PAPER • **OPEN ACCESS**

The quasi-optimality criterion in the linear functional strategy

To cite this article: Stefan Kindermann *et al* 2018 *Inverse Problems* **34** 075001

OPEN ACCESS
IOP Publishing

Inverse Problems

Inverse Problems **34** (2018) 075001 (24pp)

https://doi.org/10.1088/1361-6420/aabe4f

The quasi-optimality criterion in the linear functional strategy

Stefan Kindermann[1], **Sergiy Pereverzyev Jr**[2] **and Andrey Pilipenko**[3,4]

[1] Industrial Mathematics Institute, Johannes Kepler University of Linz, Altenbergerstrasse 69, A-4040 Linz, Austria
[2] Department of Neuroradiology, Medical University of Innsbruck, Anichstrasse 35, A-6020 Innsbruck, Austria
[3] Institute of Mathematics, National Academy of Sciences of Ukraine, Tereshchenkivska str. 3, 01601, Kyiv, Ukraine
[4] Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

E-mail: kindermann@indmath.uni-linz.ac.at, sergiy.pereverzyev@i-med.ac.at and pilipenko.ay@gmail.com

Received 28 September 2017, revised 11 December 2017
Accepted for publication 16 April 2018
Published 8 May 2018

Abstract

The linear functional strategy for the regularization of inverse problems is considered. For selecting the regularization parameter therein, we propose the heuristic quasi-optimality principle and some modifications including the smoothness of the linear functionals. We prove convergence rates for the linear functional strategy with these heuristic rules taking into account the smoothness of the solution and the functionals and imposing a structural condition on the noise. Furthermore, we study these noise conditions in both a deterministic and stochastic setup and verify that for mildly-ill-posed problems and Gaussian noise, these conditions are satisfied almost surely, where on the contrary, in the severely-ill-posed case and in a similar setup, the corresponding noise condition fails to hold. Moreover, we propose an aggregation method for adaptively optimizing the parameter choice rule by making use of improved rates for linear functionals. Numerical results indicate that this method yields better results than the standard heuristic rule.

Keywords: regularization, linear functional strategy, heuristic parameter choice rules, quasi-optimality rule, aggregation

(Some figures may appear in colour only in the online journal)

1. Introduction

The estimation of linear bounded functionals of an unknown element x from an indirect noisy observation y^δ given as

$$y^\delta = Tx + \delta\xi \tag{1.1}$$

is one of the classical problems in regularization theory [2]. Here, we assume that T is a linear, injective, not necessarily boundedly invertible operator from a solution Hilbert space X into an observation Hilbert space Y, ξ is an additive noise process, and δ is its intensity, or noise level, such that for $y = Tx$, it holds $\|y - y^\delta\| \leqslant \delta$, $\delta \in (0, 1)$. We use the same symbols $\langle\cdot,\cdot\rangle$, $\|\cdot\|$ for the inner products and the corresponding norms in both X and Y.

It is known that the problem of estimating the value $f(x) = \langle f, x\rangle$ of a linear bounded functional $f \in X$ from (1.1) is less ill-posed than the problem of estimating x, in the sense that the value $f(x)$ allows for a more accurate reconstruction than the element x in the X-norm [3, 10, 19]. A regularization of the first-named problem is usually performed by the so-called linear functional strategy [1] that is also closely related to the mollifier methods [18]. In case of a known noise intensity δ, the choice of the regularization parameters in the linear functional strategy has been extensively studied (see, e.g. [11, 19, 20] and references therein).

At the same time, in some applications, such as satellite gravity gradiometry, one cannot expect to have good knowledge of the noise model in general and of the noise intensity δ in particular (see, e.g. discussions in [5, 16]). As a remedy for this, regularization theory has an arsenal of so-called heuristic parameter choice strategies that do not require knowledge of the noise intensity and therefore can be used in the above mentioned applications. The quasi-optimality criterion [23] is one of the simplest and the oldest but still quite efficient instance among such strategies.

Of course, in the *worst case* scenario, where the noise ξ in (1.1) is assumed to be chosen by some antagonistic opponent only subject to the constraint $\|\xi\| \leqslant 1$, the quasi-optimality criterion, as well as any other heuristic parameter choice strategy, cannot guarantee convergence of the corresponding regularized approximants because of the so-called Bakushinskii veto [4]. On the other hand, it has been shown [6, 7] that for the quasi-optimality criterion, the Bakushinskii veto can be avoided if the regularization performance is measured *on average* over realizations of ξ.

At the same time, another way to overcome the Bakushinskii veto has been proposed in [15, 21], where convergence of the regularized approximants to x in the solution space norm and its rates have been established under a qualitative restriction on the noise ξ (a noise condition of Muckenhoupt type). Our intention in this paper is to extend this *restricted noise* approach in [15, 21] to the context of the linear functional strategy. We also show that for a wide class of moderately ill-posed problems (1.1) and for random noise ξ with bounded moments, the above mentioned Muckenhoupt-type condition is satisfied almost surely.

The case of severely ill-posed problems is considered as well. Note that in this case, the theoretical bounds on the convergence rates of the regularized approximants selected by the quasi-optimality criterion in the solution space norm are worse than those for the noise level-dependent parameter choice strategies. At the same time, as follows from our results, in the linear functional strategy, the above-mentioned convergence rate gap can be essentially reduced. This hints at an opportunity to use the linear functional strategy equipped with the quasi-optimality criterion for aggregating the constructed regularized approximants in a way described in [9]. Then from [9], it follows that such aggregation by the linear functional strategy can improve the accuracy compared to the aggregated regularized approximations, and this can be seen as a way to use the quasi-optimality criterion for mildly and severely ill-posed problems.

Note that a practical implementation of the quasi-optimality criterion depends on the so-called differential quadrature [8]

$$\left.\frac{\partial x_\alpha^\delta}{\partial \alpha}\right|_{\alpha=\alpha_i} \approx \sum_j a_{ij} x_{\alpha_j}^\delta \tag{1.2}$$

that is used to approximate the partial derivative $\frac{\partial x_\alpha^\delta}{\partial \alpha}$ of the regularized solution x_α^δ of (1.1), which is based on a current value of the regularization parameter $\alpha = \alpha_i$. Starting from the original paper [23], one usually uses a simple backward difference formula, where $a_{ij} = 0$ for $j \neq i, i-1$, and $a_{i,i} = -a_{i,i-1} = (\alpha_i - \alpha_{i-1})^{-1}$.

On the other hand, as it is mentioned in [8], there are many ways of determining the coefficients a_{ij} in (1.2). For example, in the backward difference formula, one can introduce correction factors such that

$$a_{i,i} = c_i \left(\alpha_i - \alpha_{i-1}\right)^{-1}, \quad a_{i,i-1} = -c_{i-1} \left(\alpha_i - \alpha_{i-1}\right)^{-1},$$

where c_ℓ, $\ell = i, i-1$, approximates the values c_ℓ^* minimizing the error $\left\|x - c_\ell^* x_{\alpha_\ell}^\delta\right\| = \min_c \left\|x - c x_{\alpha_\ell}^\delta\right\|$. It is clear that $c_\ell^* = \left\langle x, x_{\alpha_\ell}^\delta \right\rangle / \left\|x_{\alpha_\ell}^\delta\right\|^2$, and $\left\langle x, x_{\alpha_\ell}^\delta \right\rangle$ is the value of the linear bounded functional $x_{\alpha_\ell}^\delta \in X$ at the unknown solution that can be approximated by $\left\langle x_{\alpha_j}^\delta, x_{\alpha_\ell}^\delta \right\rangle$, where α_j is chosen by the quasi-optimality criterion.

The use of the backward difference formula corrected as above can be seen as an iterated quasi-optimality rule. We will demonstrate in section 5 that such a combination of the linear functional strategy—by an aggregation approach—and the quasi-optimality criterion can also improve the regularization performance as compared to the standard quasi-optimality.

The paper is organized as follows. In the next section, we present the problem setup and formulate main results. The proofs are given in section 3. In section 4, we describe random processes and investigate whether they almost surely meet the Muckenhoupt-type conditions. In section 5, we discuss a combination of the aggregation by means of the linear functional strategy with the quasi-optimality criterion and present numerical experiments.

2. The main convergence rates results

In this section, we formulate the main results. Let us introduce some standard notation. Let X, Y be Hilbert spaces, $T : X \to Y$ be a continuous linear operator such that $\mathrm{Ker}(T) = \{0\}$, $\mathrm{Ker}(T^*) = \{0\}$. Here, the assumptions of injectivity of T and T^* are only imposed for simplicity; the main results hold with modifications in the general case as well. We denote by E_λ and F_λ the spectral families for the operators T^*T and TT^*, respectively. The notion $\mathrm{R}(T)$ stands for the range and $\mathrm{Ker}(T)$ for the nullspace of the operator T. For f, g being functions or sequences, the notation $f \asymp g$ indicates that some constants c_1, c_2 exist such that $c_1 f \leqslant g \leqslant c_2 f$ for all arguments or sequence indices, where the constants in particular do not depend on δ.

Consider an ill-posed problem in the form $Tx = y$. Suppose that we observe $y^\delta \in Y$ such that $\|y^\delta - y\| \leqslant \delta$. We introduce regularized solutions obtained by a general spectral filter function g_α:

$$x_\alpha = g_\alpha(T^*T)T^*y, \;\; x_\alpha^\delta = g_\alpha(T^*T)T^*y^\delta.$$

Moreover, let $f \in X^* = X$ be a linear functional.

One aim of this paper is to obtain upper bounds for the error of linear functionals of the solutions, i.e. for the quantity $\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle$, where a parameter $\alpha(y^\delta)$ is selected in a special way and depends only on the observation y^δ. To state a smoothness/source condition for x and/or f, we use φ and $\varkappa$, which are continuous, non-negative, increasing real functions defined for positive real values (so-called index functions). Below we impose some standard assumptions on $\varphi, \varkappa, g_\alpha$.

Convergence rates estimates for the error $x_\alpha - x$ using some smoothness conditions on x are nowadays a classical topic. For instance, if δ is known, see, for example, [19], then under some natural conditions the best accuracy that can be guaranteed under the smoothness condition $x \in \mathrm{R}(\varphi(T^*T))$ is of the order $\varphi(\theta^{-1}(\delta))$, where $\theta(t) = \varphi(t)\sqrt{t}$ and θ^{-1} is its inverse function. For linear functionals, the situation can be improved: Assume that $x \in \mathrm{R}(\varphi(T^*T))$, and $f \in \mathrm{R}(\varkappa(T^*T))$, where $\varphi, \varkappa$ are index functions, then the best accuracy for the linear functionals $\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle$ is of the order $(\varkappa\varphi)(\theta^{-1}(\delta))$.

