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

Sparsity reconstruction by the standard Tikhonov method

S. Lu, S. Pereverzyev

RICAM-Report 2008-17

SPARSITY RECONSTRUCTION BY THE STANDARD TIKHONOV REGULARIZATION

SHUAI LU AND SERGEI V. PEREVERZEV

ABSTRACT. It is a common belief that Tikhonov scheme with $\|\cdot\|_{L_2}$-penalty fails to reconstruct a sparse structure with respect to a given system $\{\phi_i\}$. However, in this paper we present a procedure for sparsity reconstruction, which is totally based on the standard Tikhonov method. This procedure consists of two steps. At first Tikhonov scheme is used as a sieve to find the coefficients near ϕ_i, which are suspected to be non-zero. Within this step the performance of the standard Tikhonov method is controlled in some sparsity promoting space rather than in original Hilbert one. In the second step of proposed procedure the coefficients with indices selected in the previous step are estimated by means of data functional strategy. The choice of the regularization parameter is the crucial issue for both steps. We show that recently developed parameter choice rule called the balancing principle can be effectively used here. We also present the results of computational experiments giving the evidence of the reliability of our approach.

1. INTRODUCTION

In this paper, we will discuss a practically important problem of the recovery of an element of interest which has a sparse expansion with respect to a preassigned linearly independent system $\{\phi_i\}$. Such a problem often arise in scientific context, ranging from image reconstruction and restoration to wavelet denosing [6], to inverse bifurcation analysis [13].

In a rather general form the problem can be represented as an operator equation

$$Ax = y \tag{1.1}$$

with a linear operator $A \in \mathcal{L}(X, Y)$ acting between Hilbert spaces X and Y and having a non-closed range $R(A)$. This non-closedness is reflected in the discontinuity of the inverse operator A^{-1}, if it exists. In general, the generalized solution $A^\dagger y$, where $A^\dagger$ is the Moore-Penrose inverse of A, does not depend continuously on the right-hand side y. At the same time, in applications, usually only noisy data y^δ are available such that

$$\|y - y^\delta\|_Y \le \delta. \tag{1.2}$$

Then the problem of recovery of $A^\dagger y$ from noisy equation $Ax = y^\delta$ is ill-posed, and the task of solving it makes sense only when placed in an appropriate framework. Following Daubechies, Defrise and De Mol [6], we consider a linear inverse problem (1.1) where the solution $A^\dagger y$ is assumed to have a sparse structure. The focus in this problem is to recover $x^\dagger = A^\dagger y$ from (1.1), (1.2) under the assumption that it has a sparse expansion

$$x^\dagger = \sum_i \widehat{x}_i \phi_i \tag{1.3}$$

on the given system $\{\phi_i\}$. We define the sparsity of $x^\dagger$ by the presence of a small number $\sharp\{\widehat{x}_i \neq 0\}$ of large coefficients $\widehat{x}_i$ in (1.3) and zeroes elsewhere, although *a priori* we do not know either the number of non-zero coefficients, or their indices.

In contrary to the classical setting (c.f. [9]), in sparse reconstruction we need to recover the exact solution $x^\dagger$ as an element of some space Z_ρ promoting sparsity and equipped with an appropriate distance $\rho = \rho(u_1, u_2)$, $u_1, u_2 \in Z_\rho$. Several papers have been published recently on regularization in such space. We refer here to [3, 5, 6, 12, 21]. If, for example, $\{\phi_i\}$ is an orthonormal basis of X then following [6] one can take

$$\rho(u_1, u_2) = \|u_1 - u_2\|_p = \left(\sum_i |\langle u_1, \phi_i \rangle - \langle u_2, \phi_i \rangle|^p \right)^{1/p}. \tag{1.4}$$

where $\langle \cdot, \cdot \rangle$ is the inner product in X. It has been explained in [6] that for $1 \le p < 2$ the space equipped with such a distance really promotes sparsity. Moreover, it has been shown in [6] that a sparse structure of $A^\dagger y$ with respect to $\{\phi_i\}$ can be recovered by minimizing the functional

$$D_{\alpha,\rho}(x) = D_{\alpha,\rho}(A, y^\delta, \{\phi_i\}; x) = \|Ax - y^\delta\|_Y^2 + \alpha \|x\|_p^p, \tag{1.5}$$

and it has been mentioned that the sparsity-promoting feature of (1.5) is the more pronounced the smaller p is. Therefore, some application even use values of p with $0 < p < 1$. Since the distance (1.4) with $p < 1$ does not meet the triangle inequality, for $0 < p \le 1$ one usually uses a distance $\rho_p(u_1, u_2) := \|u_1 - u_2\|_p^p$ satisfying this inequality (see [8] for details). Nevertheless, in [6] the authors restrict themselves to $p \ge 1$ because the functional (1.5) ceases to be concave if $p < 1$. Note that even for $1 \le p < 2$ the minimization of (1.5) is not so easy. In [6] the functional (1.5) has been replaced by a sequence of surrogate functionals which are easier to minimize, and the bulk of [6] deals with an iterative algorithm to obtain minimizers for (1.5). At the same time, the quality of the recovery via minimizer of (1.5) depends on the choice of α. In [6] it has been suggested to choose $\alpha = \alpha(\delta)$ in such a way that $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$. Such a choice can only guarantee a convergence of the minimizer of (1.5) to $A^\dagger y$ in the norm of original Hilbert space X for vanishing noise level δ. Since a Hilbert space X does not promote sparsity, it is not clear how does the regularization by minimizing (1.5) compare with standard regularization techniques (which also provide convergence in X), and how α should be chosen in (1.5) to guarantee a reasonable sparsity reconstruction for a fixed noise level δ.

At this point it is worth to note that the reconstruction of a sparse structure is essentially the reconstruction of coefficients $\{\hat{x}_i\}$ in (1.3). For a system $\{\phi_i\}$ consisting of linear independent elements $\phi_i \in X$ each such a coefficient can be seen as a value of some linear functional $\hat{x}_i(x^\dagger)$ of the element $x^\dagger$, i.e. $\hat{x}_i := \langle l_i, x^\dagger \rangle$, where l_i is the generalized Ritz representer of $\hat{x}_i$ (distribution). For example, in the case of an orthonormal system $\{\phi_i\}$, $l_i = \phi_i$

From this viewpoint the sparsity reconstruction can be seen as the problem of indirect functional estimation. This problem has been extensively studied, and a few selected references are [1, 2, 7, 10, 16]. In particular, from the Corollary 3.1 of [2] it follows that the standard Tikhonov method estimating $\langle l_i, x^\dagger \rangle$ by

$$\langle l_i, x_\alpha^\delta \rangle = \langle l_i, (\alpha I + A^* A)^{-1} A^* y^\delta \rangle \tag{1.6}$$

is order optimal for a wide range of functionals l_i and elements $x^\dagger$, provided the regularization parameter α is chosen properly.

Note that the construction of a Tikhonov approximation x_α^δ, and a calculation of an estimation (1.6) for each individual l_i, are less computationally demanding than a minimization of (1.5). Of course, in this way one cannot estimate all coefficient $\hat{x}_i$ of an infinite series (1.3), but if the solution $x^\dagger$ admits a sparse representation (1.3) than

only a few of them are of interest. The indices of these non-zero coefficients are a priori unknown. Therefore, the idea is to use a standard Tikhonov approximation x_α^δ with an appropriate α for selecting the indices of "suspected" coefficients $\hat{x}_i$ which are above some threshold τ, and then estimate them more accurately using (1.6) with some other α.

Thus in this paper we are going to present a procedure for reconstructing a sparse structure which is totally based on the standard Tikhonov method. This procedure consists of two steps. At first Tikhonov scheme is used as a sieve to find coefficients which are suspected to be non-zero. Within this step the performance of the standard Tikhonov method is controlled in some sparsity promoting space rather than in original Hilbert space X. The examples of how it can be done are presented in the Section 3.

In the second step of proposed procedure the coefficients $\hat{x}_i = \langle l_i, x^\dagger \rangle$ with indices selected in the previous step are estimated by $\langle l_i, x_\alpha^\delta \rangle$. It is described in the Section 4. The choice of the regularization parameter α is the crucial issue for both steps. We show that recently developed parameter choice strategy called the balancing principle can be effectively used in each step.

2. The balancing principle

In numerical analysis there are many situation, where an element of interest $u^\dagger$ (solution of the problem or some functional of it) can be in principle approximated by an ideal element u_α depending on a positive parameter α in such a way that an appropriate distance $\rho(u^\dagger, u_\alpha)$ between them goes to zero as $\alpha \to 0$, i.e.

$$\lim_{\alpha \to 0} \rho(u^\dagger, u_\alpha) = 0. \tag{2.1}$$

In practice, however, this ideal element u_α is not available, because the data required for constructing u_α are given with error. As a result, we have at our disposal some element u_α^δ instead of u_α, where δ is a bound for the error in given data. In this paper first the role of u_α^δ will be played by a Tikhonov approximation, and then by $\langle l_i, x_\alpha^\delta \rangle$ estimating the coefficient $\hat{x}_i$ in (1.3).

In both above mentioned cases the stability of approximation u_α with respect to δ-perturbation in data can be described in the form of inequality

$$\rho(u_\alpha, u_\alpha^\delta) \le \psi(\alpha, \delta), \tag{2.2}$$

where $\psi(\alpha, \delta)$ is assumed to be a decreasing function of α.

