Multi-penalty regularization with a component-wise penalization

2013 Inverse Problems 29 075002

(http://iopscience.iop.org/0266-5611/29/7/075002)

Inverse Problems **29** (2013) 075002 (15pp) doi:10.1088/0266-5611/29/7/075002

Multi-penalty regularization with a component-wise penalization

V Naumova and S V Pereverzyev

Johann Radon Institute for Computational and Applied Mathematics (RICAM),
Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria

E-mail: valeriya.naumova@oeaw.ac.at

Received 20 February 2013, in final form 15 April 2013
Published 3 June 2013
Online at stacks.iop.org/IP/29/075002

Abstract

We discuss a new regularization scheme for reconstructing the solution of a linear ill-posed operator equation from given noisy data in the Hilbert space setting. In this new scheme, the regularized approximation is decomposed into several components, which are defined by minimizing a multi-penalty functional. We show theoretically and numerically that under a proper choice of the regularization parameters, the regularized approximation exhibits the so-called compensatory property, in the sense that it performs similar to the best of the single-penalty regularization with the same penalizing operator.

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

1. Introduction

In this paper, we address the solution of a linear ill-posed problem

$$Ax = y, \tag{1}$$

where $A : X \to Y$ is a bounded linear operator between Hilbert spaces X and Y with the non-closed range $\mathcal{R}(A)$. We denote the inner product and the corresponding norm on the Hilbert spaces by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. In the following, we assume that the operator A is injective and y belongs to $\mathcal{R}(A)$ such that there exists a unique solution $x^\dagger \in X$ of equation (1).

Moreover, typically (1) is only an idealized model in which noise has been neglected. In reality,

$$y_\delta = Ax^\dagger + \xi, \tag{2}$$

where $\xi \in Y$, $\|\xi\| \leqslant \delta$, $\delta \in (0, 1)$. Moreover, since it is assumed that $\mathcal{R}(A)$ is non-closed, the solution $x^\dagger$ does not depend continuously on data and can be reconstructed in a stable way from y_δ only by means of a regularization method [8].

Tikhonov–Phillips (TP) regularization is proved to be efficient for such a reconstruction. Recall that in this method, the regularized approximate solution x_α^δ of (1) is defined as the minimizer of the following functional:

$$TP_\alpha(x) := \|Ax - y_\delta\|^2 + \alpha\|x\|^2, \tag{3}$$

0266-5611/13/075002+15$33.00

with $\alpha > 0$ being a regularization parameter. Due to the simplicity and effectiveness of the method, this classical approach is very attractive to users and the minimizer x_α^δ can be numerically found either by solving the corresponding system of linear equations or by employing a suitable optimization tool.

At the same time, it is well known that the TP regularization suffers from a saturation effect [27, 18]. More precisely, this regularization method cannot guarantee an accuracy better than $O(\delta^{2/3})$ regardless of the smoothness of the solution $x^\dagger$.

On the other hand, this order can be potentially improved if one employs the original idea of Tikhonov [29] and changes the identity operator $\mathbb{I}$ in the penalty term in (3) for some unbounded operator B. Then, the regularized solution $x_{\beta,B}^\delta$ is defined as the minimizer of the functional

$$T_\beta(x) := \|Ax - y_\delta\|^2 + \beta\|Bx\|^2 \tag{4}$$

over the domain $\mathcal{D}(B)$ of the operator B.

In many practical applications, the operator B that influences the properties of the regularized approximant is chosen as a differential operator.

It is worthwhile emphasizing that the superiority of (4) over (3) is theoretically justified only under the assumption that the operators A and B are related by the so-called link condition. In the simplest case, this presupposes that B is a densely defined self-adjoint strictly positive-definite operator and for all $x \in X$ it holds

$$\|B^{-s}x\| \leqslant \|Ax\| \leqslant b\|B^{-s}x\|, \tag{5}$$

where $s > 0$ and $b \geqslant 1$ are some constants.

For more details, we refer to the classical paper [23]; see also [26, 22, 3] and references therein.

It is clear that condition (5) is a serious restriction and, what is even more important, the condition is sometimes hardly verifiable, as is the case, for instance, when Tikhonov regularization is used for solving nonlinear ill-posed equations [25, 28, 12]. For example, in [28] it is suggested to solve a nonlinear equation $F(x) = y_\delta$ iteratively by minimizing at each iteration a functional of the form (4) with $A = F'(x_k)$ given as the Frechét derivative of F calculated for the approximate solution x_k constructed on the previous iteration. It is clear that generally in such a situation the link condition (5) cannot be verified *a priori*.

At the same time, it may happen that the regularization (4) performs poorly when condition (5) is violated. To exemplify these kinds of difficulties, we refer to the section with numerical experiments and specifically to figure 2.

Thus, if condition (5) is not granted *a priori*, it is not clear, in general, which of the regularization methods is more suitable for a problem at hand, since the TP method (3) may not allow the accuracy of the best possible order, while the Tikhonov method (4) may fail without the link condition (5).

This opens room to more sophisticated methods such as multi-penalty (MP) regularization with a component-wise penalization, in which the following form of the regularization functional is used:

$$\Phi(\alpha, \beta; u, v) := \|A(u + v) - y_\delta\|^2 + \alpha\|u\|^2 + \beta\|Bv\|^2. \tag{6}$$

This form is inspired by the study [15] on the multiple kernel learning, where A is given as the so-called sampling operator, and the penalization of the components u and v is performed in different reproducing kernel Hilbert spaces. To the best of our knowledge, regularization based on the minimization of the functional (6) has never been studied so far in the context of regularization theory.