If the noise intensity is known, then the best order in accuracy can usually be achieved by standard means of selecting α. However, if δ is not known, the choice of the optimal α is a serious problem. For $\alpha(y^\delta)$ selected according to the quasi-optimality principle, some upper bounds for $\|x^\delta_{\alpha(y^\delta)} - x\|$ were obtained in [15, 21]. There it is proved that if $\varphi(t) = t^\mu$ and if the qualification μ_0 of the regularization g_α is such that $\mu_0 \geqslant \mu$, then

$$\|x^\delta_{\alpha(y^\delta)} - x\| = O(\delta^{\frac{2\mu}{2\mu+1}\frac{\mu}{\mu_0}}),\ \delta \to 0.$$

The main assumption on the noise was the following condition of Muckenhoupt type (noise condition):

$$\exists C > 0\ \forall \delta > 0\ \forall \alpha > 0 \qquad \alpha^2 \int_\alpha^\infty \lambda^{-1} d\|F_\lambda(y^\delta - y)\|^2 \leqslant C \int_0^\alpha \lambda d\|F_\lambda(y^\delta - y)\|^2. \tag{2.1}$$

We give some sufficient conditions that ensure (2.1) in section 4. In this paper we consider (2.1) and its generalization for the linear functional strategy. We discuss these conditions in the deterministic and random case; in particular we verify that for mildly ill-posed problems and Gaussian noise, it is satisfied almost surely. Moreover, we provide upper bounds for $\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle$, where $\alpha(y^\delta)$ is selected by the quasi-optimality principle as in [15, 21], and we also obtain some generalization of the upper bounds there. Furthermore, we prove improved bounds $\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle$, when $\alpha_\varkappa(y^\delta)$ is selected heuristically but using information about y^δ and also $\varkappa$.

For later use we introduce the quasi-optimality functional and a variant suited for functionals:

$$\psi^2(\alpha, y^\delta) = \int_0^\infty (1 - \lambda g_\alpha(\lambda))^2 \lambda g_\alpha^2(\lambda) d\|F_\lambda y^\delta\|^2 = \|(I - T^*Tg_\alpha(T^*T))x^\delta_\alpha\|^2,$$

$$\begin{aligned}\psi_\varkappa^2(\alpha, y^\delta) &= \int_0^\infty \varkappa^2(\lambda)(1 - \lambda g_\alpha(\lambda))^2 \lambda g_\alpha^2(\lambda) d\|F_\lambda y^\delta\|^2 \\ &= \|\varkappa(T^*T)(I - T^*Tg_\alpha(T^*T))x^\delta_\alpha\|^2.\end{aligned}$$

Note that the both above quantities can be seen as squares of some norms of the same element. The second one is a weighted version of the first one. Under our assumption on $\varkappa$, the second norm is weaker than the first, such that with respect to this norm a better convergence rate can be expected. This is an explanation why the value of a linear bounded functional $f(x)$,

$f \in \mathrm{R}(\varkappa(T^*T))$, allows for a more accurate reconstruction than the element x itself, and we benefit from this when aggregating several regularized solutions.

We introduce the following minimization-based heuristic parameter choice rules; the first one is the classical quasi-optimality rule as in [15, 21], while the second one is our modification:

$$\alpha(y^\delta) = \operatorname{argmin}_\alpha \psi(\alpha, y^\delta), \qquad \alpha_\varkappa(y^\delta) = \operatorname{argmin}_\alpha \psi_\varkappa(\alpha, y^\delta).$$

It is clear that $\alpha(y^\delta)$ can be computed without knowledge of δ, which is the defining feature of heuristic parameter choice rules. The novel modified rule $\alpha_\varkappa(y^\delta)$ additionally needs knowledge of the functional smoothness (via $\varkappa$). It will be shown that this additional information leads to improvements in the error bounds.

At first, we state some standard assumptions:

Assumption 1.

1. *For all* $\alpha > 0$ *we have*

$$0 \leqslant \lambda g_\alpha(\lambda) \leqslant 1, \lambda > 0, \qquad \sup_{\lambda>0} \sqrt{\lambda} g_\alpha(\lambda) \leqslant \frac{c_1}{\sqrt{\alpha}}. \tag{2.2}$$

2. *For all* $\alpha > 0$ *and* $\lambda \in (0, \alpha)$

$$(1 - \lambda g_\alpha(\lambda)) \geqslant c_2, \qquad \frac{c_3}{\alpha} \leqslant g_\alpha(\lambda) \leqslant \frac{c_4}{\alpha}. \tag{2.3}$$

3. *For any* $\lambda > 0$,

$$k(\lambda) := \inf_{\alpha \in (0, \|T\|]} \frac{(1 - \lambda g_\alpha(\lambda)) g_\alpha(\lambda)}{\alpha} > 0. \tag{2.4}$$

4. *The qualification of* g_α *covers* φ *and* $\varphi\varkappa$, *i.e. for all* $\alpha > 0$

$$\sup_{\lambda>0} |\varphi(\lambda)(1 - \lambda g_\alpha(\lambda))| \leqslant c_5 \varphi(\alpha), \tag{2.5}$$

$$\sup_{\lambda>0} |\varkappa(\lambda)\varphi(\lambda)(1 - \lambda g_\alpha(\lambda))| \leqslant c_6 \varkappa(\alpha)\varphi(\alpha). \tag{2.6}$$

5. *The function* $\varkappa$ *is covered by the qualification 1/2, i.e. for all* $\alpha > 0$

$$\sup_{\lambda>\alpha} \varkappa(\lambda)/\sqrt{\lambda} \leqslant c_7 \varkappa(\alpha)/\sqrt{\alpha}. \tag{2.7}$$

6. *The function* $\varkappa, \varphi$ *are regularly varying: For all* $c_8 > 0$ *there exists* $c_9 > 0$ *and* $\delta_0 > 0$ *such that*

$$\varphi(c_8\delta) \leqslant c_9\varphi(\delta) \quad \text{and} \quad \varkappa(c_8\delta) \leqslant c_9\varkappa(\delta) \qquad \forall \delta \in (0, \delta_0). \tag{2.8}$$

We note that in several places, condition (2.4) could be replaced by one with a more general qualification, i.e. that there exists $\mu_0 > 0$ such that for any $\lambda > 0$

$$k(\lambda) := \inf_{\alpha \in (0, \|T\|]} \frac{(1 - \lambda g_\alpha(\lambda)) g_\alpha(\lambda)}{\alpha^{\mu_0}} > 0. \tag{2.9}$$

Additionally to the structural conditions on the filter and index functions, we impose the following generalization of the noise condition (2.1):

$$\exists\delta_0 > 0 : \forall\delta \in (0,\delta_0)\forall\alpha > 0 :$$
$$\alpha^2 \int_\alpha^\infty \lambda^{-1}\varkappa^2(\lambda)d\|F_\lambda(y^\delta - y)\|^2 \leqslant c_{10}\int_0^\alpha \lambda\varkappa^2(\lambda)d\|F_\lambda(y^\delta - y)\|^2. \tag{2.10}$$

We state the main convergence result of the paper. In the sequel we denote by $\vee$ the maximum.

Theorem 1. *Suppose that $y \neq 0, x \in \mathrm{R}(\varphi(T^*T)), f \in \mathrm{R}(\varkappa(T^*T))$, where $\varphi, \varkappa$ are continuous, non-negative, increasing functions, the function $(0,\infty)^2 \ni (\lambda,\alpha) \to g_\alpha(\lambda)$ is continuous, and there are constants $c_1,\dots,c_9 > 0$ such that assumptions 1 hold. Moreover, let the noise condition (2.10) hold.*

Then, as $\delta \to 0$,

$$|x^\delta_{\alpha(y^\delta)} - x| = O\left(\varphi(\varphi(\theta^{-1}(\delta)) \vee \varphi(\theta^{-1}(\delta))\right), \tag{2.11}$$

$$\begin{aligned} |\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle| &= O\left(\varkappa(\varphi(\theta^{-1}(\delta)))\varphi(\varphi(\theta^{-1}(\delta)) \vee \varphi(\theta^{-1}(\delta))\right) \\ &= O\left((\varkappa\varphi)\circ(\varphi(\theta^{-1}(\delta))) \vee \varphi(\theta^{-1}(\delta))\right); \end{aligned} \tag{2.12}$$

$$|\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle| = O\left((\varkappa\varphi)\circ(\varkappa\varphi)(\theta^{-1}(\delta)) \vee (\varkappa\varphi)(\theta^{-1}(\delta))\right). \tag{2.13}$$

Observe, that the bound (2.13) for the modified rule $\alpha_\varkappa(y^\delta)$ is improved compared to (2.12).

Remark 1. If we replace (2.4) by the more general one, (2.9), then the convergence rates in this theorem read as

$$|x^\delta_{\alpha(y^\delta)} - x| = O\left(\varphi(\varphi^{1/\mu_0}(\theta^{-1}(\delta)) \vee \varphi(\theta^{-1}(\delta))\right), \tag{2.14}$$

$$\begin{aligned} |\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle| &= O\left(\varkappa(\varphi^{1/\mu_0}(\theta^{-1}(\delta)))\varphi(\varphi(\theta^{-1}(\delta)) \vee \varphi(\theta^{-1}(\delta))\right) \\ &= O\left((\varkappa\varphi)\circ(\varphi^{1/\mu_0}(\theta^{-1}(\delta))) \vee \varphi(\theta^{-1}(\delta))\right), \end{aligned} \tag{2.15}$$

$$|\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle| = O\left((\varkappa\varphi)\circ(\varkappa\varphi)^{1/\mu_0}(\theta^{-1}(\delta)) \vee (\varkappa\varphi)(\theta^{-1}(\delta))\right). \tag{2.16}$$

Remark 2. Formula (2.11) can be deduced using the reasoning of [15, 21] (the authors used concrete power function in their estimates). It also can be seen from our proof for $\varkappa(\lambda) \equiv 1$. To verify (2.11), actually only (2.1) is required, which is implied by (2.10) as the following remark indicates.

Remark 3. The main assumption of the theorem is (2.10). It can be considered as an analogue of (2.1) from [15, 21] for the mollified noise $\varkappa(T^*T)(y^\delta - y)$. It should be noted, that (2.10) implies (2.1). Indeed, it follows from the monotonicity of $\varkappa$ that $\frac{\varkappa(\lambda)}{\varkappa(\alpha)} \geqslant 1$ for $\lambda \geqslant \alpha$. So,

$$\alpha^2 \int_\alpha^\infty \lambda^{-1} d\|F_\lambda(y^\delta - y)\|^2 \leqslant \alpha^2 \int_\alpha^\infty \lambda^{-1}\frac{\varkappa^2(\lambda)}{\varkappa^2(\alpha)} d\|F_\lambda(y^\delta - y)\|^2.$$

Due to (2.10) the right hand side of the last inequality is less than or equal to

$$c_{10}\int_0^\alpha \lambda\frac{\varkappa^2(\lambda)}{\varkappa^2(\alpha)}d\|F_\lambda(y^\delta-y)\|^2 \leqslant c_{10}\int_0^\alpha \lambda d\|F_\lambda(y^\delta-y)\|^2,$$

where we used that $\frac{\varkappa(\lambda)}{\varkappa(\alpha)} \leqslant 1$ for $\lambda \leqslant \alpha$.

For Tikhonov's regularization $g_\alpha(\lambda) = \frac{1}{\alpha+\lambda}$, assumptions (2.2)–(2.4) are obviously satisfied, assumptions (2.5)–(2.7) are valid for $\varphi(t) = t^\mu$, $\varkappa(t) = t^\gamma$ with $\mu > 0, \gamma \in [0, 1/2], \mu + \gamma \leqslant 1$.