On the other hand, in view of (2.1) one can always find a non-decreasing function φ such that $\varphi(0) = 0$ and for any positive α

$$\rho(u^\dagger, u_\alpha) \le \varphi(\alpha). \tag{2.3}$$

Using (2.2), (2.3) and the triangle inequality we obtain the following estimation

$$\rho(u^\dagger, u_\alpha^\dagger) \le \varphi(\alpha) + \psi(\alpha, \delta), \tag{2.4}$$

which tells us that a coordination between a parameter α, governing the approximation, and the amount of data error δ is required to obtain good accuracy.

In the ideal situation such a coordination could be achieved by the choice of α solving an equation $\varphi(\alpha) = \psi(\alpha, \delta)$. The point is that the best function φ measuring the rate of convergence in (2.3) is usually unknown.

Therefore, in practical applications different parameters $\alpha = \alpha_i$ are often selected from some finite set

$$\Sigma_N = \{\alpha_i : 0 < \alpha_1 < \alpha_2 < \ldots < \alpha_N\},$$

and corresponding elements $u^\delta_{\alpha_i}$, $i = 1, 2, \dots, N$, are studied on-line.

A parameter choice rule, called the balancing principle selects $\alpha = \alpha_+$ from Σ_N as follows

$$\alpha_+ = \max\left\{\alpha_i \in \Sigma_N : \forall j = 1, 2, \dots, i; \quad \rho(u^\delta_{\alpha_i}, u^\delta_{\alpha_j}) \le 4\psi(\alpha_j, \delta)\right\}. \tag{2.5}$$

To draw a conclusion from this parameter choice we consider all possible functions φ satisfying (2.3) and $\varphi(\alpha_1) < \psi(\alpha_1, \delta)$. Any of such functions is called admissible for $u^\dagger$ and ψ, and it can be used as a measure for the convergence rate in (2.1).

Then the Corollary 1 [18] provides us with the conclusion from the parameter choice (2.5) that can be drawn in the form of the following bound

$$\rho(u^\dagger, u^\delta_{\alpha_+}) \le 6D \min\{\varphi(\alpha) + \psi(\alpha, \delta), \quad \alpha \in \Sigma_N, \quad \varphi \text{ is admissible. }\}, \tag{2.6}$$

where the constant D depends only on ψ and Σ_N and is such that $\psi(\delta, \alpha_i) \le D\psi(\delta, \alpha_{i+1})$, $i = 1, 2, \dots, N-1$. Thus, the parameter choice $\alpha = \alpha_+$ allows us to reach (up to a constant factor $6D$) the best error bound of the form (2.4) that in principle can be obtained for $\alpha \in \Sigma_N$.

We would like to stress that the parameter choice strategy (2.5) is based on the function ψ alone. One does not need to know an admissible function corresponding to the best convergence rate in (2.1), while the information about the stability, as given by ψ in (2.2), is extremely important. The balancing principle (2.5) can be implemented in any metric space and for any regularization method provided such an information is available. In the next section we discuss how Monte Carlo simulation can be used for numerical estimation of the function ψ in the stability bound (2.2).

3. Discretized Tikhonov regularization for a rough sparsity reconstruction

It is worth to notice that in practice we are able to handle only a finite section of an expansion (1.3). Therefore, in reality one tries to recover a sparse structure of a projection $P_M x^\dagger = \sum_{i=1}^M \widehat{x}_i \phi_i$. Here P_M is the orthogonal projector from X onto $\mathrm{span}\{\phi_i\}_{i=1}^M$. Note that $P_M x^\dagger$ solves the equation $Ax = Ax^\dagger - A(I - P_M)x^\dagger$. Moreover, a system $\{\phi_i\}$ has usually a reasonable approximation property such that $\|(I - P_M)x^\dagger\|_X \to 0$ and $\|(I - P_M)A^*\|_{Y\to X} \to 0$ as $M \to \infty$. Then for sufficiently large M one has $\|(I - P_M)A^*\|_{Y\to X} \le \sqrt{\delta}$, $\|(I - P_M)x^\dagger\|_X \le \sqrt{\delta}$, and

$$\begin{aligned} \|y^\delta - AP_M x^\dagger\|_Y &= \|y^\delta - (Ax^\dagger - A(I - P_M)x^\dagger)\|_Y \\ &\le \|A(I - P_M)\|_{X\to Y}\|(I - P_M)x^\dagger\|_X + \|Ax^\dagger - y^\delta\|_Y \\ &\le \|(I - P_M)A^*\|_{Y\to X}\|(I - P_M)x^\dagger\|_X + \delta \le 2\delta. \end{aligned}$$

It means that for sufficiently large M a level of a noise in the right-hand side of the equation

$$AP_M x = y^\delta \tag{3.1}$$

is of the same order of magnitude as in y^δ, and it can be used for a recovery of the sparse structure in $P_M x^\dagger$ from (3.1).

To make the further discussion more concrete we consider an example, where A is the linear integral operator

$$Ax(t) = \int_0^1 a(t, s)x(s)ds, \qquad t \in [0, 1], \tag{3.2}$$

with the Green's function

$$a(t,s) = \begin{cases} t(1-s), & s \geq t, \\ s(1-t), & s \leq t, \end{cases}$$

as a kernel. In the inverse problems community this operator is frequently used as a prototype example (see, e.g., a recent paper [20] by Neubauer). Moreover, among orthonormal systems $\{\phi_i\}$ discussed in the paper [6] by Daubechies, Defrise and De Mol, we choose the simplest one, where $\phi_i = \phi_i^M(t)$ are L_2-orthonormalized characteristic functions of the intervals $[\frac{i-1}{M}, \frac{i}{M}], i = 1, 2, \ldots, M$.

Such a system appears in several application (see, for example, [13] and numerical experiment with image deblurring presented below).

Observe that for the system $\{\varphi_i^M\}$ and $0 < p \leq 1$ the sparsity promoting distance $\rho_p(u_1, u_2) = \|u_1 - u_2\|_p^p$, which appear in (1.4), (1.5) is equivalent, up to normalizing factor $C_p = M^{\frac{p-2}{p}}$, to the L_p-distance

$$\rho(u_1, u_2) = \|u_1 - u_2\|_{L_p} := \int_0^1 |u_1(t) - u_2(t)|^p dt. \tag{3.3}$$

Therefore, for $p \in (0, 1]$ the distance (3.3) also can be considered as a sparsity promoting one. The advantage of the distance (3.3) is that it can be computed independently on the number M of system elements.

In this section we apply the standard Tikhonov regularization to the discretized equation (3.1) and control the performance of this method in the space equipped with the distance (3.3) by means of the balancing principle (2.5). We argue that it allows a significant reduction in the number of coefficients $\hat{x}_i$ suspected to be non-zero in a sparse expansion (1.3).

Recall that applying the standard Tikhonov regularization to (3.1) we obtain a regularized approximation $x_{\alpha,M}^\delta$ that can be written as

$$x_{\alpha,M}^\delta = (\alpha I + P_M A^* A P_M)^{-1} P_M A^* y^\delta = \sum_{i=1}^M \widehat{x}_{i,M} \phi_i, \tag{3.4}$$

where the vector $x_M = (\widehat{x}_{1,M}, \widehat{x}_{2,M}, \ldots, \widehat{x}_{M,M})$ of coefficients solves a system of linear algebraic equations

$$\alpha x_M + B x_M = b_\delta \tag{3.5}$$

with a matrix $B = \{\langle A\phi_i, A\phi_j \rangle_Y\}_{i,j=1}^M$ and a vector

$$b_\delta = \Big(\langle A\phi_1, y^\delta \rangle_Y, \langle A\phi_2, y^\delta \rangle_Y, \ldots, \langle A\phi_M, y^\delta \rangle_Y \Big)$$

($\langle \cdot, \cdot \rangle_Y$ is the inner product in a Hilbert space Y). It remains to choose α. The reader is encouraged to consult [9] for more detailed information on discretized Tikhonov regularization.

We describe now how the Monte Carlo approach can be used for estimating $\psi(\alpha, \delta)$ in (2.2), where $u_\alpha^\delta = x_{\alpha,M}^\delta$, and $u_\alpha = x_{\alpha,M} = x_{\alpha,M}^0$.

In view of (1.2) noisy data y^δ can be represented as $y^\delta = y + \delta\xi$, where $\|\xi\|_Y \leq 1$. Then for $\rho(u, v) = \|u - v\|_{L_p}$, $0 < p \leq 1$, the function

$$\begin{aligned} \psi(\alpha, \delta) &= \sup \int_0^1 |x_{\alpha,M}^\delta(t) - x_{\alpha,M}(t)|^p dt \\ &= \delta^p \sup_{\|\xi\| \leq 1} \|(\alpha I + P_M A^* A P_M)^{-1} P_M A^* \xi\|_{L_p} \end{aligned} \tag{3.6}$$

can be taken as a stability bound in (2.2), and the Monte Carlo approach can be used to estimate the last sup numerically. This approach can be implemented, for example, as follows.