At the same time, the idea of decomposing the solution into different components has been very popular in imaging (see, e.g., [21, 31, 32]), where y_δ would be considered as a given noisy image, while u and v would be respectively a cartoon representation and an oscillatory component consisting of texture and noise. The difference between this context and the one which is studied in this paper is that in imaging u is *a priori* assumed to contain the main features of a real original image, and $A = \mathbb{I}$. Therefore, the dependence of the results with respect to parameters α and β is not too sensitive. In contrast, for a general operator A one cannot say *a priori* which of the components, u or v, is more important, and it can be detected only with a proper choice of the regularization parameters.

Note that the MP regularization is not a new topic in modern regularization theory, where in the case of two penalties one usually deals with the minimization of the functional

$$\Psi(\alpha, \beta; x) := \|Ax - y_\delta\|^2 + \alpha\|x\|^2 + \beta\|Bx\|^2. \tag{7}$$

Here, we may refer to the papers [2, 5, 7, 13, 17]. Our present study is stimulated by the remark made in [17], where in numerical experiments the authors observed the compensatory property of the MP regularization (7): this method performed similar to the best single-penalty regularization (3) or (4). However, no theoretical justification of this effect has been provided.

The primary goal of this paper is to demonstrate theoretically the similar compensatory property of the regularization (6) that will be done in the following section. In the final section with numerical experiments, we illustrate the efficiency of the proposed approach equipped with a heuristic parameter choice rule on a number of academic examples.

2. Convergence rates for multi-penalty regularization with component-wise penalization

As already mentioned in the introduction, the MP regularization could exhibit the compensatory property, at least numerically. In this section, we provide a theoretical justification of this property for the MP regularization (6). This will be done by analyzing two cases separately. First, we consider the case when the link condition (5) is violated. As follows from paper [19], in this situation one can still rely on the so-called source condition

$$x^\dagger = \varphi(A^*A)g, \qquad \|g\| \leqslant R, \tag{8}$$

where $\varphi : [0, \|A\|^2] \to [0, 1]$ is called an index function that is assumed to be continuous, increasing and such that $\varphi(0) = 0$ and $\frac{t}{\varphi(t)}$ is non-decreasing. Then, we analyze the case when condition (5) is satisfied.

Recall that in the case when the link condition is violated, the TP regularization (3) yields the maximal rate of accuracy $O(\delta^{2/3})$ that cannot be beaten in general regardless of the smoothness of $x^\dagger$, whereas in the situation when a problem at hand meets the link condition, the saturation effect can be postponed and, thus, better accuracy order may be achieved.

Before starting our analysis, we derive the formulas for the minimizers $u^\delta_{\alpha,\beta}$ and $v^\delta_{\alpha,\beta}$ of the functional $\Phi(\alpha, \beta; u, v)$. Using the standard technique of the calculus of variations, we obtain the following system of equations for the minimizers:

$$\begin{cases} A^*A\big(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}\big) - A^*y_\delta + \alpha u^\delta_{\alpha,\beta} = 0 \\ A^*A\big(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}\big) - A^*y_\delta + \beta B^2 v^\delta_{\alpha,\beta} = 0, \end{cases}$$

that allows the representation

$$u^\delta_{\alpha,\beta} = (\alpha\mathbb{I} + A^*A)^{-1}\big(A^*y_\delta - A^*Av^\delta_{\alpha,\beta}\big), \tag{9}$$

$$v^\delta_{\alpha,\beta} = \alpha(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta, \tag{10}$$

where $\mathbb{I}$ is the identity operator.

Remark 1. Note that from the above system, it follows that $\alpha u^{\delta}_{\alpha,\beta} = \beta B^2 v^{\delta}_{\alpha,\beta}$. This simple relation is helpful in understanding the difference between the MP regularizations (6) and (7). It tells us that if in (6) one of the regularization parameters, say α, is set to zero, then the minimization of (6) is reduced to a least-squares problem $\|Au - y_\delta\| \to \min$ regardless of the other parameter. At the same time, in the case of (7) by setting one of the parameters β or α to zero, we switch from a MP regularization to the single-penalty scheme (3) or (4). This remark may be seen as an explanation why in (6) the compensatory property is achieved when one of the regularization parameters is larger than 1, while for (7) this property was numerically observed in [17] when both of the parameters α and β are small. Below, we provide a theoretical justification of the compensatory property for the MP regularization (6). In the case of (7), such a justification is still to be provided.

2.1. Error bound under violated link condition

We will follow the convention that the symbol c denotes a number that does not depend on α, β, δ and may not be the same at different occurrences.

Theorem 1. *Let condition (8) be satisfied. Then, for a sufficiently small α and $\beta > 1$ we have the bound*

$$\left\|x^\dagger - \left(u^{\delta}_{\alpha,\beta} + v^{\delta}_{\alpha,\beta}\right)\right\| \leqslant c\left(\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}\right). \tag{11}$$

In addition, if the parameter α is chosen as $\alpha_{\rm opt} = \theta^{-1}_{\varphi}(\delta)$, where $\theta_\varphi(t) = \varphi(t)\sqrt{t}$, then an order optimal error bound

$$\left\|x^\dagger - \left(u^{\delta}_{\alpha_{\rm opt},\beta} + v^{\delta}_{\alpha_{\rm opt},\beta}\right)\right\| \leqslant c\varphi\left(\theta^{-1}_{\varphi}(\delta)\right) \tag{12}$$

is obtained.