For iterated Tikhonov's regularization $g_\alpha(\lambda) = \lambda^{-1}(1 - \frac{\alpha^n}{(\alpha+\lambda)^n})$, assumptions (2.2), (2.3) and (2.9) are obviously satisfied, where $\mu_0 = n$; assumptions (2.5)–(2.7) are valid for $\varphi(t) = t^\mu$, $\varkappa(t) = t^\gamma$ with $\mu > 0, \gamma \in [0, 1/2], \mu + \gamma \leqslant \mu_0$.

Specializing the previous theorem to Tikhonov regularization and Hölder-type index functions, we find the following corollary:

Corollary 1. *Let $g_\alpha(\lambda) = \frac{1}{\alpha+\lambda}$, $\varphi(t) = t^\mu$, $\varkappa(t) = t^\gamma$ with $\mu > 0, \gamma \in [0, 1/2], \mu + \gamma \leqslant 1$. Assume that (2.10) is satisfied. Then as $\delta \to 0$,*

$$\begin{aligned}
|x^\delta_{\alpha(y^\delta)} - x| &= O\left(\delta^{\frac{2\mu}{2\mu+1}\mu}\right),\\
|\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle| &= O\left(\delta^{\frac{2\mu}{2\mu+1}(\mu+\gamma)}\right),\\
|\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle| &= O\left(\delta^{\frac{2(\mu+\gamma)^2}{2\mu+1}}\right).
\end{aligned}$$

Remark 4. If we use the generalized qualification condition (2.9) and replace the condition $\mu + \gamma \leqslant 1$ by $\mu + \gamma \leqslant \mu_0$, then the rates in corollary 1 have to be replaced by

$$\begin{aligned}
|x^\delta_{\alpha(y^\delta)} - x| &= O\left(\delta^{\frac{2\mu}{2\mu+1}\frac{\mu}{\mu_0}}\right), \qquad |\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle| = O\left(\delta^{\frac{2\mu}{2\mu+1}\frac{\mu+\gamma}{\mu_0}}\right),\\
|\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle| &= O\left(\delta^{\frac{2(\mu+\gamma)}{2\mu+1}\frac{\mu+\gamma}{\mu_0}}\right).
\end{aligned}$$

Remark 5. Under the conditions of corollary 1 the bound for $\|x^\delta_{\alpha(y^\delta)} - x\|$ in [15, 21] is $O\left(\delta^{\frac{2\mu}{2\mu+1}\mu}\right)$ (respectively, $O\left(\delta^{\frac{2\mu}{2\mu+1}\frac{\mu}{\mu_0}}\right)$ for the case with μ_0) while the order-optimal bound is $O\left(\delta^{\frac{2\mu}{2\mu+1}}\right)$. For linear functionals as in the corollary, it is known that the optimal order is $|\langle f, x^\delta_\alpha - x\rangle| = O\left(\delta^{\frac{2(\mu+\gamma)}{2\mu+1}}\right)$, as $\delta \to 0$; see [19].

3. Proof of the main result

We need the following auxiliary results. Many of them are quite standard, we provide the proofs to make the exposition self-contained. At first we provide bounds for the approximation errors.

Lemma 1. *Under assumption 1, there is $c > 0$ such that for all $\alpha > 0$ we have*

$$\begin{aligned}\|x_\alpha - x\| &\leqslant c\varphi(\alpha);\\ |\langle f, x_\alpha - x\rangle| &\leqslant c\varkappa(\alpha)\varphi(\alpha);\\ \|\varkappa(T^*T)(x_\alpha - x)\| &\leqslant c\varkappa(\alpha)\varphi(\alpha).\end{aligned}$$

Proof. Let $x = \varphi(T^*T)v_x, f = \varkappa(T^*T)u_f$. Then

$$\begin{aligned}\|x_\alpha - x\|^2 = \int_0^\infty (1-\lambda g_\alpha(\lambda))^2 d\|E_\lambda x\|^2 &= \int_0^\infty \varphi^2(\lambda)(1-\lambda g_\alpha(\lambda))^2 d\|E_\lambda v_x\|^2\\ &\leqslant K_1 \sup_\lambda(\varphi(\lambda)(1-\lambda g_\alpha(\lambda)))^2 \leqslant K_2\varphi^2(\alpha),\end{aligned}$$

which proves the first inequality. For the remain ones, we estimate

$$\begin{aligned}\langle f, x_\alpha - x\rangle^2 &= \langle \varkappa(T^*T)u_f, x_\alpha - x\rangle^2 = \langle u_f, \varkappa(T^*T)(x_\alpha - x)\rangle^2\\ &\leqslant \|u_f\|^2\|\varkappa(T^*T)(x_\alpha - x)\|^2 = \|u_f\|^2\int_0^\infty \varkappa^2(\lambda)(1-\lambda g_\alpha(\lambda))^2 d\|E_\lambda x\|^2\\ &= \|u_f\|^2\int_0^\infty \varkappa^2(\lambda)\varphi^2(\lambda)(1-\lambda g_\alpha(\lambda))^2 d\|E_\lambda v_x\|^2\\ &\leqslant K_1\sup_\lambda(\varkappa(\lambda)\varphi(\lambda)(1-\lambda g_\alpha(\lambda)))^2 \leqslant K_2\varkappa^2(\alpha)\varphi^2(\alpha),\end{aligned}$$

where we used (2.6). □

Next we bound the parameter choice functionals.

Lemma 2. *Let assumption 1 hold. Then there exists a $c > 0$ such that for all $\alpha > 0$ and all $\delta > 0$ we have*

$$\begin{aligned}\psi(\alpha, y) &\leqslant \|x_\alpha - x\|, & \psi(\alpha, y^\delta - y) &\leqslant \|x_\alpha^\delta - x_\alpha\|,\\ \psi(\alpha, y^\delta) &\leqslant \|x_\alpha - x\| + \|x_\alpha^\delta - x_\alpha\|, & &\\ \psi_\varkappa(\alpha, y) &\leqslant \|\varkappa(T^*T)(x_\alpha - x)\|, & \psi_\varkappa(\alpha, y^\delta - y) &\leqslant \|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\|,\\ \psi_\varkappa(\alpha, y^\delta) &\leqslant \|\varkappa(T^*T)(x_\alpha - x)\| + \|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\|. & &\end{aligned}$$

Proof.

$$\begin{aligned}\psi^2(\alpha, y) &= \|(I - T^*Tg_\alpha(T^*T))x_\alpha\|^2 = \|(I - T^*Tg_\alpha(T^*T))T^*Tg_\alpha(T^*T)x\|^2\\ &= \int_0^\infty (1-\lambda g_\alpha^2(\lambda))^2(\lambda g_\alpha(\lambda))^2 d\|E_\lambda x\|^2 \leqslant \int_0^\infty (1-\lambda g_\alpha^2(\lambda))^2 d\|E_\lambda x\|^2 = \|x_\alpha - x\|^2.\\ \psi(\alpha, y^\delta - y) &= \|(I - T^*Tg_\alpha(T^*T))(x_\alpha^\delta - x_\alpha)\| \leqslant \|x_\alpha^\delta - x_\alpha\|.\end{aligned}$$

The inequalities for $\psi_\varkappa$ follow in an analogous way. □

The following result is a straightforward consequence of (2.2) and $\|y^\delta - y\| \leqslant \delta$.

Lemma 3. *Let assumption 1 hold. There exists $c > 0$ such that for all $\alpha > 0$ and all $\delta > 0$ we have*

$$\|x_\alpha^\delta - x_\alpha\|^2 = \int_0^\infty \lambda g_\alpha^2(\lambda) d\|F_\lambda(y^\delta - y)\|^2 \leqslant c\left(\frac{\delta}{\sqrt{\alpha}}\right)^2.$$

Lemma 4. *Let assumption 1 hold. We have for* $\delta > 0$,

$$\psi(\alpha(y^\delta), y^\delta) = \inf_\alpha \psi(\alpha, y^\delta) \leqslant \inf_\alpha \left(\|x_\alpha - x\| + \|x_\alpha^\delta - x_\alpha\|\right) \leqslant c_0\varphi(\theta^{-1}(\delta)), \tag{3.1}$$

$$\psi_\varkappa(\alpha(y^\delta), y^\delta) \leqslant c_1\psi(\alpha(y^\delta), y^\delta) \leqslant c_2\varphi(\theta^{-1}(\delta)), \tag{3.2}$$

where c_0, c_1, c_2 *are constants independent of* δ, $\theta(t) = \varphi(t)\sqrt{t}$, *and* θ^{-1} *is its inverse function.*

Proof. Let $\bar{\alpha}$ be such that $\varphi(\bar{\alpha}) = \frac{\delta}{\sqrt{\bar{\alpha}}}$, i.e. $\bar{\alpha} = \theta^{-1}(\delta)$. Then (3.1) follows from lemmas 1 and 3, and the following calculations

$$\begin{aligned}\inf_\alpha \left(\|x_\alpha - x\| + \|x_\alpha^\delta - x_\alpha\|\right) &\leqslant C \inf_\alpha \left(\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}\right) \leqslant C\left(\varphi(\bar{\alpha}) + \frac{\delta}{\sqrt{\bar{\alpha}}}\right) \\ &= 2C\varphi(\theta^{-1}(\delta)).\end{aligned}$$

Inequality (3.2) follows from (3.1) because $\varkappa$ is bounded on $[0, \|T\|]$. □

The next lemma gives a very important consequence of (2.10), which is crucial for our proofs. In the sequel, we use the symbols $K_1, K_2, \ldots$, and C for generic constants that may take different values in different formulas.

Lemma 5. *Let assumption 1 hold and assume the generalized noise condition (2.10). Then there exist constants* K_1, K_2, *and* $\delta_0 > 0$ *such that for all* $\delta \in (0, \delta_0)$, $\alpha > 0$:

$$|\langle f, (x_\alpha^\delta - x_\alpha)\rangle| \leqslant K_1\|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\| \leqslant K_2\psi_\varkappa(\alpha, y^\delta - y).$$

Proof. The first inequality is proved similarly to lemma 1. Let us verify the second inequality. By splitting the integral we obtain

$$\|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\|^2 = \int_0^\infty \varkappa^2(\lambda)\lambda g_\alpha^2(\lambda) d\|F_\lambda(y^\delta - y)\|^2 = \int_0^\alpha [\ldots] + \int_\alpha^\infty [\ldots].$$

It follows from (2.2) and (2.10) that

$$\begin{aligned}\int_\alpha^\infty \varkappa^2(\lambda)\lambda g_\alpha^2(\lambda) d\|F_\lambda(y^\delta - y)\|^2 &= \int_\alpha^\infty \varkappa^2(\lambda)\lambda^{-1}(\lambda g_\alpha(\lambda))^2 d\|F_\lambda(y^\delta - y)\|^2 \\ &\leqslant \int_\alpha^\infty \varkappa^2(\lambda)\lambda^{-1} d\|F_\lambda(y^\delta - y)\|^2 \leqslant K_1\alpha^{-2}\int_0^\alpha \lambda\varkappa^2(\lambda) d\|F_\lambda(y^\delta - y)\|^2.\end{aligned}$$