At first one should choose a system $\{w_k\}_{k=1}^n \subset Y$ and simulate vectors $\overline{\xi_j} = (\xi_{k,j})_{k=1}^n \in R^n$, $j = 1, 2, \dots, T$, with uniformly distributed random components normalized in such a way that

$$\left\| \sum_{k=1}^{n} \xi_{k,j} w_k \right\|_Y = 1.$$

Then Monte Carlo estimate for the stability bound (3.6) can be constructed as

$$\begin{aligned} \psi(\alpha,\delta) &= \psi_{max}(\alpha,\delta) \\ &= \delta^p \max_{j=1,2,\dots,T} \|(\alpha I + P_m A^* A P_m)^{-1} P_m A^* \xi_j\|_{L_p}, \\ &\quad \alpha \in \Sigma_N, \quad m \le M, \end{aligned} \tag{3.7}$$

where

$$\xi_j = \sum_{k=1}^{n} \xi_{k,j} w_k, \quad j = 1, 2, \dots, T. \tag{3.8}$$

Another possibility is to take

$$\begin{aligned} \psi(\alpha,\delta) &= \psi_{mean}(\alpha,\delta) \\ &= \delta^p T^{-1} \sum_{j=1}^{T} \|(\alpha I + P_m A^* A P_m)^{-1} P_m A^* \xi_j\|_{L_p}. \end{aligned} \tag{3.9}$$

Both these estimates are numerically feasible and the only issue is the choice of the system $\{w_k\}$.

It is natural to take $\{w_k\}$ from the singular value decomposition of the problem operator A, i.e.

$$A = \sum_{k=1}^{\infty} s_k(A) \langle u_k, \cdot \rangle_X w_k. \tag{3.10}$$

The reason is that a noise ξ enters the bound (3.6) only through the operator A^* such that

$$A^* \xi = \sum_{k=1}^{\infty} s_k(A) \langle w_k, \xi \rangle_Y u_k,$$

which means that only the coefficients $\langle w_k, \xi \rangle_Y$ influence the stability bound $\psi(\alpha, \delta)$. Therefore, simulating the noise in the form of (3.8) with $\{w_k\}$ from (3.10) one obtains an adequate noise model.

There is another noise model that also seems to be suitable for the problem of sparsity reconstruction, especially when the elements forming SVD (3.10) are not available.

Trying to reconstruct a sparse structure with respect to a system $\{\phi_i\}$ one can restrict the image space of A to the subspace $span\{A\phi_i\}$ of linear combinations of $\{A\phi_i\}$, since only such combinations can appear when A acts on elements of the form (1.3). Of course, the noise model (3.8) with $w_k = A\phi_k$ allows a reduction of the ill-posedness of the equation $Ax = y^\delta$ in X because for such a noise the data y^δ always belong to the image space of A. But, as we will see below, in sparsity reconstruction one is more interested in indices of non-zero coefficients in (1.3) than in an approximation of $x^\dagger$ in X-norm. Therefore, a noise (3.8) with $w_k = A\phi_k$, in spite of its smoothness, can

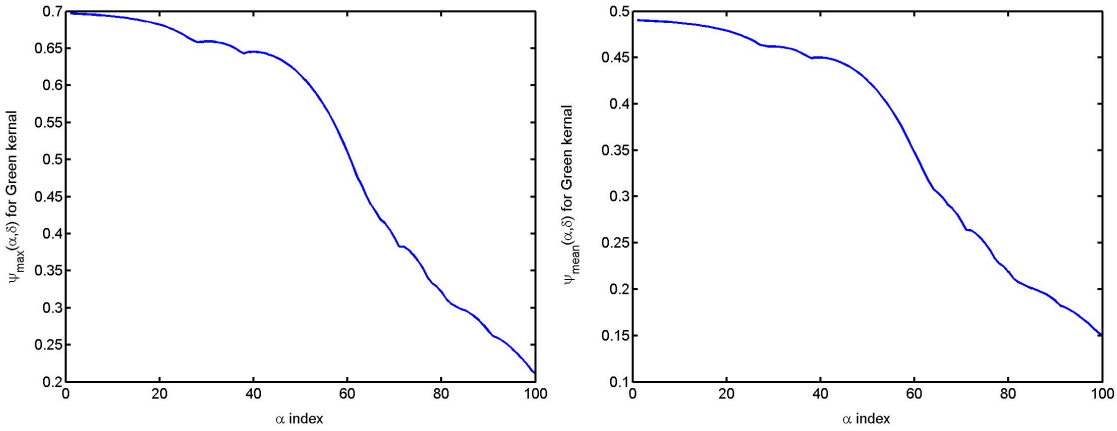

FIGURE 1. Monte Carlo estimates $\psi_{max}(\alpha,\delta)$ (left) and $\psi_{mean}(\alpha,\delta)$ (right) of the stability bound in $L_{1/2}$ for the operator (3.2) and the system $\{\phi_i^{25}\}_{i=1}^{25}$ of L_2-orthonormalized piece-wise constant functions with jumps at $t_i = i/25$, $i = 1, 2, \ldots, 25$.

essentially blur a sparse structure of $x^\dagger$. In our numerical experiments we construct estimates (3.7) and (3.9) using both above mentioned noise models.

It is known that the operator (3.2) admits the following singular value decomposition

$$A = \sum_{k=1}^{\infty} \frac{2}{(\pi k)^2} \sin(\pi k t) \langle \sin(\pi k t), \cdot \rangle_{L_2(0,1)} \tag{3.11}$$

that allows a use of the noise model (3.8) with $w_k(t) = \sqrt{2}\sin(\pi k t)$. Corresponding Monte Carlo estimates (3.7) and (3.9) for $\delta = 10^{-4}$, $p = 1/2$, $m = n = 25$, $T = 10$ are plotted in Figure 1 against the indices of $\alpha_i \in \Sigma_{100} = \{\alpha_j = \alpha_1 q^{j-1}, j = 1, 2, \ldots, 100, \alpha_1 = 10^{-10}, q = 1.1\}$.

Although these estimates of the stability bound have been obtained for the system $\{\phi_i^{25}\}_{i=1}^{25}$, they allow a reconstruction of the sparse structure with respect to other systems such as $\{\phi_i^{50}\}_{i=1}^{50}$, for example. It can been seen from Figure 2, where the exact solutions $x^\dagger = 3\phi_{13}^{50} + 3\phi_{35}^{50}$ and $x^\dagger = 6\phi_{10}^{25} + 7\phi_{12}^{25} + 8\phi_{14}^{25}$ (dashed lines) are displayed together with their Tikhonov's approximations $x_{\alpha,50}^\delta$, $\alpha = \alpha_{60} = 3.0448 \times 10^{-8}$, and $x_{\alpha,25}^\delta$, $\alpha = \alpha_{12} = 3.1384 \times 10^{-10}$ (solid lines). The regularization parameters have been chosen here in accordance with the balancing principle (2.5) corresponding to $L_{1/2}$-distance and $\psi_{mean}(\alpha,\delta)$ as in Figure 1 (right). In both cases Tikhonov's approximations $x_{\alpha_+,M}^\delta$ hint at a sparse structure.

Now we present the results of numerical experiments which show that the form of the bound $\psi(\alpha,\delta)$ for sparsity promoting spaces L_p, $0 < p \le 1$, is really operator- and system-dependent.

To this end we consider the Abel integral operator

$$Ax(t) = \int_0^t \frac{x(s)}{\sqrt{t-s}} ds, \quad t \in [0,1], \tag{3.12}$$

which is also used in the inverse problems theory as a prototype example (see, e.g., [9]). Moreover, we also change a system $\{\phi_i\}$ and consider the recovery of a sparse structure

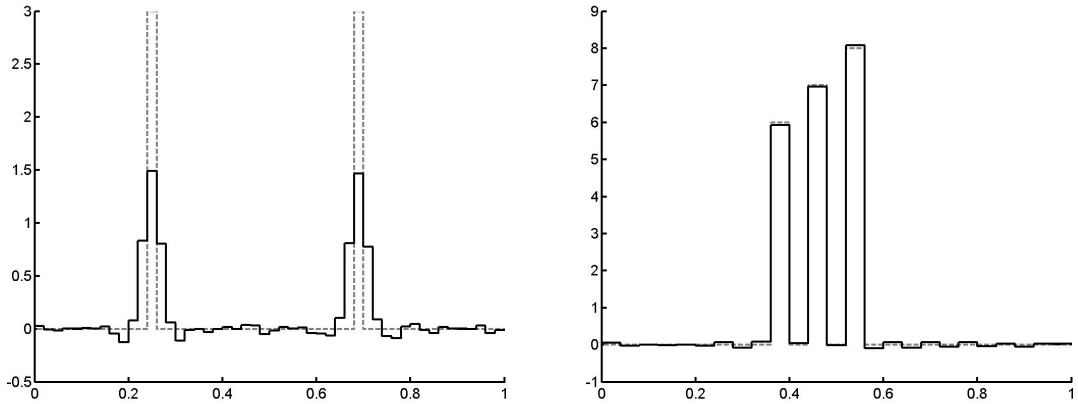

FIGURE 2. Orthonormal basis: reconstruction with the stability estimation displayed in Figure 1 (right) used in the standard Tikhonov regularization. The exact solutions are $x^\dagger = 3\phi_{13}^{50} + 3\phi_{35}^{50}$ (left) and $x^\dagger = 6\phi_{10}^{25} + 7\phi_{12}^{25} + 8\phi_{14}^{25}$ (right). In both figures, the dashed line is the exact solution and the solid line is the reconstruction, vertical axes have the scales $\sqrt{M}$.

with respect to the system of piece-wise linear B-splines

$$\phi_i(t) = \phi_i^M(t) = \begin{cases} M(t - \frac{i-1}{M}), & t \in [\frac{i-1}{M}, \frac{i}{M}], \\ M(\frac{i+1}{M} - t), & t \in [\frac{i}{M}, \frac{i+1}{M}], \\ 0, & t \notin [\frac{i-1}{M}, \frac{i+1}{M}], \end{cases}$$
$$i = 1, 2, \dots, M-1.$$

This system is also discussed in the context of a sparsity recovery (see, e.g., the dissertation [14] by Malioutov). It is not an orthogonal system, but the version (3.4), (3.5) of the ordinary Tikhonov regularization can be also used in the considered case without changes. We just need a stability estimation to implement the balancing principle (2.5) with L_p-distance for $u_{\alpha_i}^\delta = x_{\alpha_i,M}^\delta$, $\alpha_i \in \Sigma_{100} = \{\alpha_i = \alpha_0 \times (1.2)^i, \alpha_0 = 10^{-8}, i = 1, 2, \dots, 100\}$.