Proof. Note that the bound (12) is a consequence of (11), and its optimality under condition (8) is proven in [20]. Therefore, only (11) needs to be proven.

From (9) and (10), it follows that

$$x^\dagger - \left(u^{\delta}_{\alpha,\beta} + v^{\delta}_{\alpha,\beta}\right) = x^\dagger - x^{\delta}_{\alpha} + (\alpha\mathbb{I} + A^*A)^{-1}A^*Av^{\delta}_{\alpha,\beta} - v^{\delta}_{\alpha,\beta},$$

where $x^{\delta}_{\alpha} = (\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta$ is the minimizer of the functional (3).

It is known [20] that

$$\left\|x^\dagger - x^{\delta}_{\alpha}\right\| \leqslant c\left(\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}\right).$$

Moreover, using spectral calculus and (10) we have

$$\begin{aligned}\left\|(\alpha\mathbb{I} + A^*A)^{-1}A^*Av^{\delta}_{\alpha,\beta} - v^{\delta}_{\alpha,\beta}\right\| &\leqslant \alpha\|(\alpha\mathbb{I} + A^*A)^{-1}\|\left\|v^{\delta}_{\alpha,\beta}\right\| \leqslant \left\|v^{\delta}_{\alpha,\beta}\right\| \\ &\leqslant \alpha\|(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}\|\|(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta\|.\end{aligned}$$

In addition, for $\beta > 1$ and $0 < \alpha < \frac{1}{2\|B^{-2}\|}$, the following bound holds true:

$$\begin{aligned}\|(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}\| &\leqslant \frac{\|B^{-2}\|}{\beta - \|B^{-2}\|\|\alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1}\|} \\ &\leqslant \frac{\|B^{-2}\|}{\beta - \alpha\|B^{-2}\|} < 2\|B^{-2}\|.\end{aligned}$$

Then, from (8) and (10) it follows that

$$\begin{aligned}\left\|v^{\delta}_{\alpha,\beta}\right\| &\leqslant 2\alpha\|B^{-2}\|\left(\|(\alpha\mathbb{I} + A^*A)^{-1}A^*Ax^\dagger\| + \|(\alpha\mathbb{I} + A^*A)^{-1}A^*\xi\|\right) \\ &\leqslant 2\alpha\|B^{-2}\|\left(\|x^\dagger\| + \frac{\delta}{\sqrt{\alpha}}\right) \leqslant c\left(\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}\right).\end{aligned}$$

Summing up, we finally arrive at (11). □

2.2. Error bound under satisfied link condition

It is well known [18] that in the TP regularization, saturation occurs when in (8) the index function $\varphi(t)$ tends to 0 faster than t. In this case, one can try to postpone saturation by using penalization in terms of an operator B meeting the link condition (5).

In this subsection, we will assume that (5) is satisfied with $s > \frac{1}{2}$ and, moreover, $Bx^\dagger$ is well defined as an element of X. To illustrate that this assumption is not so restrictive in the present context, we consider the situation when A and B^{-1} have a common orthonormal system $\{v_n\}$ in their singular-value decomposition, i.e.

$$A = \sum_k a_k \langle v_k, \cdot \rangle u_k \qquad B^{-1} = \sum_k b_k \langle v_k, \cdot \rangle v_k, \tag{13}$$

where $\{u_k\}$ is some other complete orthonormal system and $\{a_k\}$, $\{b_k\}$ denote sets of eigenvalues of the self-adjoint operators $(A^*A)^{1/2}$ and B^{-1} correspondingly. Then, in view of (5) we have

$$a_k \asymp b_k^s, \quad k = 1, 2, \ldots . \tag{14}$$

From the source condition (8), it follows that in this situation the element

$$Bx^\dagger = \sum_k b_k^{-1} \varphi(a_k^2) \langle v_k, g \rangle v_k$$

is well defined in X since $\varphi(t)$ is assumed to go to zero faster than t, so that $b_k^{-1}\varphi(a_k^2) \asymp a_k^{-1/s}\varphi(a_k^2)$ is bounded for $s > \frac{1}{2}$.

In this subsection, we assume that $\alpha > 1$ and introduce a linear compact operator

$$C_\alpha = (\alpha \mathbb{I} + AA^*)^{-1/2} A B^{-1}.$$

From [19], it follows that for $Bx^\dagger \in X$ one can find an index function ψ and $g_\alpha \in X$ such that

$$Bx^\dagger = \psi(C_\alpha^* C_\alpha) g_\alpha.$$

In the following, we will rely on the following assumption.

Assumption 1. *Let $\mathcal{A}$ be a sufficiently large number and $\alpha \in (1, \mathcal{A})$. Assume that there exist a positive constant R and an index function ψ meeting Δ_2-condition such that*

$$Bx^\dagger = \psi(C_\alpha^* C_\alpha) g_\alpha, \qquad \|g_\alpha\| \leqslant R. \tag{15}$$

The essence of assumption 1 is that in (15), the functions ψ and R are independent of α. To illustrate that this assumption is really not restrictive, we again consider the operators (13). The result of [19] ensures that for $Bx^\dagger \in X$ there are $g \in X$ and an index function ψ such that

$$Bx^\dagger = \psi(B^{-1}A^*AB^{-1}) g = \sum_k \psi\left(a_k^2 b_k^2\right) \langle v_k, g \rangle v_k.$$