The second assumption in (2.3) yields that $g_\alpha^2(\lambda) \leqslant \frac{\text{const}}{\alpha^2}$ for $\lambda \in (0, \alpha)$. Thus,

$$\int_0^\alpha \varkappa^2(\lambda)\lambda g_\alpha^2(\lambda) d\|F_\lambda(y^\delta - y)\|^2 \leqslant K_2\alpha^{-2}\int_0^\alpha \varkappa^2(\lambda)\lambda d\|F_\lambda(y^\delta - y)\|^2,$$

and consequently

$$\|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\|^2 \leqslant K_3\alpha^{-2}\int_0^\alpha \varkappa^2(\lambda)\lambda d\|F_\lambda(y^\delta - y)\|^2.$$

Since $(1-\lambda g_\alpha(\lambda)) \geqslant \text{const} > 0$ and $g_\alpha^2(\lambda) \geqslant \frac{\text{const}}{\alpha^2} > 0$ for $\lambda \in (0,\alpha)$, see (2.3), we have

$$\alpha^{-2}\int_0^\alpha \varkappa^2(\lambda)\lambda d\|F_\lambda(y^\delta-y)\|^2 \leqslant K_4\int_0^\alpha \varkappa^2(\lambda)\lambda g_\alpha^2(\lambda)(1-\lambda g_\alpha(\lambda))^2 d\|F_\lambda(y^\delta-y)\|^2$$
$$\leqslant K_4\int_0^\infty \varkappa^2(\lambda)\lambda g_\alpha^2(\lambda)(1-\lambda g_\alpha(\lambda))^2 d\|F_\lambda(y^\delta-y)\|^2 = K_4\psi_\varkappa^2(\alpha,y^\delta-y).$$

□

Lemma 6. *Let $y\neq 0$. Then there exist $C>0$ and $\delta_0>0$ such that for all* $\delta\in(0,\delta_0)$ and $\alpha\in(0,1)$

$$\psi(\alpha,y^\delta)\geqslant C\alpha \qquad \text{and} \qquad \psi_\varkappa(\alpha,y^\delta)\geqslant C\alpha. \tag{3.3}$$

Proof. Let us only verify the second inequality. We follow the course of the proof from [15, 21]. Let $\bar\alpha$ be fixed. It follows from (2.3) that

$$\begin{aligned}\psi_\varkappa^2(\alpha,y^\delta) &= \int_0^\infty \varkappa^2(\lambda)(1-\lambda g_\alpha(\lambda))^2\lambda g_\alpha^2(\lambda)d\|F_\lambda y^\delta\|^2\\
&\geqslant \int_{\bar\alpha}^\infty \varkappa^2(\lambda)\alpha^2(\frac{(1-\lambda g_\alpha(\lambda))g_\alpha(\lambda)}{\alpha})^2\lambda d\|F_\lambda y^\delta\|^2\\
&\geqslant \int_{\bar\alpha}^\infty \varkappa^2(\lambda)\alpha^2 \inf_{a\in(0,\|T\|]}\left(\frac{(1-\lambda g_a(\lambda))g_a(\lambda)}{a}\right)^2\lambda d\|F_\lambda y^\delta\|^2\\
&= \alpha^2\int_{\bar\alpha}^\infty \varkappa^2(\lambda)k^2(\lambda)\lambda d\|F_\lambda y^\delta\|^2\\
&\geqslant \alpha^2\left(\int_{\bar\alpha}^\infty \varkappa^2(\lambda)k^2(\lambda)\lambda d\left[2^{-1}\|F_\lambda y\|^2-\|F_\lambda(y^\delta-y)\|^2\right]\right),\end{aligned}$$

where k is from (2.3). Set $h(\lambda)=\varkappa^2(\lambda)k^2(\lambda)\lambda, \lambda>0$; the function h is positive. It follows from the definition of k that $h(\lambda)\leqslant\varkappa^2(\lambda)(1-\lambda g_1^2(\lambda))^2\lambda g_1^2(\lambda)$. So, all considered integrals are finite. Select $\bar\alpha>0$ such that $\int_{\bar\alpha}^\infty d\|F_\lambda y\|^2>0$. Then $\int_{\bar\alpha}^\infty h(\lambda)d2^{-1}\|F_\lambda y\|^2>0$. Since

$$\lim_{\delta\to0}\int_{\bar\alpha}^\infty h(\lambda)d\|F_\lambda(y^\delta-y)\|^2=0,$$

there is $\delta_0>0$ such that

$$\int_{\bar\alpha}^\infty h(\lambda)d\|F_\lambda(y^\delta-y)\|^2\leqslant 4^{-1}\int_{\bar\alpha}^\infty h(\lambda)d\|F_\lambda y\|^2.$$

Hence we get the second inequality in (3.3) with $C=4^{-1}\int_{\bar\alpha}^\infty h(\lambda)d\|F_\lambda y\|^2$. □

Bounds for $\psi_\varkappa(\alpha_\varkappa(y^\delta),y^\delta)$ are given in the following statement.

Lemma 7. *We have*

$$\psi_\varkappa(\alpha_\varkappa(y^\delta),y^\delta)\leqslant C\varkappa(\theta^{-1}(\delta))\varphi(\theta^{-1}(\delta)), \tag{3.4}$$

where C is a constant independent of δ.

Proof. Lemmas 1 and 2 yield that

$$\psi_\varkappa(\alpha_\varkappa(y^\delta), y^\delta) = \inf_\alpha \psi_\varkappa(\alpha, y^\delta) \leqslant \inf_\alpha \left(\|\varkappa(T^*T)(x_\alpha - x)\| + \|\varkappa(T^*T)(x_\alpha^\delta - x_\alpha)\|\right).$$

It follows from [19, p 107] that the right hand side is less than or equal to $C\varkappa(\theta^{-1}(\delta))\varphi(\theta^{-1}(\delta))$. □

Proof of theorem 1. By lemmas 1, 5 and 2,

$$\begin{aligned}
|\langle f, x^\delta_{\alpha(y^\delta)} - x\rangle| &\leqslant |\langle f, x^\delta_{\alpha(y^\delta)} - x_{\alpha(y^\delta)}\rangle| + |\langle f, x_{\alpha(y^\delta)} - x\rangle| \\
&\overset{\text{Lemma 1}}{\leqslant} |\langle f, x^\delta_{\alpha(y^\delta)} - x_{\alpha(y^\delta)}\rangle| + K_1\varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta)) \\
&\overset{\text{Lemma 5}}{\leqslant} K_2\left(\psi_\varkappa(\alpha(y^\delta), y^\delta - y) + \varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta))\right) \\
&\leqslant K_2\left(\psi_\varkappa(\alpha(y^\delta), y^\delta) + \psi_\varkappa(\alpha(y^\delta), y) + \varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta))\right) \\
&\overset{\text{Lemmas 1, 2}}{\leqslant} K_2\left(\psi_\varkappa(\alpha(y^\delta), y^\delta) + K_3\varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta)) + \varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta))\right) \\
&\leqslant K_4\left(\psi_\varkappa(\alpha(y^\delta), y^\delta) + \varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta))\right) \qquad (3.5)
\end{aligned}$$

$$\overset{\text{see (3.2)}}{\leqslant} K_5\left(\varphi(\theta^{-1}(\delta)) + \varkappa(\alpha(y^\delta))\varphi(\alpha(y^\delta))\right). \qquad (3.6)$$

It follows from lemma 6 and (3.1) that

$$\alpha(y^\delta) \leqslant K_8\psi(\alpha(y^\delta), y^\delta) \leqslant K_9\varphi(\theta^{-1}(\delta))$$

for sufficiently small $\delta > 0$. Thus, the monotonicity of $\varkappa$ and φ and (2.8) yields that the right hand side of (3.6) does not exceed

$$K_{10}\left(\varphi(\theta^{-1}(\delta)) + \varkappa(\varphi(\theta^{-1}(\delta)))\varphi(\varphi(\theta^{-1}(\delta)))\right).$$

This proves (2.12).

The proof of (2.13) is identical to that of (2.12). Similarly to (3.5) we get

$$|\langle f, x^\delta_{\alpha_\varkappa(y^\delta)} - x\rangle| \leqslant K_{11}\left(\psi_\varkappa(\alpha_\varkappa(y^\delta), y^\delta) + \varkappa(\alpha_\varkappa(y^\delta))\varphi(\alpha_\varkappa(y^\delta))\right). \qquad (3.7)$$

It follows from lemma 6 that $\alpha_\varkappa(y^\delta) \leqslant \psi_\varkappa(\alpha_\varkappa(y^\delta), y^\delta)$. The proof of the theorem 1 now follows from (3.7), (3.4). □

4. Case studies of noise conditions

In order to understand (2.1) and (2.10), we study situations, when these inequalities hold or fail; in particular for the case of random noise.

In this section, we specialize to the case when T is a compact operator, thus it allows for a singular system λ_k, v_k, u_k, i.e. $\lambda_k > 0$, $Tv_k = \lambda_k u_k$, $T^*u_k = \lambda_k v_k$. Then (2.1) and (2.10) can be equivalently rephrased as

$$\exists C : \forall n \geqslant 1 \quad \lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \langle y - y^\delta, u_k \rangle^2 \leqslant C \sum_{k=n+1}^{\infty} \lambda_k^2 \langle y - y^\delta, u_k \rangle^2 \tag{4.1}$$

and

$$\exists C : \forall n \geqslant 1 \quad \lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \varkappa^2(\lambda_k^2) \langle y - y^\delta, u_k \rangle^2 \leqslant C \sum_{k=n+1}^{\infty} \lambda_k^2 \varkappa^2(\lambda_k^2) \langle y - y^\delta, u_k \rangle^2, \tag{4.2}$$

respectively.

As an example, we now assume a polynomially decaying deterministic noise, i.e.

$$\langle y - y^\delta, u_k \rangle^2 \asymp \delta^2 k^{-\rho}, \qquad \rho > 0. \tag{4.3}$$

Then, the following tables exemplify some sufficient conditions for the noise condition (4.1) for different degrees of ill-posedness:

Ill-posedness		Noise	Sufficient condition for (4.1)
Mildly	$\lambda_k^2 \asymp k^{-\beta}$	(4.3)	$\rho > 1, \beta > \rho - 1,$
Severely $a \in (0,1)$	$\lambda_k^2 \asymp a^k,$	(4.3)	$\rho > 1.$

A similar results can be stated for the modified noise condition (4.2):

Ill-posedness		noise	$\varkappa$	Sufficient condition for (4.2)
Mildly	$\lambda_k^2 \asymp k^{-\beta}$	(4.3)	$\varkappa(t) \asymp t^\gamma$ (or $\varkappa^2(\lambda_k^2) \asymp k^{-2\gamma\beta}$)	$\rho > 1, \gamma > 0,$ $\beta > 2\gamma\beta + \rho - 1,$
Severely $a \in (0,1)$	$\lambda_k^2 \asymp a^k,$	(4.3)	$\varkappa(t) \asymp t^\gamma$	$\rho > 1, \gamma \in (0,1).$
Severely $a \in (0,1)$	$\lambda_k^2 \asymp a^k,$	(4.3)	$\varkappa(t) \asymp (\log t^{-1})^{-\gamma}$	$\rho > 1, \gamma > 0.$

In contrast to the deterministic case, we now investigate the case of random noise. We assume that the noise is random and of the form

$$y^\delta - y = \sum_{k=1}^{\infty} \sigma_k(\delta) \xi_k u_k, \tag{4.4}$$

where $\xi_k = \xi_k(\omega), \omega \in \Omega$ are independent random variables given on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, with

$$\mathrm{E}\xi_k = 0, \qquad \mathrm{Var}(\xi_k) = 1, \tag{4.5}$$

and analogous to (4.3), we assume that

$$\sigma_k^2(\delta) \asymp \delta^2 k^{-\rho}, \qquad \rho > 1. \tag{4.6}$$

Note that $\mathrm{E}(y^\delta - y) = 0$ and $\mathrm{Var}(y^\delta - y) \asymp \delta^2$.