Keeping in mind that for the Abel integral operator an analytical form of the singular value decomposition is unknown, we follow the reason presented above and calculate the Monte Carlo estimates (3.7), (3.9) for $p = 1/2$ and $\delta = 0.02$ using the noise model (3.8) with $w_k = A\phi_k^n$, $k = 1, 2, \dots, n$. Corresponding graphs are presented in Figure 3.

To test a reliability of these estimates of $L_{1/2}$-stability we incorporate them into the balancing principle (2.5) and use it for recovering a sparse structure with respect to other system of B-splines $\{\phi_i^{100}\}$ (recall that the estimates were obtained for $\{\phi_i^{25}\}$).

Typical results are presented in Figure 4, where the graph of the exact solution $x^\dagger = 3\phi_{38}^{100} + 4\phi_{40}^{100} + 3\phi_{72}^{100}$ is display together with its Tikhonov approximation $x_{\alpha,100}^\delta$, $\alpha = \alpha_{82}$ and $\alpha = \alpha_{77}$. Here $\delta = 0.02$ and the regularization parameters have been chosen from Σ_{100} in accordance with the balancing principle based on $\psi(\alpha, \delta) = \psi_{max}(\alpha, \delta)$ (Figure 4, left) and $\psi(\alpha, \delta) = \psi_{mean}(\alpha, \delta)$ (Figure 4, right). Note that the test presented in Figure 4 is rather hard, since the modes ϕ_{38}^{100} and ϕ_{40}^{100} are very close to each other (narrow band problem). Nevertheless, the reconstruction given by the standard Tikhonov scheme is of the same quality as in the tests by Malioutov [14] (see Fig.4.1-4.3 there), where a

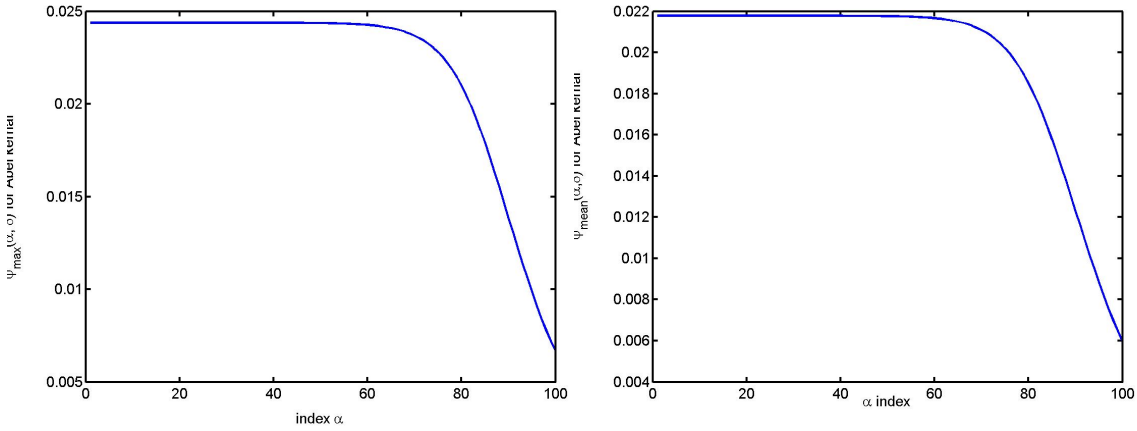

FIGURE 3. Monte Carlo estimates $\psi_{max}(\alpha,\delta)$ (left) and $\psi_{mean}(\alpha,\delta)$ (right) of the stability bound in $L_{1/2}$ for the Abel integral operator and the system $\{\phi_i^{25}\}_{i=1}^{25}$ of piece-wise linear B-splines with the knots at $t_i = i/25$, $i = 1, 2, \ldots, 24$.

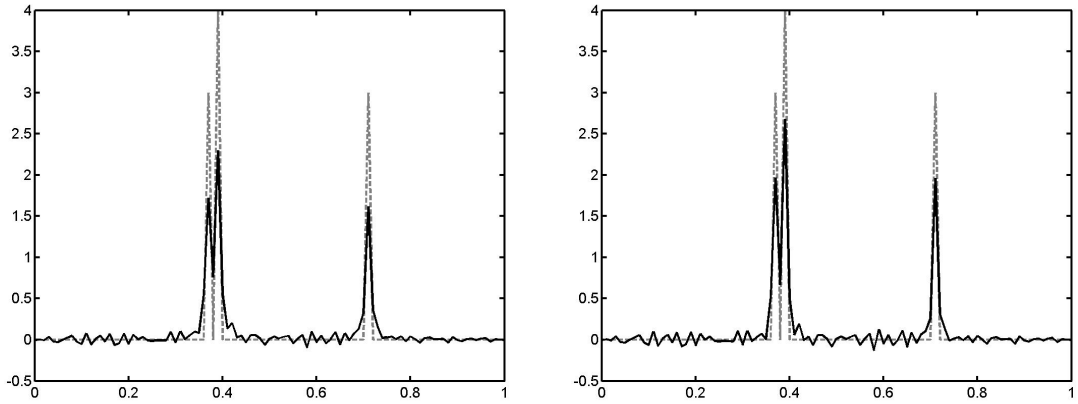

FIGURE 4. Reconstruction of the solution $x^\dagger = 3\phi_{38}^{100} + 4\phi_{40}^{100} + 3\phi_{72}^{100}$ (dashed line) of Abel integral equation obtained by means of the standard Tikhonov regularization. Regularization parameters $\alpha = \alpha_{82} = 0.0311$ (left) and $\alpha = \alpha_{77} = 0.0125$ (right) for the approximate solution $x_{\alpha,100}^\delta$ are chosen in accordance with (2.5) for $\psi(\alpha,\delta) = \psi_{max}(\alpha,\delta)$ and $\psi(\alpha,\delta) = \psi_{mean}(\alpha,\delta)$ respectively.

regularization via minimization of a Tikhonov type functional with l_1-penalty $\sum |\widehat{x}_i|$ has been used.

The stability bounds displayed in the Figures 1 and 3 are essentially different. They have been obtained for two operator equations regularized in the same space $L_{1/2}$. Of course, they are of a numerical origin, but in the combination with the balancing principle they seem to be reliable. So, if they are so different for two model problems then, in contrast to the classical Hilbert space setting, one can not rely on a problem independent stability bound when dealing with a regularization in L_p, $0 < p \leq 1$.

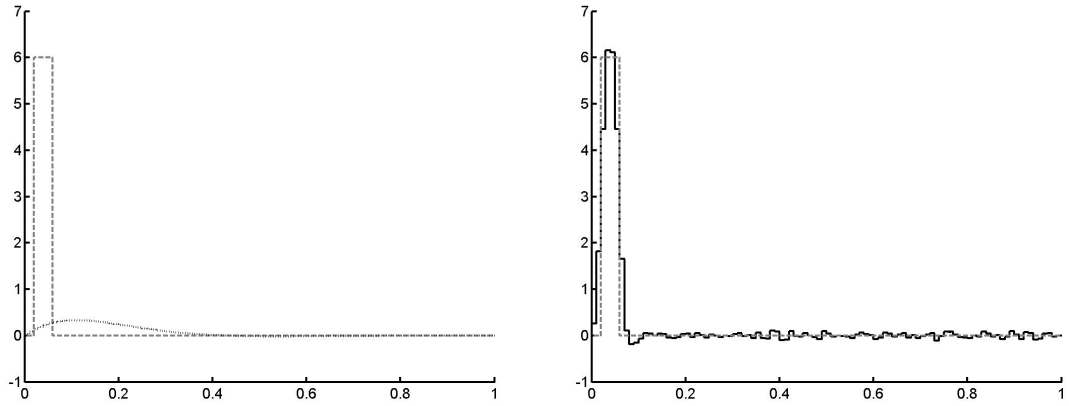

FIGURE 5. Orthonormal basis: comparison between Tikhonov approximations corresponding to different stability estimations with $x^\dagger = 6(\phi_2^{50} + \phi_3^{50})$. Dashed line is the exact solution. Dotted line in left figure is the reconstruction under a L_2 stability bound; solid line in right figure is the reconstruction under a stability bound given by Monte Carlo simulation, vertical axis has the scale $\sqrt{2M} = 10$.

Remark 3.1. To provide an evidence of the reliability of Monte Carlo approach to the stability estimation we can show the results of simulation for estimating

$$\begin{aligned} \psi(\alpha,\delta) &= \|x^\delta_{\alpha,M} - x_{\alpha,M}\|_{L_2} \\ &= \delta \sup_{\|\xi\|_{L_2}\le 1} \|(\alpha I + P_M A^* A P_M)^{-1} P_M A^* \xi\|_{L_2}, \end{aligned} \tag{3.13}$$

where the theory gives us the bound

$$\|x^\delta_{\alpha,M} - x_{\alpha,M}\|_{L_2} \le \psi(\alpha,\delta) := c\frac{\delta}{\sqrt{\alpha}}, \tag{3.14}$$

which is valid for a wide variety of regularization methods including the Tikhonov one. We refer to Ch.4 of the book [9] by Engl, Hanke and Neubauer for further details concerning the dependence of the constant c in (3.14) on concrete method.