Without loss of generality, we may assume that ψ meets Δ_2-condition. Then, as was observed, for example in [20], for $\gamma \in (0, 1)$ one can find a constant c depending only on ψ and γ such that $c\psi(t) \leqslant \psi(\gamma t)$. Therefore, for $\alpha \in (1, \mathcal{A})$ we have

$$c\psi\left(a_k^2 b_k^2\right) \leqslant \psi\left(\frac{a_k^2 b_k^2}{\mathcal{A} + \|A\|^2}\right) \leqslant \psi\left(\frac{a_k^2 b_k^2}{\alpha + a_k^2}\right) \leqslant \psi\left(a_k^2 b_k^2\right).$$

Consider now

$$g_\alpha = \sum_k \frac{\psi\left(a_k^2 b_k^2\right)}{\psi\left(\frac{a_k^2 b_k^2}{\alpha + a_k^2}\right)} \langle v_k, g \rangle v_k.$$

It is clear that $\|g_\alpha\| \leqslant \|g\|/c$ and

$$Bx^\dagger = \sum_k \psi\left(\frac{a_k^2 b_k^2}{\alpha + a_k^2}\right)\langle v_k, g_\alpha\rangle v_k = \psi(C_\alpha^* C_\alpha) g_\alpha$$

that gives us (15) with α-independent ψ and R. As a by product, we have

$$\mathrm{Range}(\psi(C_\alpha^* C_\alpha)) = \mathrm{Range}(\psi(B^{-1}A^*AB^{-1})) = \mathrm{Range}(\psi(B^{-(2s+2)})). \tag{16}$$

Theorem 2. *Let the link condition (5) and assumption 1 be satisfied. Assume also that $\frac{\sqrt{t}}{\psi(t)}$ is non-decreasing. Then, for $\alpha \in (1, \mathcal{A})$ and sufficiently small β we have*

$$\left\|x^\dagger - \left(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta\right)\right\| \leqslant c\left(\beta^{\frac{1}{2(1+s)}}\psi(\beta) + \delta\beta^{-\frac{s}{2(1+s)}}\right). \tag{17}$$

In addition, if β_{opt} is chosen such that $\beta_{opt} = \theta_\psi^{-1}(\delta)$, where $\theta_\psi(t) = \psi(t)\sqrt{t}$, then

$$\left\|x^\dagger - \left(u_{\alpha,\beta_{\mathrm{opt}}}^\delta + v_{\alpha,\beta_{\mathrm{opt}}}^\delta\right)\right\| \leqslant c\psi\left(\theta_\psi^{-1}(\delta)\right)\left(\theta_\psi^{-1}(\delta)\right)^{\frac{1}{2(1+s)}}. \tag{18}$$

Proof. Keeping in mind that $y_\delta = Ax^\dagger + \xi$ and $\alpha > 1$, we can deduce from (9) that

$$\begin{aligned}\|u_{\alpha,\beta}^\delta\| &\leqslant \|(\alpha\mathbb{I} + A^*A)^{-1}A^*A\left(x^\dagger - v_{\alpha,\beta}^\delta\right)\| + \|(\alpha\mathbb{I} + A^*A)^{-1}A^*\xi\| \\ &\leqslant \|x^\dagger - v_{\alpha,\beta}^\delta\| + \frac{\delta}{2\sqrt{\alpha}} \leqslant \|x^\dagger - v_{\alpha,\beta}^\delta\| + \frac{\delta}{2}.\end{aligned}$$

Moreover, by the definition of the operator C_α, we can rewrite (10) as follows:

$$\begin{aligned}v_{\alpha,\beta}^\delta &= \alpha B^{-1}(\beta\mathbb{I} + \alpha B^{-1}A^*(\alpha\mathbb{I} + AA^*)^{-1}AB^{-1})^{-1}B^{-1}(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta \\ &= \alpha B^{-1}(\beta\mathbb{I} + \alpha C_\alpha^* C_\alpha)^{-1}B^{-1}A^*(\alpha\mathbb{I} + AA^*)^{-1}y_\delta \\ &= \alpha B^{-1}(\beta\mathbb{I} + \alpha C_\alpha^* C_\alpha)^{-1}C_\alpha^*(\alpha\mathbb{I} + AA^*)^{-1/2}(Ax^\dagger + \xi).\end{aligned} \tag{19}$$

Note that

$$\left\|x^\dagger - v_{\alpha,\beta}^\delta\right\| \leqslant \left\|x^\dagger - v_{\alpha,\beta}^0\right\| + \left\|v_{\alpha,\beta}^0 - v_{\alpha,\beta}^\delta\right\|, \tag{20}$$

where $v_{\alpha,\beta}^0$ is given by (19) with $\xi = 0$.

Now, we are going to use the well-known interpolation inequality [16] of the form

$$\|x\| \leqslant \|B^{-s}x\|^{\frac{1}{1+s}}\|Bx\|^{\frac{s}{1+s}}, \tag{21}$$

which is valid for $s > 0$ and $x \in \mathrm{Range}(B^{-1})$.

