lectures presented at the meeting held at the Fields Institute for Research in Mathematical Sciences, Waterloo, Ontario, September 1994, Edited by Peter Lancaster.

[5] Gottfried Bruckner, Siegfried Prössdorf, and Gennadi Vainikko. Error bounds of discretization methods for boundary integral equations with noisy data. *Appl. Anal.*, 63(1-2):25–37, 1996.

[6] L. Cavalier and A. B. Tsybakov. Sharp adaptation for inverse problems with random noise. manuscript, 1998.

[7] Pao-Liu Chow, Ildar A. Ibragimov, and Rafail Z. Khasminskii. Statistical approach to some ill-posed problems for linear partial differential equations. *Probab. Theory Related Fields*, 113(3):421–441, 1999.

[8] Wolfgang Dahmen, Angela Kunoth, and Reinhold Schneider. Operator equations, multiscale concepts and complexity. In *The mathematics of numerical analysis (Park City, UT, 1995)*, pages 225–261. Amer. Math. Soc., Providence, RI, 1996.

[9] V. Dicken and P. Maass. Wavelet-Galerkin methods for ill-posed problems. *J. Inverse Ill-Posed Probl.*, 4(3):203–221, 1996.

[10] D.L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.

[11] Milutin R. Dostanić. Asymptotic behavior of the singular values of fractional integral operators. *J. Math. Anal. Appl.*, 175(2):380–391, 1993.

[12] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems.* Kluwer Academic Publishers Group, Dordrecht, 1996.

[13] Heinz W. Engl and Andreas Neubauer. Convergence rates for Tikhonov regularization in finite-dimensional subspaces of Hilbert scales. *Proc. Amer. Math. Soc.*, 102(3):587–592, 1988.

[14] G. Golubev and Khasiminskii R. Statistical approach to inverse boundary value problems for partial differential equations. WIAS Preprint 370, 1997.

[15] M. Hegland. Variable Hilbert scale and their interpolation inequalities with applications to Tikhonov regularization. *Applic. Analysis*, 59:207–223, 1995.

[16] V.A. Il'in and E.G. Poznjak. Основы математического анализа. Nauka, Moscow, 1973.

[17] Iain M. Johnstone and Bernard W. Silverman. Discretization effects in statistical inverse problems. *J. Complexity*, 7(1):1–34, 1991.

[18] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*, volume 82 of *Lect. Notes Statist.* Springer, New York, 1993.

[19] K. C. Li. Minimaxity of the method of regularization of stochastic processes. *Ann. Statist.*, 10:937–942, 1982.

[20] O. A. Liskovets. Theory and methods of solving ill-posed problems. In *Mathematical analysis, Vol. 20*, pages 116–178, 264. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, 1982.

[21] Alfred Karl Louis. *Inverse und schlecht gestellte Probleme.* B. G. Teubner, Stuttgart, 1989.

[22] M.A. Lukas. Comparisons of parameter choice methods for regularization with discrete noisy data. *Inverse Problems*, 14:161–184, 1998.

[23] B. A. Mair. Tikhonov regularization for finitely and infinitely smoothing operators. *SIAM J. Math. Anal.*, 25(1):135–147, 1994.

[24] Bernard A. Mair and Frits H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[25] Peter Mathé. s-numbers in information-based complexity. *J. Complexity*, 6(1):41–66, 1990.

[26] V. A. Morozov. *Methods for solving incorrectly posed problems.* Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.

[27] Frank Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numer. Math.*, 28(3):329–341, 1977.

[28] Frank Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal.*, 18(1-2):29–37, 1984.

[29] Andreas Neubauer. An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM J. Numer. Anal.*, 25(6):1313–1326, 1988.

[30] M. Nussbaum. Degrees of ill-posedness in stochastic and deterministic noise models. in Algorithms and Complexity for Continuous Problems. Dagstuhl-Seminar-Report 101, 1994.

[31] M. Nussbaum and S. V. Pereverzev. The degree of ill-posedness in stochastic and deterministic noise models. WIAS Preprint No. 509, 1999.

[32] Douglas W. Nychka and Dennis D. Cox. Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.*, 17(2):556–572, 1989.

[33] S. V. Pereverzev and S. G. Solodky. An efficient discretization for solving ill-posed problems. In *The mathematics of numerical analysis (Park City, UT, 1995)*, pages 643–649. Amer. Math. Soc., Providence, RI, 1996.

[34] Albrecht Pietsch. *Operator ideals.* North-Holland Publishing Co., Amsterdam, 1980. Translated from German by the author.

[35] R. Plato and G. Vainikko. On the regularization of projection methods for solving ill-posed problems. *Numer. Math.*, 57(1):63–79, 1990.

[36] S. G. Samko, A. A. Kilbas, and O. I. Marichev. Интегралы и производные дробного порядка и некоторые их приложения. "Nauka i Tekhnika", Minsk, 1987. Edited and with a preface by S. M. Nikol′skiĭ.

[37] J. Saranen. A modified discrete spectral collocation method for first kind integral equations with logarithmic kernel. *J. Integral Equations Appl.*, 5(4):547–567, 1993.

[38] E. Schock. Ritz-regularization versus least-square-regularization. Solution methods for integral equations of the first kind. *Z. Anal. Anwendungen*, 4(3):277–284, 1985.

[39] P. Speckman. Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Math.*, 13:970–984, 1985.

[40] Ulrich Tautenhahn. Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.*, 33(6):2120–2130, 1996.

[41] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[42] Nicole Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals.* Longman Scientific & Technical, Harlow, 1989.

[43] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-based complexity.* Academic Press Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boult.

[44] A. B. Tsybakov. Pointwise and sup–norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469, 1998.

[45] G. M. Vaĭnikko and U. A. Khyamarik. Projection methods and self-regularization in ill-posed problems. *Izv. Vyssh. Uchebn. Zaved. Mat.*, (10):3–17, 84, 1985.

[46] G. M. Vaĭnikko and A. Yu. Veretennikov. Итерационные процедуры в некорректных задачах. "Nauka", Moscow, 1986.

[47] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667, 1977.

[48] Arthur G. Werschulz. *The computational complexity of differential and integral equations.* The Clarendon Press Oxford University Press, New York, 1991. An information-based approach, Oxford Science Publications.