In Figure 6 we present the Monte Carlo estimate for (3.13) plotted against the indices of $\alpha_i \in \Sigma_{100}$, where $\delta = 10^{-4}$, $M = 25$, and the noise model (3.8) with $w_k(t) = \sqrt{2}\sin(\pi k t)$ is used. The operators A and P_M are the same as in our first experiment.

In Figure 6 one can easily recognized the graph of the function $\psi(\alpha,\delta)$ from (3.14) with $c = 1, \delta = 10^{-4}$, $\alpha \in \Sigma_{100}$. Thus, in case of L_2-distance the Monte Carlo approach described above produces a stability bound that is in agreement with the theory, and it can be seen as an evidence of its reliability in the situations, where no theory is available.

At the same time, the standard Tikhonov method with a regularization parameter chosen in accordance with the balancing principle (2.5) implemented for $\rho(u,v) = \|u - v\|_{L_2}$ does not allow the reconstruction of a sparse structure. It can be seen from Figure 5 (left) displaying the graph (dashed line) of the exact solution $x^\dagger = P_M x^\dagger = 6(\varphi_2^{50} + \varphi_3^{50})$ together with the graph (dotted line) of $x^\delta_{\alpha,50}$ given by (3.4), where perturbed data y^δ corresponds to $\delta = 10^{-4}$ (a noise is simulated as in our first experiment), and $\alpha = 1.3781 \times 10^{-6}$ is chosen from Σ_{100} in accordance with (2.5) for $\rho(u,v) = \|u - v\|_{L_2}$ and

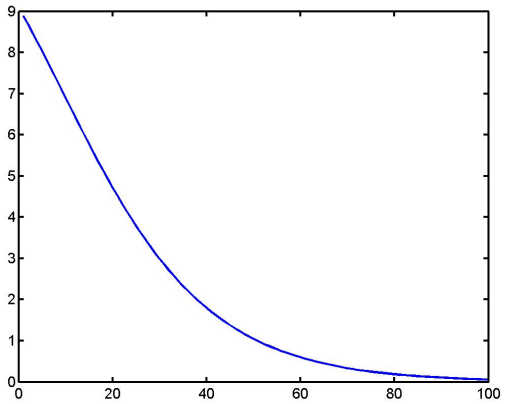

FIGURE 6. The plot of $\psi(\alpha,\delta)$ from L_2-stability bound $\|x^\delta_{\alpha,M} - x_{\alpha,M}\|_{L_2} \le \psi(\alpha,\delta)$ given by Monte Carlo simulation for $\delta = 10^{-4}, \alpha \in \Sigma_{100}$, $M = 25$, and the operator (3.2).

$\psi(\alpha,\delta) = \delta\alpha^{-1/2}$ as in (3.14). It is clear that no sparse structure can be reconstructed from such $x^\delta_{\alpha,M}$. By the way, a similar situation appears in the case of α chosen in accordance with the classical discrepancy principle as

$$\alpha = \sup\{\alpha > 0 : \|Ax^\delta_{\alpha,M} - y^\delta\|_{L_2} \le c\delta\}, \tag{3.15}$$

where c is some fixed constant.

At the same time, in Figure 5 (right) one can also see the graph (solid line) of $x^\delta_{\alpha,50}$ with $\alpha = 2.8102 \times 10^{-9}$ chosen from Σ_{100} in accordance with (2.5), where a stability bound $\psi(\alpha,\delta)$ is found using Monte Carlo approach for $\rho(u,v) = \|u - v\|_{L_1}$. This time $x^\delta_{\alpha,M}$ hints at a sparse structure.

Remark 3.2. Some of our numerical experiments hint that Monte Carlo estimations of the stability bound (3.6) can be used for a priori assessment of the efficiency of the standard Tikhonov method in sparsity reconstruction.

For example, one of our tests was performed for operator (3.2) and the system of piece-wise linear B-splines. It happened that for $p = 1$ the Monte Carlo estimation of the stability bound (3.6) was of the same order $\delta/\sqrt{\alpha}$ as the obvious estimation of L_1-stability via L_2-stability (see (3.14)):

$$\|x^\delta_{\alpha,M} - x_{\alpha,M}\|_{L_1} \le \|x^\delta_{\alpha,M} - x_{\alpha,M}\|_{L_2} \le c\frac{\delta}{\sqrt{\alpha}}.$$

Ad hoc interpretation was that in such a case the choice of the regularization parameter α would not be able to force the Tikhonov method to perform in L_1 better than in L_2. This a priori conclusion was confirmed by numerical tests, where Tikhonov method exhibited a poor performance, similar to the Figure 5 (left). So, if for given operator A and system $\{\phi_i\}$ a Monte Carlo estimation of L_1-stability bound (3.6) is of order $\delta/\sqrt{\alpha}$ then the standard Tikhonov method fails to reconstruct a sparse structure with respect to $\{\phi_i\}$.

At the end of the section we present a numerical experiment with two-dimensional deblurring problem to demonstrate that a combination of the standard Tikhonov method

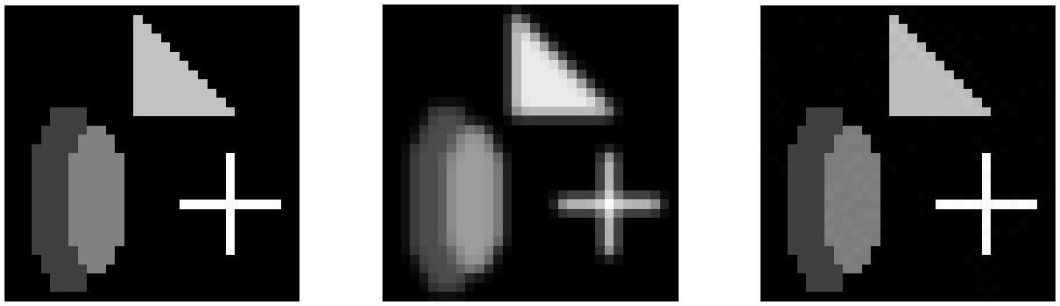

FIGURE 7. 2-Dimensional test for $M = 1024$, real image (left), blurred image (middle) and the reconstruction (right).

with the balancing principle implemented in a sparsity promoting space can be used for the recovery of sparse solutions of severely ill-posed multidimensional problems.

Recall that the image deblurring problem consists in the reconstruction of a so-called brightness function $x(t,\tau)$ of original digital image from the brightness function $y(t,\tau)$ of a blurred one. In Figure 7 we present a test example of the deblurring problem borrowed from [11]. In this figure the left picture is the original image, while a blurred one can be seen in the middle. The brightness function $x(t,\tau)$ is a piece-wise constant. It takes the value 4 at white picksels, the value 0 at black pixels, and the values between 0 and 4 at gray pixels. So, for an image located in the domain $\Omega = [0,1] \times [0,1]$ and formed by $M = m^2$ pixels this function admits a sparse expansion of the form (1.3) on the orthonormal system of box function

$$\varphi_i(t,\tau) = \varphi_i^M(t,\tau) = \varphi_k^m(t)\varphi_l^m(\tau), \qquad i = (k-1)m + l \\ k,l = 1,2,\dots,m, \quad i = 1,2\dots,M, \tag{3.16}$$

where $\varphi_k^m(t)$, $\varphi_l^m(\tau)$ are L_2-orthonormalized characteristic functions of the intervals $[\frac{k-1}{m}, \frac{k}{m}]$. The brightness function of the image displayed in the Figure 7 (left) admits a sparse expansion on the system (3.16) with $m = 2^5$, $M = 2^{10}$.

Following [11] we simulate the blurring process as a convolution with the Gaussian point-spread function

$$a(u,v) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2+v^2}{2\sigma^2}\right), \qquad \sigma = 0.7,$$

i.e. the brightness functions $x(t,\tau)$ and $y(t,\tau)$ are assumed to be related by the equation

$$Ax(t,\tau) := \int_\Omega a(t-u,\tau-v)x(u,v)dudv = y(t,\tau). \tag{3.17}$$

Note that in (3.17) the kernel of the operator A is an analytic function, while the solution $x(t,\tau)$ is expected to be a piece-wise constant function. Thus, the problem (3.17) is a first kind Fredholm integral equation with an analytic kernel and a discontinuous solution. Such a problem is known to be severely ill-posed.

Nevertheless, a sparse structure of the solution of this severely ill-posed problem can be reconstructed by means of the standard Tikhonov method in the same way as it has been described above.

To show this we at first obtain an estimation of the stability bound (3.6) using Monte Carlo approach. This time we are interested in (3.6) with $p = 1$ and use the noise model (3.8) with $w_k = \varphi_k^M(t,\tau)$, where $M = m^2 = 2^8$. As a result, we obtain the functions displayed in the Figure 8.