We will also use the fact (see, for example, [20]) that if for $t \in [0, d]$, $d > 0$, a function $\varphi(t)$ is continuous, increasing and such that $\varphi(0) = 0$, but $\frac{t}{\varphi(t)}$ is non-decreasing, then for any $\lambda \in [0, 1]$ holds

$$\sup_{t\in[0,d]}\left|\frac{\lambda}{\lambda + t}\varphi(t)\right| \leqslant \varphi(\lambda). \tag{22}$$

Then, under the conditions of the theorem, from (19), (22) it follows that

$$\begin{aligned}\left\|B\left(x^\dagger - v_{\alpha,\beta}^0\right)\right\| &= \left\|Bx^\dagger - \alpha(\beta\mathbb{I} + \alpha C_\alpha^* C_\alpha)^{-1}C_\alpha^* C_\alpha Bx^\dagger\right\| \\ &= \left\|\left(\mathbb{I} - \left(\frac{\beta}{\alpha}\mathbb{I} + C_\alpha^* C_\alpha\right)^{-1} C_\alpha^* C_\alpha\right)Bx^\dagger\right\| \\ &\leqslant R\left\|\left(\mathbb{I} - \left(\frac{\beta}{\alpha}\mathbb{I} + C_\alpha^* C_\alpha\right)^{-1} C_\alpha^* C_\alpha\right)\psi(C_\alpha^* C_\alpha)\right\| \\ &\leqslant R\sup_t\left|\frac{\frac{\beta}{\alpha}}{\frac{\beta}{\alpha} + t}\psi(t)\right| \leqslant c\psi\left(\frac{\beta}{\alpha}\right).\end{aligned}$$

Moreover, using the link conditions (5) and (22) we can continue as follows:

$$\begin{aligned}\left\|B^{-s}\left(x^{\dagger}-v^{0}_{\alpha,\beta}\right)\right\| &\leqslant \left\|AB^{-1}\left(Bx^{\dagger}-Bv^{0}_{\alpha,\beta}\right)\right\| \\ &= \left\|(\alpha\mathbb{I}+AA^{*})^{1/2}C_{\alpha}\left(\mathbb{I}-\left(\frac{\beta}{\alpha}\mathbb{I}+C^{*}_{\alpha}C_{\alpha}\right)^{-1}C^{*}_{\alpha}C_{\alpha}\right)Bx^{\dagger}\right\| \\ &\leqslant R(\alpha+\|A\|^{2})^{1/2}\left\|\left(\mathbb{I}-\left(\frac{\beta}{\alpha}\mathbb{I}+C_{\alpha}C^{*}_{\alpha}\right)^{-1}C_{\alpha}C^{*}_{\alpha}\right)\psi(C_{\alpha}C^{*}_{\alpha})(C_{\alpha}C^{*}_{\alpha})^{1/2}\right\| \\ &\leqslant R(\alpha+\|A\|^{2})^{1/2}\sup_{t}\left|\frac{\frac{\beta}{\alpha}}{\frac{\beta}{\alpha}+t}\psi(t)\sqrt{t}\right| \leqslant c(\alpha+\|A\|^{2})^{1/2}\psi\left(\frac{\beta}{\alpha}\right)\sqrt{\frac{\beta}{\alpha}},\end{aligned}$$

where we use (22) with $\varphi(t)=\psi(t)\sqrt{t}$.

Thus, we arrive at the bound

$$\left\|B^{-s}\left(x^{\dagger}-v^{0}_{\alpha,\beta}\right)\right\| \leqslant c\sqrt{\beta}\psi\left(\frac{\beta}{\alpha}\right).$$

Applying the same argument to

$$v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}=B^{-1}\left(\frac{\beta}{\alpha}\mathbb{I}+C^{*}_{\alpha}C_{\alpha}\right)^{-1}C^{*}_{\alpha}(\alpha\mathbb{I}+AA^{*})^{-1/2}\xi,$$

we also have

$$\left\|B\left(v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right)\right\| \leqslant \left\|\left(\frac{\beta}{\alpha}\mathbb{I}+C^{*}_{\alpha}C_{\alpha}\right)^{-1}(C^{*}_{\alpha}C_{\alpha})^{1/2}\right\|\,\|(\alpha\mathbb{I}+AA^{*})^{-1/2}\xi\| \leqslant c\frac{\delta}{\sqrt{\beta}},$$

$$\left\|B^{-s}\left(v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right)\right\| \leqslant \delta\left\|(\alpha\mathbb{I}+AA^{*})^{1/2}\left(\frac{\beta}{\alpha}\mathbb{I}+C_{\alpha}C^{*}_{\alpha}\right)^{-1}C_{\alpha}C^{*}_{\alpha}\right\| \leqslant c\delta.$$

Then, the interpolation inequality (21) gives us

$$\begin{aligned}\left\|x^{\dagger}-v^{0}_{\alpha,\beta}\right\| &\leqslant \left\|B^{-s}\left(x^{\dagger}-v^{0}_{\alpha,\beta}\right)\right\|^{\frac{1}{1+s}}\left\|B\left(x^{\dagger}-v^{0}_{\alpha,\beta}\right)\right\|^{\frac{s}{1+s}} \leqslant c\beta^{\frac{1}{2(1+s)}}\psi\left(\frac{\beta}{\alpha}\right), \\ \left\|v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right\| &\leqslant \left\|B^{-s}\left(v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right)\right\|^{\frac{1}{1+s}}\left\|B\left(v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right)\right\|^{\frac{s}{1+s}} \leqslant c\delta\beta^{-\frac{s}{2(1+s)}}.\end{aligned}$$