The stochastic analogue of the inequality (4.1) is of the following form: for almost all ω there is a constant $C = C(\omega)$ such that

$$\forall n \geqslant 1 \qquad \lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \sigma_k^2(\delta) \xi_k^2 \leqslant C \sum_{k=n+1}^{\infty} \lambda_k^2 \sigma_k^2(\delta) \xi_k^2 \tag{4.7}$$

or

$$\sup_{n \geqslant 1} \frac{\lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \sigma_k^2(\delta) \xi_k^2}{\sum_{k=n+1}^{\infty} \lambda_k^2 \sigma_k^2(\delta) \xi_k^2} < \infty \text{ almost surely.}$$

The stochastic analogue of (4.2) can be considered similarly with the natural modifications.

Theorem 2. *Assume a mildly ill-posed case, i.e.* $\lambda_k^2 \asymp k^{-\beta}$, *with* $\beta > 0$. *Moreover, let the noise satisfy (4.4)–(4.6), and assume that the random variables* $\{\xi_k\}$ *have moments of all orders:*

$$\forall p \geqslant 1 \quad \sup_k \mathrm{E}|\xi_k|^p < \infty.$$

Then, if $\beta > \rho - 1$

$$\sup_{n \geqslant 1} \frac{\lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \sigma_k^2(\delta) \xi_k^2}{\sum_{k=n+1}^{\infty} \lambda_k^2 \sigma_k^2(\delta) \xi_k^2} < \infty \text{ almost surely.} \tag{4.8}$$

The proof of this theorem is given in the appendix. The assumptions on $\{\xi_k\}$ hold in particular for independent Gaussian $N(0, 1)$-random variables. Thus, for the mildly ill-posed operators, the stochastic case is completely similar to the deterministic one and the analogous convergence rates results hold true (almost surely).

This, however, is not true for the severely ill-posed case as the following theorem shows.

Theorem 3. *Assume a severely ill-posed case, i.e.* $\lambda_k^2 \asymp a^k$, *with* $a \in (0, 1)$ *and let (4.4) and (4.6) hold, where* $\{\xi_k\}$ *are independent Gaussian* $N(0, 1)$ *random variables. Then*

$$\mathrm{P}\left(\sup_{n \geqslant 1} \frac{\lambda_n^4 \sum_{k=1}^{n} \lambda_k^{-2} \sigma_k^2(\delta) \xi_k^2}{\sum_{k=n+1}^{\infty} \lambda_k^2 \sigma_k^2(\delta) \xi_k^2} = \infty \right) = 1. \tag{4.9}$$

In particular, in this situation, the noise condition (4.1) fails almost surely. This shows that the difference between stochastic and deterministic cases may be very essential.

Proof of theorem 3. Introduce the Markov moment $\tau_p := \inf\{n \geqslant 1 \ : \ \xi_n^2 > p\}$. Obviously, $\tau_p < \infty$ almost surely. Then

$$a^{2\tau_p} \sum_{k=1}^{\tau_p} a^{-k} k^{-\rho} \xi_k^2 \geqslant a^{2\tau_p} a^{-\tau_p} {\tau_p}^{-\rho} p = a^{\tau_p} {\tau_p}^{-\rho} p.$$

We have

$$\sum_{k=\tau_p+1}^{\infty} a^k k^{-\rho} \xi_k^2 = a^{\tau_p} \sum_{k=1}^{\infty} a^k (k + \tau_p)^{-\rho} \xi_{k+\tau_p}^2 \leqslant a^{\tau_p} {\tau_p}^{-\rho} \sum_{k=1}^{\infty} a^k \xi_{k+\tau_p}^2.$$

Since τ_p is a finite Markov moment,

$$\mathrm{E}\xi_{k+\tau_p}^2 = \sum_{n \geqslant 0} \mathrm{E}(\mathbb{I}_{\tau_p = n} \xi_{k+n}^2) = \sum_{n \geqslant 0} \mathrm{E}\mathbb{I}_{\tau_p = n} \mathrm{E}\xi_{k+n}^2 = \sum_{n \geqslant 0} \mathrm{E}\mathbb{I}_{\tau_p = n} = 1.$$

Hence $\mathrm{E} \sum_{k=1}^{\infty} a^k \xi_{k+\tau_p}^2 = \sum_{k=1}^{\infty} a^k = (1 - a)^{-1}$. By Chebyshev's inequality we have

$$\mathrm{P}(\sum_{k=1}^{\infty} a^k \xi_{k+\tau_p}^2 \geqslant \sqrt{p}) \leqslant ((1-a)\,p)^{-1/2}.$$

Therefore for any $p \geqslant 1$

$$\mathrm{P}\left(\exists n \geqslant 1 \quad \frac{a^{2n}\sum_{k=1}^{n} a^{-k}k^{-\rho}\xi_k^2}{\sum_{k=n+1}^{\infty} a^k k^{-\rho}\xi_k^2} \geqslant \sqrt{p}\right) \geqslant \mathrm{P}\left(\frac{a^{2\tau_p}\sum_{k=1}^{\tau_p} a^{-k}k^{-\rho}\xi_k^2}{\sum_{k=\tau_p+1}^{\infty} a^k k^{-\rho}\xi_k^2} \geqslant \sqrt{p}\right)$$
$$\geqslant \mathrm{P}\left(\frac{a^{\tau_p}\tau_p{}^{-\rho}p}{a^{\tau_p}\tau_p{}^{-\rho}\sum_{k=1}^{\infty} a^k\xi_{k+\tau_p}^2} \geqslant \sqrt{p}\right) = \mathrm{P}\left(\frac{p}{\sum_{k=1}^{\infty} a^k\xi_{k+\tau_p}^2} \geqslant \sqrt{p}\right)$$
$$= \mathrm{P}\left(\sum_{k=1}^{\infty} a^k\xi_{k+\tau_p}^2 \leqslant \sqrt{p}\right) \geqslant 1 - ((1-a)\,p)^{-1/2}.$$

This yields (4.9). □

In this proof we used the following properties of the sequence $\{\xi_k\}$:

(i) independence, (ii) $\sup_k \mathrm{E}\xi_k^2 < \infty$, (iii) $\limsup_{k\to\infty} |\xi_k| = +\infty$ a.s.

Remark 6. It may be conjectured that if $\{\xi_k\}$ are uniformly bounded random variables, for example, if $\{\xi_k\}$ have the uniform distribution on $[-1, 1]$, then (4.8) would hold. However, this conjecture is wrong. Problems may arise if $\{\xi_k\}$ are i.i.d. and 0 belongs to the support of the ξ_k's distribution, i.e. if $\mathrm{P}(|\xi_k| < \varepsilon) > 0$ for any $\varepsilon > 0$.

Indeed, let $m, p \geqslant 1$ be fixed. Select $c > 0$ such that $\mathrm{P}(|\xi_k| > c) > 0$. Set

$$\tau_{mp} := \inf\Big\{n \geqslant 1 \ : \ |\xi_{n-p}| > c, |\xi_{n-m+1}| < p^{-1}, |\xi_{n-m+2}| < p^{-1}, \dots$$
$$\dots, |\xi_{n-1}| < p^{-1}, |\xi_n| < p^{-1}\Big\}.$$

Since $\mathrm{P}(|\xi_k| > c) > 0$ and $\mathrm{P}(|\xi_k| < p^{-1}) > 0$, the random variable τ_{mp} is finite almost surely.

Similarly to the reasoning above we get the inequalities

$$a^{2(\tau_{mp}-m)}\sum_{k=1}^{\tau_{mp}-m} a^{-k}k^{-\rho}\xi_k^2 \geqslant a^{2\tau_{mp}-m}a^{-\tau_{mp}-m}\tau_{mp}{}^{-\rho}c = a^{\tau_{mp}}(\tau_{mp}-m)^{-\rho}c,$$
$$\sum_{k=\tau_{mp}-m+1}^{\infty} a^k k^{-\rho}\xi_k^2 \leqslant a^{\tau_{mp}-m}(\tau_{mp}-m)^{-\rho}(m/p + a^m\sum_{k=1}^{\infty} a^k\xi_{k+\tau_{mp}}^2).$$

Chose $m \in (-\frac{\log p}{\log a}, \sqrt{p}/2)$, i.e. $m/p < 1/(2\sqrt{p})$ and $a^m < 1/p$. Then

$$
\begin{aligned}
&\mathrm{P}\left(\exists n \geqslant 1 \quad \frac{a^{2n}\sum_{k=1}^{n} a^{-k}k^{-\rho}\xi_k^2}{\sum_{k=n+1}^{\infty} a^k k^{-\rho}\xi_k^2} \geqslant c\sqrt{p}\right) \\
&\geqslant \mathrm{P}\left(\frac{a^{\tau_{mp}}(\tau_{mp}-m)^{-\rho}c}{a^{\tau_{mp}-m}(\tau_{mp}-m)^{-\rho}(m/p + a^m\sum_{k=1}^{\infty} a^k\xi_{k+\tau_{mp}}^2)} \geqslant c\sqrt{p}\right) \\
&\geqslant \mathrm{P}\left(1/(2\sqrt{p}) + 1/p\sum_{k=1}^{\infty} a^k\xi_{k+\tau_{mp}}^2 \leqslant 1/\sqrt{p}\right) \\
&\geqslant \mathrm{P}\left(\sum_{k=1}^{\infty} a^k\xi_{k+\tau_{mp}}^2 \leqslant \sqrt{p}/2\right) \to 1, \quad \text{as } p \to \infty
\end{aligned}
$$

and we again obtain (4.9), the failure of the noise condition.

The conclusion from the above reasoning is that if $\{\xi_k\}$ are i.i.d. and $\lambda_k^2 \asymp a^k$ where $a \in (0,1)$, then assumption (4.8) is true if $\mathrm{P}(|\xi_k| \in [\varepsilon, \varepsilon^{-1}]) = 1$ for some $\varepsilon > 0$, that is, the support of ξ_k is separated from 0 and ∞. The sufficiency follows from the deterministic statement.

Remark 7. It is interesting that the Muckenhoupt-type condition fails for a typical random noise in the case of severely ill-posed problems. This observation, however, is in line with numerical investigation on the performance of heuristic rules done, for instance, by Hämarik, Palm, and Raus [12], in particular in [22]. Typically, for mildly ill-posed problems, the quasi-optimality principle is amongst the most efficient heuristic rules. However, for the backward heat equation (which is severely ill-posed), it performs worse compared to competitors such as the Hanke–Raus rules (see, e.g. [13, 14]) which by our results can be understood as caused by the failure of the noise condition. Note that the convergence theory for the latter rules is based on a weaker Muckenhoupt-type condition which might not suffer from the negative result in theorem 3. Thus, the restricted noise analysis clearly reveals the behaviour of heuristic rules, which was quite mysterious for a long time.