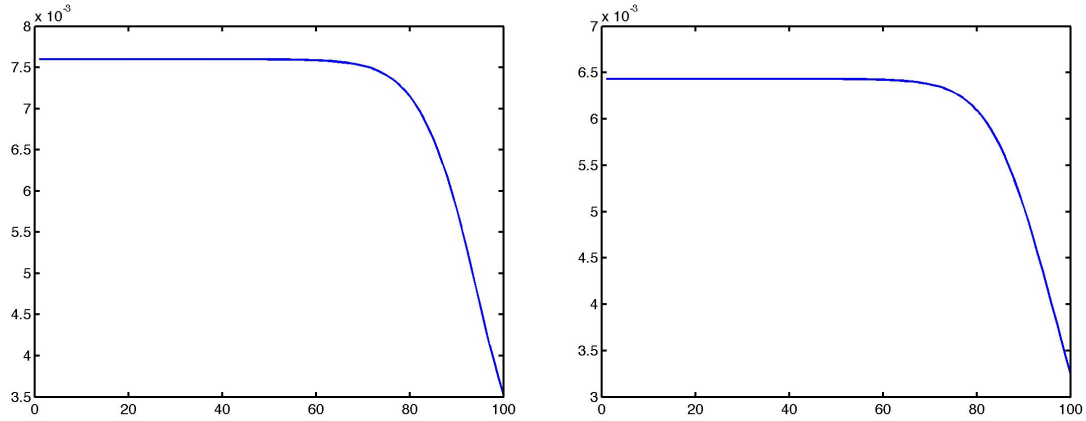

FIGURE 8. The plots of $\psi_{max}(\delta,\alpha)$ (left) and $\psi_{mean}(\alpha,\delta)$ (right) for the operator A from (3.17) and the system (3.16); $\delta = 10^{-3}$, $M = 16^2$.

Noisy data y^δ are simulated in the form

$$y^\delta(t,\tau) = \sum_{i=1}^{M}(y_i + \delta\xi_i)\varphi_i^M(t,\tau), \quad M = 2^{10}, \delta = 10^{-3},$$

where $(y_1,\dots,y_M)$ is the columnwise stacked version of the blurred image, and $(\xi_1,\dots,\xi_M)$ is a normally distributed random vector with zero mean and the standard deviation 1.

Using these noisy data, we construct Tikhonov regularized solution (3.4) and choose a regularization parameter $\alpha \in \Sigma_{100}$ in accordance with the balancing principle (2.5), where $u_\alpha^\delta = x_{\alpha,M}^\delta(t,\tau)$, $\rho(u,v)$ means L_1-distance, and the stability bound $\psi(\alpha,\delta)$ is displayed in the Figure 8 (right).

It gives us the value $\alpha = \alpha_{82} = 2.5923 \times 10^{-4}$. The image corresponding to the reconstructed brightness function $x_{\alpha_{82},M}^\delta$, $M = 2^{10}$, can be seen in the Figure 7 (right). The reconstruction is of good quality, the maximal value of $|x_{\alpha_{82},M}^\delta - x|$ is 0.062520, where $x(t,\tau)$ is the brightness function of the original image displayed in Figure 7 (left).

This experiment shows that in principle the standard Tikhonov method with properly chosen regularization parameter can be used for a sparsity reconstruction even in case of severely ill-posed problem.

4. Local regularization

Recall that for a system $\{\varphi_i\}$ consisting of linear independent elements each coefficient $\hat{x}_i$ in (1.3) is a value of some linear functional defined on a Hilbert space X as $\hat{x}_i = \langle l_i, x^\dagger\rangle$. For example, in the case of the system $\{\varphi_i^M\}$ of piece-wise linear B-splines we have $l_i = \delta_{i/M}$, where δ_t is a Dirac delta function concentrated at point t, i.e. $\hat{x}_i = \langle \delta_{i/M}, x^\dagger\rangle = x^\dagger(\frac{i}{m})$.

Numerical experiments presented in the Section 3 show that the standard Tikhonov regularization method equipped with L_p-balancing principle, $0 < p \le 1$, can be used as a sieve to find "suspected" coefficients $\hat{x}_i$ which are above some threshold τ. For example, using Figure 2 (left) one can easily realize that there are only a few suspected coefficients and their indices are $i = 12, 13, 14$ and $i = 34, 35, 36$.

Let $\widehat{x}_i = \langle l_i, x^\dagger \rangle$ be one of such "suspected" coefficients. It means that

$$\left| \langle l_i, x^\delta_{\alpha_+,M} \rangle \right| \geq \tau, \tag{4.1}$$

where $x^\delta_{\alpha_+,M}$ is a Tikhonov regularized approximation (3.4) corresponding to a regularization parameter $\alpha = \alpha_+$ selected in accordance with L_p-balancing principle (2.5) for an appropriate stability bound $\psi(\alpha, \delta)$.

On the other hand, using the standard Tikhonov method one can estimate $\widehat{x}_i$ by

$$\langle l_i, x^\delta_\beta \rangle = \langle l_i, (\beta I + A^*A)^{-1} A^* y^\delta \rangle, \quad \beta \in \Sigma_N. \tag{4.2}$$

Then

$$\widehat{x}_i - \langle l_i, x^\delta_\beta \rangle = \langle l_i, , x^\dagger - x_\beta \rangle + \langle l_i, x_\beta - x^\delta_\beta \rangle, \tag{4.3}$$

where x_β is an ideal Tikhonov approximation corresponding to noise free data $y = Ax^\dagger$. It is known that $\|x^\dagger - x_\beta\|_X \to 0$ as $\beta \to 0$. Then under rather general assumptions the first term in (4.3) also converges to zero, and there exists a non-decreasing continuous admissible function $\varphi_{l_i} : [0, \alpha_N] \to [0, \infty)$ such that $\varphi_{l_i}(0) = 0$ and for any $\beta \in [0, \alpha_N]$

$$\left| \langle l_i, x^\dagger - x_\beta \rangle \right| \leq \varphi_{l_i}(\beta). \tag{4.4}$$

To estimate the second term in (4.3) we note that

$$\langle l_i, x_\beta - x^\delta_\beta \rangle = \langle A(\beta I + A^*A)^{-1} l_i, y - y^\delta \rangle_Y.$$

Then in view of (1.2) and the obvious relation

$$\sup_{y_\delta : \|y - y_\delta\|_Y \leq \delta} \langle A(\beta I + A^*A)^{-1} l_i, y - y^\delta \rangle_Y = \delta \|(\beta I + AA^*)^{-1} A l_i\|_Y$$

the best possible bound for the second term is given by the inequality

$$\left| \langle l_i, x_\beta - x^\delta_\beta \rangle \right| \leq \psi_{l_i}(\beta, \delta), \tag{4.5}$$

where

$$\psi_{l_i}(\beta, \delta) = \delta \|(\beta I + AA^*)^{-1} A l_i\|_Y.$$

For each concrete functional l_i the values of the function $\psi_{l_i}(\beta, \delta)$ at the points $\beta \in \Sigma_N$ can be easily found either numerically or analytically. Just to give an example, in Figure 9 we plot the values of this function for $X = Y = L_2(0, 1)$, and for two operators (3.2), (3.12), and typical coefficient-functionals $\widehat{x}_i = \langle \phi_i^{50}, x^\dagger \rangle$, $\widehat{x}_i = \langle \delta_{i/100}, x^\dagger \rangle = x^\dagger(\frac{i}{100})$, $i = 2$, discussed above. In both cases the values have been obtained numerically for discretized operators $P_M A P_M$.

With the function $\psi_{l_i}(\beta, \delta)$ in hand one can easily reformulate the balancing principle (2.5) for choosing a regularization parameter β in estimating the values $\widehat{x}_i$ of the linear functional l_i at $x^\dagger$. The parameter of choice is β_+ defined as follows

$$\begin{aligned} \beta_+ = \beta^i_+ \quad = \quad & \max\{\beta \in \Sigma_N : \forall \alpha \in \Sigma_N, \alpha < \beta, \\ & |\langle l_i, x^\delta_\beta \rangle - \langle l_i, x^\delta_\alpha \rangle| \leq 4 \psi_{l_i}(\alpha, \delta)\}. \end{aligned} \tag{4.6}$$

Note that instead of computing Tikhonov approximations x^δ_β and then evaluating $\langle l_i, x^\delta_\beta \rangle$ for all $\beta \in \Sigma_N$ one can precompute data-functionals $z^i_\beta = (\beta I + AA^*)^{-1} A l_i$ in advance, use them to calculate the bound $\psi_{l_i}(\beta, \delta) = \delta \|z^i_\beta\|_Y$, and then find estimates $\langle l_i, x^\delta_\beta \rangle$ for the values of $\widehat{x}_i$ applying z_β directly to the data y^δ, since $\langle l_i, x^\delta_\beta \rangle_X = \langle z^i_\beta, y^\delta \rangle_Y$.

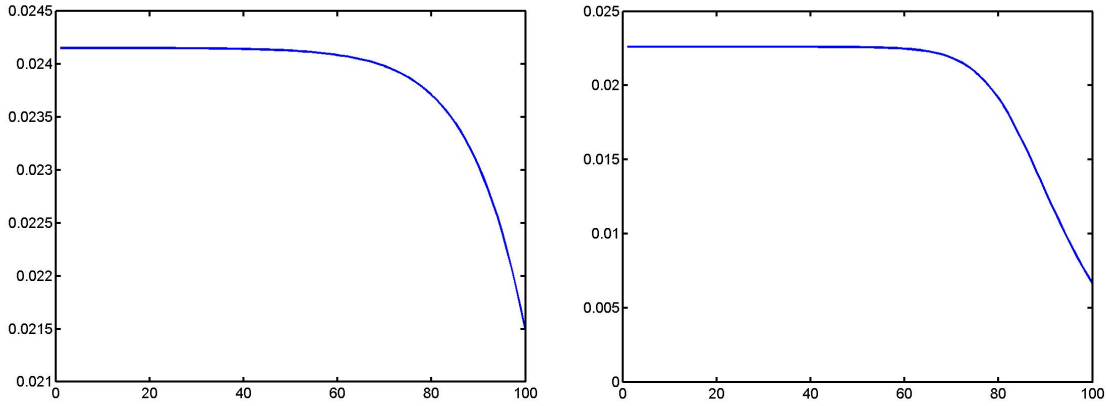

FIGURE 9. The plots of $\psi_{l_i}(\beta,\delta)$, $i=2$, for the operator (3.2), $l_i=\phi_i^{50}$, $\delta=10^{-4}$ (left); and for Abel operator, $l_i=\delta_{i/100}$, $\delta=0.02$ (right).