Combining the above estimations in (20) and recalling that $\alpha > 1$, we finally arrive at

$$\begin{aligned}\left\|x^{\dagger}-v^{\delta}_{\alpha,\beta}\right\| &\leqslant \left\|x^{\dagger}-v^{0}_{\alpha,\beta}\right\|+\left\|v^{0}_{\alpha,\beta}-v^{\delta}_{\alpha,\beta}\right\| \leqslant c\left(\beta^{\frac{1}{2(1+s)}}\psi\left(\frac{\beta}{\alpha}\right)+\delta\beta^{-\frac{s}{2(1+s)}}\right) \\ &\leqslant c\left(\beta^{\frac{1}{2(1+s)}}\psi(\beta)+\delta\beta^{-\frac{s}{2(1+s)}}\right)\end{aligned}$$

that leads to (18) for $\beta=\beta_{\mathrm{opt}}$. □

Remark 2. Note that under the condition of theorem 2, the order of the error bound (18) cannot be improved in general. For example, for $\psi(t)=t^{p}$ we have

$$\psi\left(\theta^{-1}_{\psi}(\delta)\right)\left(\theta^{-1}_{\psi}(\delta)\right)^{\frac{1}{2(s+1)}}=\delta^{\frac{2p(s+1)+1}{(2p+1)(s+1)}}. \tag{23}$$

At the same time, in view of (16) for the operators (13) and (14), assumption 1 means that

$$x^{\dagger}\in\mathrm{Range}(B^{-(2p(s+1)+1)}). \tag{24}$$

On the other hand, it is known [23] that under the link condition (5), the solution $x^{\dagger}\in\mathrm{Range}(B^{-\mu})$ cannot be in general reconstructed in X from noisy data y_{δ} with the order of accuracy better than $O\left(\delta^{\frac{\mu}{\mu+s}}\right)$, which for $\mu=2p(s+1)+1$ coincides with the bound (23) given in the considered case by theorem 2.

3. Numerical examples

In this section, we present numerical experiments that illustrate the compensatory property of the considered MP regularization (6). Recall that by this we mean that the method performs similar to the best single-penalty regularization (3) or (4). At this point, it is also worthwhile to mention that, in accordance with the analysis presented in the previous section, the method (9), (10) exhibits the compensatory property when one of the regularization parameters is greater than 1, independently of a noise level δ. Therefore, to demonstrate the above-mentioned feature, we will employ the so-called quasi-optimality criterion, which does not require any knowledge of the noise level. This heuristic approach was originally proposed in [30] and has been recently advocated in [14].

3.1. Quasi-optimality criterion

Recall that in the case of the method (3), the quasi-optimality criterion chooses a regularization parameter $\alpha = \alpha_l$ from a set

$$Q_N^{\alpha} = \{\alpha = \alpha_i = \alpha_0 q^i, i = 0, 1, 2, \ldots, N\}, \quad q > 1, \tag{25}$$

such that

$$\left\|x_{\alpha_l}^{\delta} - x_{\alpha_{l-1}}^{\delta}\right\| = \min\left\{\left\|x_{\alpha_i}^{\delta} - x_{\alpha_{i-1}}^{\delta}\right\|, i = 1, 2, \ldots, N\right\}.$$

In the similar way, one can apply the quasi-optimality criterion to a set of parameters

$$P_M^{\beta} = \{\beta = \beta_j = \beta_0 p^j, j = 0, 1, 2, \ldots, M\}, \qquad p > 1, \tag{26}$$

and choose $\beta = \beta_k \in P_M^{\beta}$ such that

$$\left\|x_{\beta_k,B}^{\delta} - x_{\beta_{k-1},B}^{\delta}\right\| = \min\left\{\left\|x_{\beta_j,B}^{\delta} - x_{\beta_{j-1},B}^{\delta}\right\|, j = 1, 2, \ldots, M\right\}.$$

Note that in general, the quasi-optimality criterion can be used for choosing regularization parameters in various regularization methods [1, 8]. The idea behind this heuristic approach is based on the belief that the minimum of the distance between regularized approximate solutions corresponding to two successive values of a regularization parameter will be attained near the 'cross-over point' where an approximation error and a propagated data error have about the same order of magnitude (see, e.g., [8, p 125]). For the MP regularization, the quasi-optimality criterion can be implemented as follows. First, for every $\beta = \beta_j \in P_M^{\beta}$ we choose $\alpha = \alpha_l = \alpha(\beta_j)$ from the set (25) such that

$$\left\|x_{\alpha_l,\beta_j}^{\delta} - x_{\alpha_{l-1},\beta_j}^{\delta}\right\| = \min\left\{\left\|x_{\alpha_i,\beta_j}^{\delta} - x_{\alpha_{i-1},\beta_j}^{\delta}\right\|, i = 1, 2, \ldots, N\right\},$$

where here and below $x_{\alpha,\beta}^{\delta} = u_{\alpha,\beta}^{\delta} + v_{\alpha,\beta}^{\delta}$.

Next, we apply the quasi-optimality criterion to the sequence $\{x_{\alpha(\beta_j),\beta_j}^{\delta}\}$ parametrized by $\beta_j \in P_M^{\beta}$. More specifically, we select $\beta_k \in P_M^{\beta}$ such that

$$\left\|x_{\alpha(\beta_k),\beta_k}^{\delta} - x_{\alpha(\beta_{k-1}),\beta_{k-1}}^{\delta}\right\| = \min\left\{\left\|x_{\alpha(\beta_j),\beta_j}^{\delta} - x_{\alpha(\beta_{j-1}),\beta_{j-1}}^{\delta}\right\|, j = 1, 2, \ldots, M\right\}.$$

Then, a regularized approximate solution $x_{\alpha,\beta}^{\delta}$ of our choice is defined by (9) and (10) with $\alpha = \alpha(\beta_k)$ and $\beta = \beta_k$.