5. The quasi-optimality criterion in the aggregation of the regularized approximants: numerical illustration

In this section, we illustrate how the quasi-optimality criterion can be used in the aggregation of the regularized approximants by means of the linear functional strategy. Recall that the idea of such an aggregation is to approximate the best linear combination

$$x_{\mathrm{agg}}^s = \sum_{j=1}^{s} c_j^s x_{\alpha_j}^\delta$$

of the constructed regularized approximants $x_{\alpha_j}^\delta$ of x, where 'best' means that x_{agg}^s solves the minimization problem

$$\|x - x_{\mathrm{agg}}^s\| = \min_{c_j}\left\| x - \sum_{j=1}^{s} c_j x_{\alpha_j}^\delta \right\|.$$

It is clear that the vector $\mathbf{c}^s = (c_1^s, c_2^s, \ldots, c_s^s) \in \mathbb{R}^s$ satisfies the system of linear equations $\mathbf{Gc} = \mathbf{p}$ with the Gram matrix $\mathbf{G} = \left(\left\langle x_{\alpha_i}^\delta, x_{\alpha_j}^\delta \right\rangle : \, i,j = 1,2,\ldots,s\right)$ and the vector $\mathbf{p} = \left(\left\langle x, x_{\alpha_i}^\delta \right\rangle : \, i = 1,2,\ldots,s\right)$. Since $x_{\alpha_j}^\delta, \ j = 1,2,\ldots,s$, are already found, the matrix $\mathbf{G}$ can be computed and the calculation of the inverse matrix $\mathbf{G}^{-1}$ can be controlled. However, the vector $\mathbf{p}$ involves the unknown solution x, and therefore, the system $\mathbf{Gc} = \mathbf{p}$ cannot be solved directly.

At the same time, each component $\left\langle x, x_{\alpha_i}^\delta \right\rangle$ of the vector $\mathbf{p}$ is a value of a bounded linear functional $x_{\alpha_i}^\delta$, and the linear functional strategy allows us to estimate $\left\langle x, x_{\alpha_i}^\delta \right\rangle$, $i = 1,2,\ldots,s$, more accurately than x in $\|\cdot\|$. For example, if $x \in \mathrm{R}\left(\varphi\left(T^*T\right)\right)$ and $x_\alpha^\delta = (\alpha I + T^*T)^{-1} T^* y^\delta$, then under the conditions of theorem 1, we have

$$\left\| x - x_{\alpha(y^\delta)}^\delta \right\| = O\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right), \tag{5.1}$$

while for each α_i, the quasi-optimality criterion in the linear functional strategy gives us $\alpha_i\left(y^\delta\right) = \alpha_{\varkappa_i}\left(y^\delta\right)$ such that

$$\left| \left\langle x, x_{\alpha_i}^\delta \right\rangle - \left\langle x_{\alpha_i(y^\delta)}^\delta, x_{\alpha_i}^\delta \right\rangle \right| = o\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right), \tag{5.2}$$

where $\varkappa_i$ is an index function for which $x_{\alpha_i}^\delta \in \mathrm{R}\left(\varkappa_i\left(T^*T\right)\right)$.

Consider now

$$\mathbf{p}_{y^\delta} = \left(\left\langle x_{\alpha_i(y^\delta)}^\delta, x_{\alpha_i}^\delta \right\rangle, \ i = 1,2,\ldots,s \right), \quad \mathbf{c}_{y^\delta}^s = \left(c_{1,y^\delta}^s, c_{2,y^\delta}^s, \ldots, c_{s,y^\delta}^s \right) = \mathbf{G}^{-1}\mathbf{p}_{y^\delta}$$

and

$$x_{\mathrm{agg},y^\delta}^s = \sum_{j=1}^s c_{j,y^\delta}^s x_{\alpha_j}^\delta. \tag{5.3}$$

Note that $x_{\mathrm{agg},y^\delta}^s$ can be effectively computed because it only uses access to T and y^δ. Then by the same arguments as in the proof of theorem 3.7 in [9], it follows from (5.2) that

$$\begin{aligned} \left\| x - x_{\mathrm{agg},y^\delta}^s \right\| &= \min_{c_j} \left\| x - \sum_{j=1}^s c_j x_{\alpha_j}^\delta \right\| + o\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right) \\ &= \left\| x - x_{\mathrm{agg}}^s \right\| + o\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right). \end{aligned} \tag{5.4}$$

If

$$\alpha\left(y^\delta\right) \in \{\alpha_j, \ j = 1,2,\ldots,s\}, \tag{5.5}$$

then the accuracy of x_{agg}^s may only be better than the one of $x_{\alpha(y^\delta)}^\delta$. Moreover, from (5.1), (5.4), it follows that the error of the effectively computed aggregator $x_{\mathrm{agg},y^\delta}^s$ differs from the error of x_{agg}^s by a quantity of higher order than the accuracy guaranteed by the standard quasi-optimality criterion. In this way, a combination of the linear functional strategy and the quasi-optimality criterion resulting in (5.3) may improve the accuracy of the latter one. Such improvement indeed is observed in the numerical illustrations below.

Note that the family of the regularized approximations $\left\{ x_{\alpha_j}^\delta \right\}$ may consist only of a single approximant $x_{\alpha_i}^\delta$. Then the value of

Table 1. Performance in terms of errors.

Error	Mean value	Simulation on figure 1
e_{qo}	0.079	0.076
e_{best}	0.067	0.064
$e_{\mathrm{qo},2}$	0.075	0.060
$e_{\mathrm{best},2}$	0.065	0.058
e_{agg}	0.065	0.057

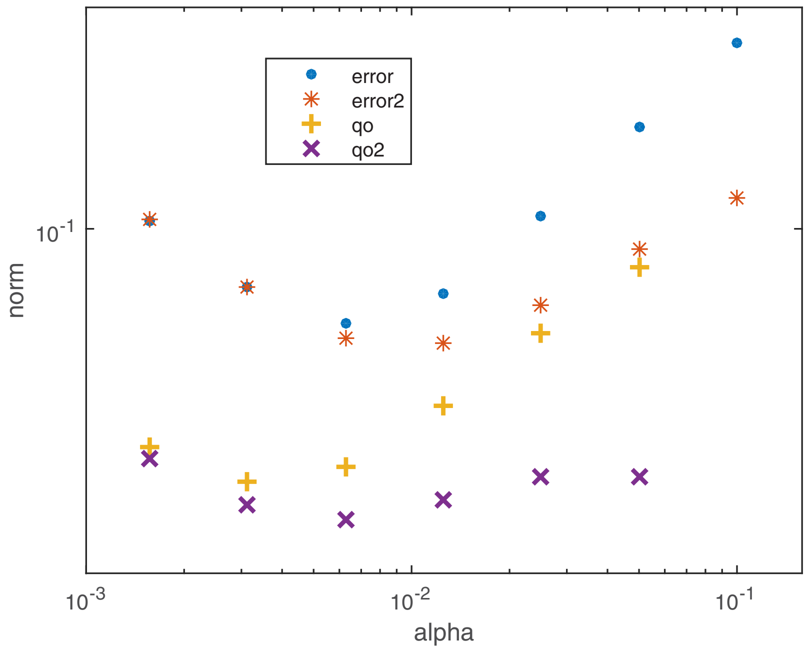

Figure 1. The quantities observed in a particular simulation: $\left\|x - x_{\alpha_i}^{\delta}\right\|$ (error), $\left\|x - \bar{x}_{\alpha_i}^{\delta}\right\|$ (error2), $\left\|x_{\alpha_i}^{\delta} - x_{\alpha_{i-1}}^{\delta}\right\|$ (qo), $\left\|\bar{x}_{\alpha_i}^{\delta} - \bar{x}_{\alpha_{i-1}}^{\delta}\right\|$ (qo2), plotted against the corresponding values of α_i, $i = 1, 2, \ldots, 7$. The regularization parameter $\alpha_1 = 0.1$ is on the right side of the picture, and α_7 is on the left side. In this simulation, the regularization parameter α_6 is selected by the quasi-optimality criterion (5.8) (the smallest qo-norm), and the regularization parameter α_5 is selected by the iterated quasi-optimality criterion (5.9) (the smallest qo2-norm).

$$c_i^* = \operatorname*{argmin}_c \left\|x - c x_{\alpha_i}^{\delta}\right\|$$

can be explicitly written as

$$c_i^* = \frac{\left\langle x, x_{\alpha_i}^{\delta}\right\rangle}{\left\|x_{\alpha_i}^{\delta}\right\|^2},$$

and can be interpreted as a correction factor for $x_{\alpha_i}^{\delta}$. If a value $\alpha = \alpha\left(y^{\delta}\right)$ has been already selected by the quasi-optimality criterion, then c_i^* can be approximated by

$$c_{i,y^\delta} = \frac{\left\langle x^\delta_{\alpha(y^\delta)}, x^\delta_{\alpha_i} \right\rangle}{\left\| x^\delta_{\alpha_i} \right\|^2}, \tag{5.6}$$

and under the conditions of theorem 1, we have

$$\left| c_i^* - c_{i,y^\delta} \right| = o\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right).$$

After calculating (5.6) for each considered α_i, we can construct a corrected family of regularized approximants $\left\{ \bar{x}^\delta_{\alpha_i} = c_{i,y^\delta} x^\delta_{\alpha_i} \right\}$ such that

$$\left\| x - \bar{x}^\delta_{\alpha_i} \right\| = \min_c \left\| x - c x^\delta_{\alpha_i} \right\| + o\left(\varphi\left(\varphi\left(\theta^{-1}(\delta)\right)\right) \right).$$

If (5.5) is satisfied, then by the same reason as above, the corrected family $\left\{ \bar{x}^\delta_{\alpha_i} \right\}$ may contain elements approximating x better than $x^\delta_{\alpha(y^\delta)}$ that suggests second or iterated application of the quasi-optimality criterion, this time to the corrected family $\left\{ \bar{x}^\delta_{\alpha_i} \right\}$. This iterated quasi-optimality criterion will also be illustrated below.