This approach called data-functional strategy was proposed in [1] and studied in [2, 10, 16]. The following optimality property of the parameter choice (4.6) can be proven in the same way as in Theorem 2.1 [2].

Theorem 4.1. *Assume that $\psi_{l_i}(\beta,\delta)$ decreases at most at a power rates, i.e. $\delta\beta^{-r_1}\le\psi_{l_i}(\beta,\delta)\le\delta\beta^{-r_2}$ for some r_1, $r_2>0$ and $\beta\in(0,\alpha_N]$. Then for $\widehat{x}_i=\langle l_i,x^\dagger\rangle$*

$$|\widehat{x}_i-\langle z^i_{\beta_+},y^\delta\rangle|\le c\min\{\varphi_{l_i}(\beta)+\psi_{l_i}(\beta,\delta),\quad \beta\in\Sigma_N,\quad \varphi_{l_i}\ \textit{is admissible}\},$$

where c depends only on l_i and $x^\dagger$.

Thus, the parameter choice rule (4.6) is capable of achieving the order-optimal balance between unknown convergence rate in (4.4) and known rate of the noise propagation $\psi_{l_i}(\beta,\delta)$.

In course of our discussion we have presented a procedure for reconstructing a sparse structure which is totally based on the standard Tikhonov method. This procedure consists of two steps. At first Tikhonov scheme, equipped with L_p-balancing principle, $0<p\le 1$, and with the threshold rule (4.1), is used to select the indices of the coefficients $\widehat{x}_i$ which are supposed to be non-zero. Then Tikhonov scheme is used within the framework of data-functional strategy to estimate selected coefficients individually. This time it is equipped with the balancing principle given in the form (4.6).

It is easy to realize that the performance of this two steps procedure very much depends on the choice of the threshold τ in (4.1). Ideal threshold should be equal to the best possible accuracy that can be guaranteed for reconstructing the value of a functional l_i at $x^\dagger$ from indirect noisy data y^δ under a fixed noise level δ. Coefficients below such a threshold cannot be distinguished from a noise any way.

The achievable accuracy for estimating a functional $\widehat{x}_i=\langle l_i,x^\dagger\rangle$ is essentially determined by the smoothness of the unknown solution $x^\dagger$, and the smoothness of the Ritz representer l_i. In rather general form a smoothness of $x^\dagger$ can be expressed as a source condition by

$$x^\dagger\in A_\varphi(R):=\{x\in X: x=\varphi(A^*A)u,\quad \|x\|_\varphi:=\|u\|_X\le R\}.$$

The variety of classes constructed in this way has been studied frequently [4, 17].

Note that the set $A_\varphi(R)$ is the ball of a radius R in a Hilbert space $A_\varphi = \{x : \|x\|_\varphi < \infty\}$, and

$$A_{\varphi_1} \hookrightarrow A_{\varphi_2} \text{ whenever } 0 < \varphi_1(t) \le \varphi_2(t), \quad t \in (0, \|A\|^2), \tag{4.7}$$

where $U \hookrightarrow V$ means that U is embedded in V.

Note also that the dual space of A_φ is given by $A_{1/\varphi}$. Therefore, one can always assume that the coefficient functional l_i obeys $l_i \in A_\lambda$ for some λ such that $0 < \lambda(t) \le 1/\varphi(t)$, $t \in (0, \|A\|^2)$, in order to ensure that $A_\lambda \hookrightarrow (A_\varphi)^* = A_{1/\varphi}$, and the functional $\langle l_i, x^\dagger\rangle$ is well-defined for $x^\dagger \in A_\varphi$.

Thus, given particular $x^\dagger$ and l_i, one can consider them as elements of some smoothness classes $A_\varphi(R)$ and $A_\lambda(R_1)$, $0 < \lambda \le 1/\varphi$. Then the best guaranteed accuracy of the estimation of $\widehat{x}_i = \langle l_i, x^\dagger\rangle$ from noisy data y^δ is defined to be the minimal uniform error over these classes,

$$e_\delta(A_\varphi(R), A_\lambda(R_1)) = \sup_{l\in A_\lambda(R_1)} \inf_{z} \sup_{x\in A_\varphi(R)} \sup_{y^\delta:\|Ax-y^\delta\|_Y\le\delta} \left|\langle l, x\rangle_X - \langle z, y^\delta\rangle_Y\right|.$$

The following result has been proven in [2] (see Corollary 3.1 there)

Theorem 4.2. *Assume that*

- **(a).** *there is a constant $\sigma > 0$ such that for the singular values $s_k(A)$ of A we have $s_{k+1}(A)/s_k(A) \ge \sigma$, $k = 1, 2, \ldots$;*
- **(b).** *the function $\lambda(t)/\sqrt{t}$ is non-increasing and the function $\sqrt{t}\lambda(t)$ is non-decreasing;*
- **(c).** *the functions φ and λ meet Δ_2 condition;*
- **(d).** *the function $\varphi^2\left((\theta_\varphi^2)^{-1}(t)\right)\lambda^2\left((\theta_\varphi^2)^{-1}(t)\right)$ is concave, where $\theta_\varphi(t) = \sqrt{t}\varphi(t)$.*

Then

$$e_\delta(A_\varphi(R), A_\lambda(R_1)) \asymp \varphi(\theta_\varphi^{-1}(\delta))\lambda(\theta_\varphi^{-1}(\delta)).$$

Moreover, if $x^\dagger \in A_\varphi(R)$, $l_i \in A_\lambda(R_1)$ then

$$\left|\langle l_i, x^\dagger\rangle_X - \langle z^i_{\beta_+}, y^\delta\rangle\right| \le c\varphi(\theta_\varphi^{-1}(\delta))\lambda(\theta_\varphi^{-1}(\delta)),$$

and it means that, up to a constant factor, the best guaranteed accuracy is realized by the data-functional strategy z^i_β with the regularization parameter chosen according to (4.6).

(Recall that f meets Δ_2-condition whenever $f(t) \asymp f(2t)$, and $a(u) \asymp b(u)$ means that $c_1 a(u) \le b(u) \le c_2 a(u)$, where c_1, c_2 do not depend on u).

From the Theorem 4.2 and our discussion above it follows that an order-optimal choice of the threshold level in (4.1) is

$$\tau \asymp \varphi(\theta_\varphi^{-1}(\delta))\lambda(\theta_\varphi^{-1}(\delta)) \tag{4.8}$$

whenever it is a priori known that $x^\dagger \in A_\varphi$ and $l_i \in A_\lambda$.

We present now the results of numerical experiments supporting the choice (4.8).

At first we revisit the example, where the system $\phi_i = \phi_i^M$, $i = 1, 2, \ldots, M$, consists of L_2-orthonormalized piece-wise constant functions, and A is given by (3.2).

It is well-known [19] that this operator acts along the Hilbert scale of Sobolev spaces of 1-periodic functions $\{W_2^\mu\}$ as isomorphism between pairs $W_2^{\mu-2}$ and W_2^μ, $\mu \in \mathbb{R}$. Moreover, from its singular value decomposition (3.11) it follows that $A_{t^\mu} = W_2^{4\mu}$. On the other hand, if $x^\dagger$ has a sparse expansion on the system $\{\phi_i^M\}$ then it is a discontinuous function, and as such it belongs to $W_2^{1/2}$ at most. The same is true for $l_i = \phi_i^M$.

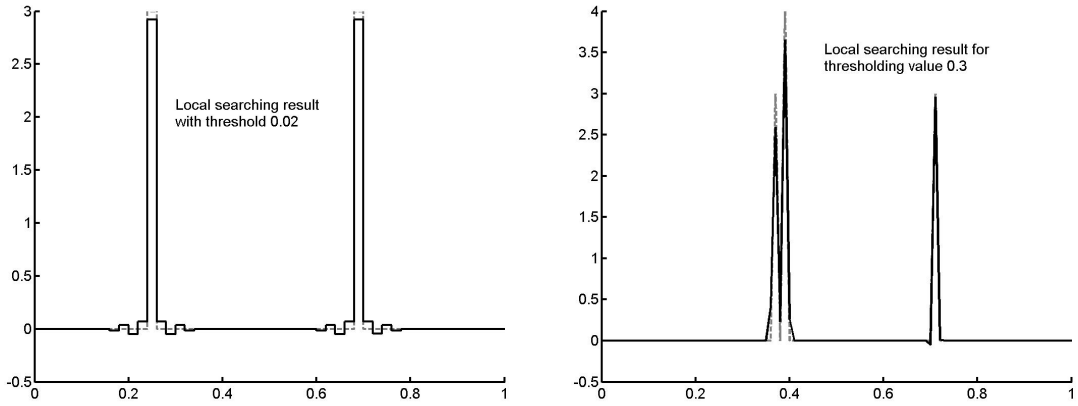

FIGURE 10. Order optimal reconstruction of the sparse structure given by the standard Tikhonov regularization used within the framework of data-functional strategy. Regularization parameters $\beta = \beta_+$ are chosen in accordance with the balancing principle (4.6).