3.2. Numerical illustrations and comparison: operators with known singular-value expansion

Similar to [1] in our first numerical experiment, we consider compact operators A and B^{-1} that are related as in (13). Note that the knowledge of the singular-value expansion of the operators allows us to verify easily whether the link condition (5) is violated or not. In the first experiment,

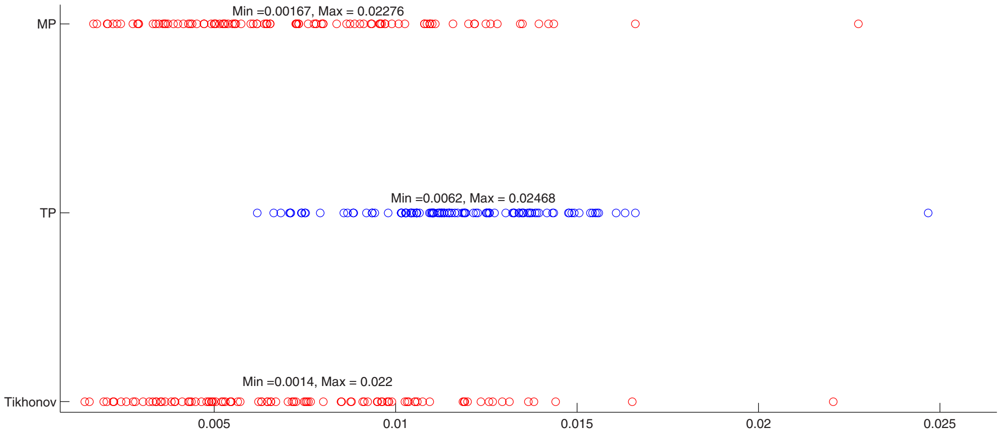

Figure 1. Numerical illustration (first experiment). The figure presents REs (circles) for 100 simulations of y_δ with 1% noise.

the operators A and B^{-1} are given as diagonal matrices of size n. The matrix corresponding to the operator A has diagonal elements $a_k = k^{-r}, k = 1, 2, \ldots, n, n = 50, r = 3$. Further, we assume that the source condition (8) is satisfied with $\varphi(t) = t^p$, $p = 2$, and the solution $x^\dagger$ is given in the form of the n-dimensional vector

$$x^\dagger = (A^*A)^2 g, \tag{27}$$

where g is a random vector whose components are uniformly distributed on [0, 1] and such that $\|g\| = 10$; here and below $\|\cdot\|$ means the standard norm in the n-dimensional Euclidean space $\mathbb{R}^n$. Then, the exact right-hand side is produced as $y = Ax^\dagger$.

Noisy data y_δ are simulated in the form $y_\delta = y + \xi$, where $\xi = \delta \frac{\epsilon}{\|\epsilon\|}$ and ϵ is another random vector with uniformly distributed components. Both vectors g and ϵ are generated 100 times, so that we have 100 problems of the form (1) with noisy data y_δ, and the noise level δ is given as $\delta = 0.01\|Ax^\dagger\|$ that corresponds to 1% data noise.

In accordance with the theory, under the source condition (27), the TP regularization (3) can suffer from saturation. On the other hand, this effect may be relaxed by using the Tikhonov regularization (4) with a proper choice of a regularization operator B for which condition (5) is satisfied. First, we choose the self-adjoint operator B such that the corresponding diagonal matrix has the elements $b_{kk} = b_k = k, k = 1, 2, \ldots, n$. For the considered A, the chosen operator B satisfies (5) with $s = 3$. In the experiment, we use the quasi-optimality criterion as the parameter choice rule with $\alpha_0 = \beta_0 = 10^{-4}$, $q = p = 1.25$ and $N = M = 45$, in the way described above.

To assess the obtained results and compare the performance of the considered regularization schemes, we measure the relative error (RE)

$$\frac{\|x - x^\dagger\|}{\|x^\dagger\|}$$

for $x = x^\delta_{\alpha,\beta}$, $x = x^\delta_\alpha$ and $x = x^\delta_{\beta,B}$.

The results are displayed in figure 1, where each circle exhibits a RE in solving the problems with one of 100 simulated data, for each of three regularization methods: the MP regularization, the TP regularization and the Tikhonov regularization (Tikhonov). Note that such a form of graphical illustration of the comparative performance of different regularization algorithms is rather common (see, e.g., [11]). Moreover, in table 1 the statistical measures