Recall that the usual way (see [23]) of implementing the quasi-optimality criterion consists in selecting $\alpha = \alpha\left(y^\delta\right) = \alpha_\ell$ from a geometric sequence

$$\left\{ \alpha_j = \alpha_1 q^{j-1},\ j = 1, 2, \ldots, M \right\},\ 0 < \alpha_1,\ q < 1, \tag{5.7}$$

such that

$$\left\| x^\delta_{\alpha_\ell} - x^\delta_{\alpha_{\ell-1}} \right\| = \min \left\{ \left\| x^\delta_{\alpha_j} - x^\delta_{\alpha_{j-1}} \right\|,\ j = 2, 3, \ldots, M \right\}. \tag{5.8}$$

In the same spirit, we can implement the above mentioned iterated quasi-optimality criterion suggesting $\alpha = \bar{\alpha}\left(y^\delta\right) = \alpha_k$ such that

$$\begin{aligned} \left\| \bar{x}^\delta_{\alpha_k} - \bar{x}^\delta_{\alpha_{k-1}} \right\| &= \min \Big\{ \left\| \bar{x}^\delta_{\alpha_j} - \bar{x}^\delta_{\alpha_{j-1}} \right\| \\ &= \left\| c_{j,y^\delta} x^\delta_{\alpha_j} - c_{j-1,y^\delta} x^\delta_{\alpha_{j-1}} \right\|,\ j = 2, 3, \ldots, M \Big\}. \end{aligned} \tag{5.9}$$

Note that the rule (5.8) is in fact a discretization of the quasi-optimality criterion considered above because $\psi\left(\alpha, y^\delta\right)$ can be written (see, e.g. [17]) as

$$\psi\left(\alpha, y^\delta\right) = \alpha \left\| \frac{\partial x^\delta_\alpha}{\partial \alpha} \right\|,$$

and (5.8) is just a backward difference approximation of the derivative $\frac{\partial x^\delta_\alpha}{\partial \alpha}$ on the mesh nodes (5.7), i.e.

$$\alpha \frac{\partial x^\delta_\alpha}{\partial \alpha}\bigg|_{\alpha=\alpha_j} \approx \alpha_j \frac{x^\delta_{\alpha_j} - x^\delta_{\alpha_{j-1}}}{\alpha_j - \alpha_{j-1}} = (q-1)^{-1} \left(x^\delta_{\alpha_j} - x^\delta_{\alpha_{j-1}} \right). \tag{5.10}$$

From this view point, the iterated quasi-optimality criterion (5.9) can be seen as the use of another difference formula to approximate $\frac{\partial x^\delta_\alpha}{\partial \alpha}$, i.e.

$$\alpha \frac{\partial x^\delta_\alpha}{\partial \alpha}\bigg|_{\alpha=\alpha_j} \approx \alpha_j \frac{c_{j,y^\delta} x^\delta_{\alpha_j} - c_{j-1,y^\delta} x^\delta_{\alpha_{j-1}}}{\alpha_j - \alpha_{j-1}} = (q-1)^{-1} \left(\bar{x}^\delta_{\alpha_j} - \bar{x}^\delta_{\alpha_{j-1}} \right). \tag{5.11}$$

The quasi-optimality criterion in the linear functional strategy is associated with the function $\psi_{\varkappa}\left(\alpha, y^{\delta}\right)$ that is a particular form of the quantity used in the so-called weighted quasi-optimality criterion discussed in [6] (see definition 2.5 there). At the same time, $\psi_{\varkappa}\left(\alpha, y^{\delta}\right)$ is, up to a constant multiplier, the upper bound for all functions

$$U_f\left(\alpha, y^{\delta}\right) = \alpha \left|\left\langle f, \frac{\partial x_{\alpha}^{\delta}}{\partial \alpha}\right\rangle\right|$$

with $f \in \mathrm{R}\left(\varkappa\left(T^*T\right)\right)$. Therefore, in view of (5.10), (5.11), for a given f, say $f = x_{\alpha_i}^{\delta}$, it is reasonable to use the following discretized version of the quasi-optimality criterion in the linear functional strategy: choose $\alpha_i\left(y^{\delta}\right) = \alpha_{\kappa_i}$ from (5.7) such that

$$\left|\left\langle x_{\alpha_i}^{\delta}, x_{\alpha_{\kappa_i}}^{\delta} - x_{\alpha_{\kappa_i - 1}}^{\delta}\right\rangle\right| = \min\left\{\left|\left\langle x_{\alpha_i}^{\delta}, x_{\alpha_j}^{\delta} - x_{\alpha_{j-1}}^{\delta}\right\rangle\right|, \ j = 2, 3, \ldots, M\right\}. \tag{5.12}$$

To illustrate the quasi-optimality criterion in the aggregation (5.3), (5.12), we simulate the data by (1.1), where T is a matrix $T = (t_{ij})$, where $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ with the non-zero entries $t_{kk} = a^k$, $0 < a < 1$, x is a vector $x = (x_j = j^{-\mu}\eta_j,\ j = 1, 2, \ldots, n)$, and η_j are randomly sampled from the uniform distribution on $[-1, 1]$. We take $a = 0.5$, $\mu = 2$, $n = 100$, $m = 150$.

Our simulation mimics a severely ill-posed problem because the singular values $\lambda_k^2 = t_{kk}^2 = a^{2k}$ of T^*T decrease exponentially, while the Fourier coefficients x_j of x in the corresponding basis decrease only polynomially. A reason to consider this case is that, as it can be seen from theorem 1, for severely ill-posed problems, the difference between the estimation of the solution and the functional estimation is the most noticeable. For example, if $\varphi(\lambda) = \log^{-\nu}\frac{1}{\lambda}$, $\nu > 0$, which corresponds to the severely ill-posed case, then the quasi-optimality criterion can guarantee an accuracy of order $O\left(\log^{-\nu}\log\frac{1}{\delta}\right)$ for an approximation of x, while the value of a bounded linear functional $\langle f, x\rangle$ can be estimated with the use of the quasi-optimality criterion much more accurately, say with the accuracy of order $O\left(\delta^{2\gamma^2}\log^{-\nu\left(1+\gamma-2\gamma^2\right)}\frac{1}{\delta}\right)$ when $f \in \mathrm{R}\left(\left(T^*T\right)^{\gamma}\right)$, $0 < \gamma < 1/2$.

Numerical illustrations below demonstrate that in the considered simulation scenario, the aggregation (5.3) and (5.12), which is based on the quasi-optimality criterion and the linear functional strategy, improves the accuracy resulting from the quasi-optimality criterion and performs at the level of the best (but unknown) regularization parameter choice.

To guarantee almost surely that the Muckenhoupt-type condition (2.10) on the noise ξ is satisfied in our test, we simulate ξ as $\xi = (\xi_i,\ i = 1, 2, \ldots, m)$, $m = 150$, where ξ_i are randomly sampled from the uniform distribution on $[-1, -\delta] \cup [\delta, 1]$, $\delta > 0$, such that the noise support is separated from 0 and ∞, as it is suggested in remark 6 discussed in the previous section.

The random simulations of ξ and x are performed 10 times, and the noise intensity is chosen as $\delta = 0.01$. The regularized approximants $x_{\alpha_i}^{\delta}$ are constructed by the Tikhonov regularization, i.e.

$$x_{\alpha_i}^{\delta} = \left(\alpha_i I + T^*T\right)^{-1} T^* y^{\delta},$$

where α_i are taken from (5.7) with $\alpha_1 = 0.1$, $q = 0.5$, $M = 20$. Moreover, in each simulation, the quasi-optimal regularization parameters $\alpha = \alpha\left(y^{\delta}\right)$, $\alpha = \bar{\alpha}\left(y^{\delta}\right)$ are chosen according to (5.8), (5.9). To guarantee condition (5.5), we aggregate in (5.3) the regularized approximants

$x^{\delta}_{\alpha_i}$ with $\alpha_i \geqslant \alpha\left(y^{\delta}\right)$. An aggregation on a wider set of approximants does not improve the accuracy, as it has been observed.

The performance of the regularized approximants is measured in terms of the following quantities:

$$\begin{aligned} e_{\text{qo}} &= \left\| x - x^{\delta}_{\alpha(y^{\delta})} \right\|, \quad e_{\text{qo},2} = \left\| x - \bar{x}^{\delta}_{\bar{\alpha}(y^{\delta})} \right\|, \\ e_{\text{best}} &= \min\left\{ \left\| x - x^{\delta}_{\alpha_i} \right\|,\ i = 1, 2, \ldots, M \right\}, \\ e_{\text{best},2} &= \min\left\{ \left\| x - \bar{x}^{\delta}_{\alpha_i} \right\|,\ i = 1, 2, \ldots, M \right\}, \\ e_{\text{agg}} &= \left\| x - x^{s}_{\text{agg},y^{\delta}} \right\|, \end{aligned}$$

where $s = \max\left\{ i:\ \alpha_i \geqslant \alpha\left(y^{\delta}\right) \right\}$, and $x^{s}_{\text{agg},y^{\delta}}$ is given by (5.3), (5.12). The mean values of the considered quantities over the performed simulations are given in table 1. The table also reports the values observed in a particular simulation displayed in figure 1.

The presented illustration confirms that for severely ill-posed problems, the aggregation based on the linear functional strategy is able to perform at the level of the best, but unknown, regularization parameter choice.

Acknowledgments

This research was partially supported by the AMMODIT project 645672 (Approximation Methods for Molecular Modelling and Diagnosis Tools) in the frame of Horizon 2020 program. Sergiy Pereverzyev Jr gratefully acknowledges the support of the Austrian Science Fund (FWF): project P 29514-N32. Stefan Kindermann is supported by the Austrian Science Fund (FWF) project P 30157-N31.

Appendix. Proof of theorem 2

We first need the following known result.

Lemma A.1. *Assume that random variables $\{Y_n\}$ have the finite second moment and*

$$\lim_{n\to\infty} \mathrm{E}Y_n = 0, \quad \sum_{n=1}^{\infty} \mathrm{Var}Y_n < \infty.$$

Then

$$Y_n \to 0, \text{ as } n \to \infty \text{ almost surely.} \tag{A.1}$$

Proof. Indeed,

$$\mathrm{E}\sum_n (Y_n - \mathrm{E}Y_n)^2 = \sum_n \mathrm{E}(Y_n - \mathrm{E}Y_n)^2 = \sum_n \mathrm{Var}Y_n < \infty.$$

So, $Y_n - \mathrm{E}Y_n \to 0$ as $n \to \infty$ almost surely, and we get (A.1) because $\lim_{n\to\infty} \mathrm{E}Y_n = 0$. □

Proof of theorem 2. We show (4.8) if $\beta > \rho - 1$. In particular, (4.1) is a particular case of (4.8) if $\mathrm{P}(\xi_k = \pm 1) = 1/2$.

To prove (4.8), it suffices to verify that

$$P\left(\sup_{n\geqslant 1}\frac{n^{-2\beta}\sum_{k=1}^{n}k^{\beta}k^{-\rho}\xi_k^2}{\sum_{k=n+1}^{\infty}k^{-\beta}k^{-\rho}\xi_k^2}<\infty\right)=P\left(\sup_{n\geqslant 1}\frac{n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}\xi_k^2}{\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\xi_k^2}<\infty\right)=1. \tag{A.2}$$

Set $\eta_k=\xi_k^2-1$. Recall that $E\eta_k=0$. We have

$$n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}\xi_k^2=n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}(1+(\xi_k^2-1))$$
$$=n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}+n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}\eta_k=\frac{n^{-\beta-\rho+1}}{\beta-\rho+1}(1+o(1))+n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}\eta_k.$$

Analogously,

$$\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\xi_k^2=\sum_{k=n+1}^{\infty}k^{-\beta-\rho}+\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k=\frac{n^{-\beta-\rho+1}}{-\beta-\rho+1}(1+o(1))+\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k.$$

So, equation (A.2) will be verified if we prove that

$$n^{\beta+\rho-1}n^{-2\beta}\sum_{k=1}^{n}k^{\beta-\rho}\eta_k=n^{-\beta+\rho-1}\sum_{k=1}^{n}k^{\beta-\rho}\eta_k\to 0\quad\text{as } n\to\infty\ \text{ almost surely} \tag{A.3}$$

and

$$n^{\beta+\rho-1}\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k\to 0\quad\text{as } n\to\infty\ \text{ almost surely.} \tag{A.4}$$