Thus, in considered case $x^\dagger$ and l_i are the elements of $A_{t^{1/8}} = W_2^{1/2}$ at most. Then $\varphi(t) = \lambda(t) = t^{1/8}$, $\theta(t) = t^{5/8}$, and in accordance with (4.8) we should take a threshold level $\tau \asymp \delta^{2/5}$ to be sure that we will not lose the coefficients which can be in principle distinguished from the noise.

Recall that in our experiments with the operator (3.2) a noise simulation has been done for $\delta = 10^{-4}$, and corresponding Tikhonov approximation $x^\delta_{\alpha_+,50}$ has been displayed in Figure 2 (left). In accordance with (4.1), (4.8) we take into account only its coefficients above the threshold $\tau = 0.02 \approx 10^{-8/5}$, and estimate each such coefficient $\widehat{x}_i = \langle \phi_i^M, x^\dagger \rangle$ by $\langle z^i_{\beta_+}, y^\delta \rangle$, where $z^i_{\beta_+} = (\beta_+ I + AA^*)^{-1} A\phi_i^M$, and $\beta_+ = \beta^i_+$ is chosen in accordance with (4.6).

Finally we construct a sparse approximation for $x^\dagger$ as follows

$$x^\delta_{sparse} = \sum_{i:\langle l_i, x^\delta_{\alpha_+,M}\rangle \geq \tau} \langle z^i_{\beta_+}, y^\delta \rangle \phi_i^M. \tag{4.9}$$

In the Figure 10 (left) one can see the graph of x^δ_{sparse} corresponding to the values of parameters indicated above.

It is interesting to compare this graph with the Figure 2 (left), where $x^\delta_{\alpha_+,50}$ is shown. The latter has comparatively large coefficients near spurious modes ϕ_i^{50}, $i = 9, \ldots, 17, 31, \ldots, 39$, and they have gone over threshold τ. But in the final approximation x^δ_{sparse} the values of these coefficients have the order of the threshold (for example, the coefficient near ϕ_{12}^{50} is equal to 0.07). From this view point x^δ_{sparse} can be seen as an order-optimal reconstruction of the sparse structure, because all its coefficient estimate the real values of $\widehat{x}_i$ with the best guaranteed order of accuracy.

Similar analysis can be performed for our second example, where $\phi_i = \phi_i^{100}$, $i = 1, 2, \ldots, 100$, are the piece-wise linear B-splines, and $A : L_2(0,1) \to L_2(0,1)$ is the Abel integral operator.

This operator and the adjoint of it act continuously from $L_2(0,1)$ into $H_2^{1/2}$, where H_2^μ, $0 < \mu \leq 1$, is the space of functions $f \in L_2(0,1)$ with L_2-modulus of continuity

$\omega_2(f,h) = O(h^\mu)$. In terms of spaces A_φ it can be expressed as $A_{t^{1/2}} \hookrightarrow H_2^{1/2}$. If $x^\dagger$ has a sparse expansion on the system $\{\phi_i^M\}$ of piece-wise linear functions then $x^\dagger \in H_2^1$, and from [15], Section 7, it follows that such $x^\dagger$ meets a source condition $x^\dagger \in A_\varphi$ with $\varphi(t) = t$.

At the same time, to ensure that a coefficient-functional $\widehat{x}_i = \langle l_i, x^\dagger\rangle = x^\dagger(\frac{i}{M})$ is well-defined on some A_ψ one should assume $\psi(t) \le t^{1/2}$, because for $\psi(t) > t^{1/2}$ even an inclusion $A_\psi \subset H_2^{1/2}$ cannot be guaranteed, although $H_2^{1/2}$ contains discontinuous functions, and so it is too wide to be a domain for $l_i = \delta_{i/M}$. Therefore, $l_i \in A_\lambda = (A_\psi)^* = A_{1/\psi} \Rightarrow \lambda(t) = 1/\psi(t) \ge t^{-1/2}$.

If $\varphi(t) = t$ and $\lambda(t) \ge t^{-1/2}$ then in accordance with (4.8) the threshold level should be at least $\delta^{1/3}$. For a noise level $\delta = 0.02$ used in our simulations it gives us $\tau = 0.3$.

And again, the data-functional strategy based on the standard Tikhonov method and equipped with the parameter choice rule $\beta = \beta_+ = \beta_+^i$, $i : x^\delta_{\alpha_+,100}(\frac{i}{100}) > 0.3$, produces an order optimal sparse reconstruction x^δ_{sparse} and automatically reduces the coefficients near spurious modes to the level of the threshold order. It can be seen from Figure 10 (right) compared to Fig. 4 (in Fig. 10 (right) the mode ϕ_{37}^{100} has the coefficient $\hat{x}_{37} = 0.4$, for example).

Thus, numerical experiments presented above support our claim that the standard Tikhonov regularization combined with an appropriate parameter choice can be effectively used for reconstruction of the sparse structure.

Remark 4.1. Calculating threshold levels for our experiments we have transformed the assumption about the sparse structure of the solution into a priori assumption about solution smoothness given in terms of source conditions. It is well-known that such smoothness assumptions allow a priori choice of the regularization parameter for Tikhonov method which is order-optimal in the sense of accuracy measured in L_2-norm.

But in the context of sparsity reconstruction such a priori chosen parameter is not of interest since L_2-space does not promote a sparsity, as it can be seen from Fig. 5 (left), for example. Therefore, one is in need of a posteriori parameter choice rule to orient the standard Tikhonov method towards a regularization in an appropriate sparsity promoting space.

At the same time, using Theorem 4.2 one can use above mentioned a priori information about solution smoothness for choosing a threshold τ, and as we have shown, it essentially improves the quality of sparsity reconstruction

Acknowledgement

This research is supported by the Austrian Fonds Zur Förderung der Wissenschaftlichen Forschung (FWF), Grant P20235-N18.

References

[1] R. Anderssen (1986). The linear functional strategy for improperly posed problems. Inverse problems (Oberwolfach, 1986), 11–30, Internat. Schriftenreihe Numer. Math., 77.

[2] F. Bauer, P. Mathé and S.V. Pereverzev (2007) Local solutions to inverse problems in geodesy: The impact of the noise covariance structure upon the accuracy of estimation, *J. Geodesy*, **81**, 39–51

[3] T. Bonesky, K. Bredies, D.A. Lorenz and P. Maass (2007) A generalized conditional gradient method for nonlinear operator equation with sparsity constraints, *Inverse Problems*, **23**, 2041–2058.

[4] A. Böttcher, B. Hofmann, U. Tautenhahn and M. Yamamoto (2006) Convergence rates for Tikhonov regularization from different kinds of smoothness conditions, *Appl. Anal.*, **85**, 555–578.
[5] M. Burger and S. Osher (2004) Convergence rates of convex variational regularization, *Inverse Problems*, **20**, 1411–1421.
[6] I. Daubechies, M. Defrise and C. De Mol (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.* 57, 1413–1457.
[7] D. Donoho (1994) Statistical estimation and optimal recovery, *Ann. Statist.*, **22**, 238–270.
[8] D. Donoho (2006) Compressed sensing, *Information Theory, IEEE Transactions on*, **52**, 1289–1306.
[9] H. W. Engl, M. Hanke and A. Neubauer (1996) *Regularization of Inverse Problems*, Kluwer Academic Publishers.
[10] A. Goldenshluger and S. V. Pereverzev (2000) Adaptive estimation of linear functions in Hilbert scales from indirect white noise observation, *Probab. Theory Relat. Fields*, **118**, 169–186.
[11] P. C. Hansen (1994) Regularization tools, a MATLAB package for analysis of discrete regularization problems, *Numer. Algorithms*, **6**, 1–35.
[12] B. Hofmann, B. Kaltenbacher, C. Pöschl C and O. Scherzer (2007) A convergence rates results in Banach spaces with non-smooth operators, *Inverse Problems*, **23**, 987–1010.
[13] J. Lu, H. W. Engl, R. Machne and P. Schuster (2007) Inverse bifurcation analysis of a model for mammalian G_1/S regulatory module, *Lecture Notes in Bioinformatics*, 4414, 168–184.
[14] D. M. Malioutov (2003) A sparse signal reconstruction perspective for source localization with sensor arrays, *Massachusetts Institute of Technology*, Dissertation.
[15] P. Mathé and S. V. Pereverzev (2001) Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods, *SIAM J. Numer. Anal.* 38, 1999–2021.
[16] P. Mathé and S. V. Pereverzev (2002) Direct estimation of linear functionals from indirect noisy observations, *Journal of Complexity*, **18**, 500–516.
[17] P. Mathé and S. V. Pereverzev (2003) Geometry of linear ill-posed problems in variable Hilbert scales, *Inverse Problems* 19, 789–803.
[18] P. Mathé and S. V. Pereverzev (2006) Regularization of some linear ill-posed problems with discretized random noisy data, *Math. Comp.* 75, 1913–1929 .
[19] A. Neubauer (1988) An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates, *SIAM J. Numer. Anal.*, **25** 1313–1326.
[20] A. Neubauer (2006) Estimation of discontinuous solutions of ill-posed problems via adaptive grid regularization, *J. Inv. Ill-posed problems*, 14, 705–716.
[21] R. Ramlau and G. Teschke (2006) A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints, *Numer. Math.*, **104**, 177–203 (2006).

Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Science, Altenbergerstrasse 69, A-4040 Linz, Austria
E-mail address: shuai.lu@oeaw.ac.at, sergei.pereverzyev@oeaw.ac.at