Consider (A.4). Set $Y_n:=\left(n^{\beta+\rho-1}\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k\right)^2$. Since $E\eta_k=0$, we have

$$EY_n=E\left(n^{\beta+\rho-1}\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k\right)^2=\mathrm{Var}\left(n^{\beta+\rho-1}\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k\right)$$
$$=n^{2\beta+2\rho-2}\sum_{k=n+1}^{\infty}\mathrm{Var}(k^{-\beta-\rho}\eta_k)=n^{2\beta+2\rho-2}\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}\mathrm{Var}(\eta_k)$$
$$\leqslant C_1n^{2\beta+2\rho-2}\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}\leqslant C_2n^{2\beta+2\rho-2}n^{-2\beta-2\rho+1}$$
$$=C_2n^{-1}\to 0,\quad\text{as } n\to\infty.$$

$$
\begin{aligned}
\mathrm{Var}Y_n &= \mathrm{Var}\left[\left(n^{\beta+\rho-1}\sum_{k=n+1}^{\infty}k^{-\beta-\rho}\eta_k\right)^2\right]\\
&= n^{4\beta+4\rho-4}\mathrm{Var}\left[\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}\eta_k^2+2\sum_{n+1\leqslant i<j}i^{-\beta-\rho}j^{-\beta-\rho}\eta_i\eta_j\right]\\
&\leqslant n^{4\beta+4\rho-4}\mathrm{Var}\left[\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}(\eta_k^2-\mathrm{E}\eta_k)+2\sum_{n+1\leqslant i<j}i^{-\beta-\rho}j^{-\beta-\rho}\eta_i\eta_j\right]\\
&= n^{4\beta+4\rho-4}\mathrm{E}\left[\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}(\eta_k^2-\mathrm{E}\eta_k)+2\sum_{n+1\leqslant i<j}i^{-\beta-\rho}j^{-\beta-\rho}\eta_i\eta_j\right]^2\\
&\leqslant 2n^{4\beta+4\rho-4}\left(\mathrm{E}\left[\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}(\eta_k^2-\mathrm{E}\eta_k^2)\right]^2+\mathrm{E}\left[2\sum_{n+1\leqslant i<j}i^{-\beta-\rho}j^{-\beta-\rho}\eta_i\eta_j\right]^2\right).
\end{aligned}
$$

If we expand the brackets in the last sum, then the expectation $\mathrm{E}(\eta_i\eta_j\eta_{i_1}\eta_{j_1})$ is equal to zero if $(i,j)\neq(i_1,j_1)$ and $(i,j)\neq(j_1,i_1)$. Thus the right hand side of the last expression equals

$$
\begin{aligned}
&n^{4\beta+4\rho-4}\left(\mathrm{Var}\left[\sum_{k=n+1}^{\infty}k^{-2\beta-2\rho}(\eta_k^2-\mathrm{E}\eta_k^2)\right]+8\sum_{n+1\leqslant i<j}\mathrm{E}\left[i^{-\beta-\rho}\eta_i\right]^2\mathrm{E}\left[j^{-\beta-\rho}\eta_j\right]^2\right)\\
&= n^{4\beta+4\rho-4}\left(\sum_{k=n+1}^{\infty}k^{-4\beta-4\rho}\mathrm{Var}\left[(\eta_k^2-\mathrm{E}\eta_k^2)\right]\right.\\
&\qquad\qquad\left.+8\sum_{n+1\leqslant i<j}\mathrm{Var}\left[i^{-\beta-\rho}\eta_i\right]\mathrm{Var}\left[j^{-\beta-\rho}\eta_j\right]\right)\\
&\leqslant C_3n^{4\beta+4\rho-4}\left(\sum_{k=n+1}^{\infty}k^{-4\beta-4\rho}+\sum_{n+1\leqslant i<j}i^{-2\beta-2\rho}\mathrm{Var}(\eta_i)j^{-2\beta-2\rho}\mathrm{Var}(\eta_j)\right)\\
&\leqslant C_4n^{4\beta+4\rho-4}\left(n^{-4\beta-4\rho+1}+(\sum_{n+1\leqslant i}i^{-2\beta-2\rho})^2\right)\\
&\leqslant C_5n^{4\beta+4\rho-4}\left(n^{-4\beta-4\rho+1}+(n^{-2\beta-2\rho+1})^2\right)\leqslant C_6n^{4\beta+4\rho-4}n^{-4\beta-4\rho+2}=C_6n^{-2}.
\end{aligned}
$$

This proves (A.4).

Consider (A.3). Set $Y_n:=n^{-2\beta+2\rho-2}(\sum_{k=1}^{n}k^{\beta-\rho}\eta_k)^2$ in lemma A.1. Similarly to the above calculations we get $\lim_{n\to\infty}\mathrm{E}Y_n=0$ and

$$
\begin{aligned}
\mathrm{Var}(Y_n) &= n^{-4\beta+4\rho-4}\,O\left(\sum_{k=1}^{n}k^{4\beta-4\rho}+(\sum_{k=1}^{n}k^{2\beta-2\rho})^2\right)\\
&= n^{-4\beta+4\rho-4}\,O\left(\left(\sum_{k=1}^{n}k^{2\beta-2\rho}\right)^2\right).
\end{aligned}
\tag{A.5}
$$

In contrast to (convergent) sums of the form $\sum_{k=n+1}^{\infty} k^{-\theta} \asymp n^{-\theta+1}$, the asymptotic of $\sum_{k=1}^{n} k^{-\theta}$ is different:

$$\sum_{k=1}^{n} k^{-\theta} \asymp \begin{cases} n^{-\theta+1}, & \theta < 1; \\ \log n, & \theta = 1; \\ 1 = n^0, & \theta > 1. \end{cases} = \begin{cases} n^{(-\theta+1)\vee 0}, & \theta \neq 1; \\ \log n, & \theta = 1. \end{cases}$$

That is why, we have to be careful in (A.5). In any case, $\lim_{n\to\infty} \mathrm{Var}(Y_n) = 0$ and the series $\sum_n \mathrm{Var}(Y_n)$ is convergent if $-4\beta + 4\rho - 4 < -1$ or $\beta > \rho - 1 + \frac{1}{4} = \rho - \frac{3}{4}$. Thus, we have already proved (A.3) for $\beta > \rho - \frac{3}{4}$. To verify (A.3) for $\beta > \rho - 1$, we have to consider moments of higher orders.

Considering $Y_n^{(m)} := n^{-2m\beta+2m\rho-2m}(\sum_{k=1}^{n} k^{\beta-\rho}\eta_k)^{2m}$ and performing similar calculation as above we get

$$\mathrm{Var}(Y_n^{(m)}) = n^{-4m\beta+4m\rho-4m}\, O\Bigg(\sum_{k_1+\ldots+k_p=2m,\ k_i\geqslant 2} \sum_{i_1=1}^{n} i_2^{k_1(\beta-\rho)} \sum_{i_2=1}^{n} i_2^{k_2(\beta-\rho)} \ldots \\ \ldots \sum_{i_p=1}^{n} i_p^{k_p(\beta-\rho)} \Bigg).$$

It can be seen that $\sum_n \mathrm{Var} Y_n^{(m)} < \infty$ if $-4m\beta + 4m\rho - 4m > -1$ or $\beta > \rho - 1 + \frac{1}{m}$. So, we have (A.3) for $\beta > \rho - 1 + \frac{1}{m}$. Since $m \geqslant 1$ is arbitrary, this yields (A.3) for $\beta > \rho - 1$. □

ORCID iDs

Stefan Kindermann https://orcid.org/0000-0002-3688-5125
Sergiy Pereverzyev Jr https://orcid.org/0000-0002-8627-3995
Andrey Pilipenko https://orcid.org/0000-0002-8786-1617

References

[1] Anderssen R 1986 The linear functional strategy for improperly posed problems *Inverse Problems* (*Int. Series of Numerical Mathematics* vol 77) ed J R Cannon and U Hornung (Basel: Birkhüser) pp 11–30

[2] Anderssen R S 1980 On the use of linear functionals for Abel-type integral equations in applications *The Application and Numerical Solution of Integral Equations* ed F De Hoog and M A Lukas (Alphen aan den Rijn: Sijthoff and Noordhof International Publishers) pp 195–221

[3] Anderssen R S and Engl H W 1991 The role of linear functionals in improving convergence rates for parameter identification via Tikhonov regularization *Inverse Problems in Engineering Sciences, Satellite Conf. Proc.* ed M Yamaguti (Berlin: Springer) pp 1–10

[4] Bakushinskii A B 1984 Remarks on choosing regularization parameter using the quasi-optimality and ratio criterion *USSR Comput. Math. Math. Phys.* **24** 181–2

[5] Bauer F, Mathé P and Pereverzev S 2007 Local solutions to inverse problems in geodesy *J. Geod.* **81** 39–51

[6] Bauer F and Reiß M 2008 Regularization independent of the noise level: an analysis of quasi-optimality *Inverse Problems* **24** 055009

[7] Becker S M A 2011 Regularization of statistical inverse problems and the Bakushinskii veto *Inverse Problems* **27** 115010

[8] Bellman R, Kashef B G and Casti J 1972 Differential quadrature: a technique for the rapid solution of nonlinear partial differential equations *J. Comput. Phys.* **10** 40–52
[9] Chen J, Pereverzyev S Jr and Xu Y 2015 Aggregation of regularized solutions from multiple observation models *Inverse Problems* **31** 075005
[10] Engl H W and Neubauer A 1988 A parameter choice strategy for (iterated) Tikhonov regularization of ill-posed problems leading to superconvergence with optimal rates *Appl. Anal.* **27** 5–18
[11] Goldenshluger A and Pereverzev S V 2000 Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations *Probab. Theory Relat. Fields* **118** 169–86
[12] Hämarik U, Palm R and Raus T 2011 Comparison of parameter choices in regularization algorithms in case of different information about noise level *Calcolo* **48** 47–59
[13] Hanke M and Raus T 1996 A general heuristic for choosing the regularization parameter in ill-posed problems *SIAM J. Sci. Comput.* **17** 956–72
[14] Jin Q 2016 Hanke–Raus heuristic rule for variational regularization in Banach spaces *Inverse Problems* **32** 085008
[15] Kindermann S and Neubauer A 2008 On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization *Inverse Problems Imaging* **2** 291–9
[16] Kusche J and Klees R 2002 Regularization of gravity field estimation from satellite gravity gradients *J. Geod.* **76** 359–68
[17] Leonov A S 1991 On the accuracy of Tikhonov regularizing algorithms and quasioptimal selection of a regularization parameter *Sov. Math.—Dokl.* **44** 711–6
[18] Louis A K and Maass P 1990 A mollifier method for linear operator equations of the first kind *Inverse Problems* **6** 427–40
[19] Lu S and Pereverzev S V 2013 *Regularization Theory for Ill-Posed Problems: Selected Topics* (Boston, MA: Walter de Gruyter)
[20] Mathé P and Pereverzev S V 2002 Direct estimation of linear functionals from indirect noisy observations *J. Complexity* **18** 500–16
[21] Neubauer A 2008 The convergence of a new heuristic parameter selection criterion for general regularization methods *Inverse Problems* **24** 055005
[22] Palm R 2010 Numerical comparison of regularization algorithms for solving ill-posed problems *PhD Thesis* Institute of Computer Science, University of Tartu
[23] Tikhonov A N and Glasko V B 1965 Use of the regularization method in non-linear problems *USSR Comput. Math. Math. Phys.* **5** 93–107