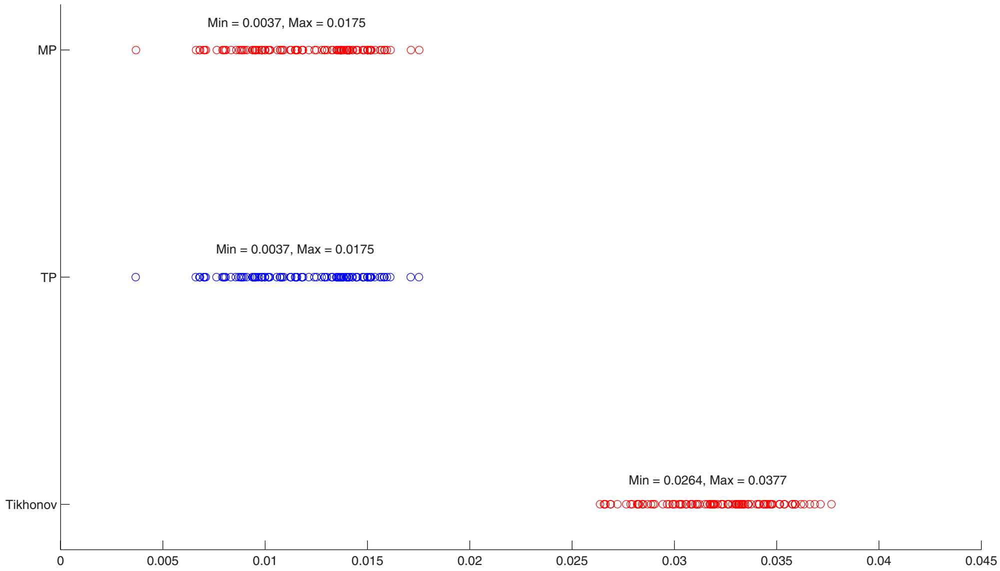

Figure 2. Numerical illustration (second experiment). The figure presents REs (circles) for 100 simulations of y_δ with 1% noise.

Table 1. Numerical illustration (first experiment). Statistical performance measures for the regularized approximations $x^\delta_{\alpha,\beta}, x^\delta_\alpha, x^\delta_{\beta,B}$ and 100 simulations of y_δ with 1% noise.

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^\delta_{\alpha,\beta}$	0.0071	0.0063	0.0036	$\alpha = 6.02, \beta = 0.002$
x^δ_α	0.0117	0.0117	0.0027	0.007
$x^\delta_{\beta,B}$	0.0072	0.0061	0.0035	0.002

Table 2. Numerical illustration (second experiment). Statistical performance measures for the regularized approximations $x^\delta_{\alpha,\beta}, x^\delta_\alpha, x^\delta_{\beta,B}$ and 100 simulations of y_δ with 1% noise.

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^\delta_{\alpha,\beta}$	0.0118	0.0119	0.0029	$\alpha = 0.0066, \beta = 12.1$
x^δ_α	0.0118	0.0119	0.0028	0.0066
$x^\delta_{\beta,B}$	0.0319	0.0322	0.0028	0.0247

such as mean values, median values, standard deviation of the RE as well as mean values of the regularization parameters are given for each of the methods.

The numerical results confirm the theoretical conclusion that the saturation of the method (3) can be potentially relaxed by the use of method (4). Moreover, in the considered case the MP method (9), (10) performs similar to method (4), as predicted by theorem 2.

On the other hand, if we consider the operator B, corresponding to the diagonal matrix with elements

$$b_k = \begin{cases} k, & k = 1, 3, \ldots, 2j-1, \\ 1/k, & k = 2, 4, \ldots, 2j,\ j = n/2, \end{cases}$$

then, from figure 2 and table 2 we can see that the saturation cannot be relaxed by the Tikhonov method (4) due to the fact that for the considered B the link condition (5) is violated

($\|B^{-s}\| \geqslant (n-1)^s \geqslant 1 \geqslant \|A\|$). At the same time, in both cases we can observe that the MP regularization (6) with a proper choice of the regularization parameters demonstrates the performance at the level of the best single-penalty regularization.

3.3. Numerical illustrations and comparison: first kind Fredholm integral equations

In this subsection, we are going to show that the proposed method exhibits similar performance in a more general case, when the singular-value expansion of the operators is not known. Similar to [17], we generate the test problems of the form (1) by using the functions *shaw*(n) and *ilaplace*$(n, 1)$ from the Matlab regularization toolbox [10]. These functions occur as the results of a discretization of the first kind Fredholm integral equation of the form

$$\int_a^b k(s,t)f(t)\,\mathrm{d}t = g(s), \qquad s \in [a,b], \tag{28}$$

with a known solution $f(t)$. As in two previous experiments, the operator A and the solution $x^\dagger$ are given as an $n \times n$-matrix and n-dimensional vector, respectively. The noisy data y_δ are simulated 100 times in the same way as above, i.e. $y_\delta = Ax^\dagger + \xi$ with the noise level δ corresponding to 1% of data noise.

Moreover, the penalizing operator is given as an $n \times n$-matrix and defined as $B = (D^*D)^{1/2}$, where

$$D = \begin{pmatrix} 1 & -1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & & & 1 & -1 \end{pmatrix}$$

is a discrete approximation of the first derivative on a regular grid with n points.

We perform an experiment with the function *shaw*(n) that is a discretization of equation (28) with $a = -\pi/2$ and $b = \pi/2$. The kernel and the solution are given as

$$k(s,t) = (\cos(s) + \cos(t))^2 \left(\frac{\sin(u)}{u}\right)^2, \qquad u = \pi(\sin(s) + \sin(t)),$$
$$f(t) \quad = 2\,\mathrm{e}^{-6(t-0.8)^2} + \mathrm{e}^{-2(t+0.5)^2}.$$

The corresponding equation (28) is discretized by a simple quadrature with n equidistant points. Similar to [17], we take $n = 100$. This time the quasi-optimality criterion is implemented with $\alpha_0 = 0.0001$, $\beta_0 = 0.0005$, $q = 1.1$ and $p = 1.3$, respectively.

The results are displayed in figure 3 and table 3.

In the last experiment, we consider the function *ilaplace*$(n, 1)$ which occurs in a discretization of the inverse Laplace transformation by means of the Gauss–Laguerre quadrature with n knots and corresponds to equation (28) with $a = 0, b = \infty, k(s,t) = \mathrm{e}^{-st}, f(t) = \mathrm{e}^{-t/2}, g(s) = (s + 1/2)^{-1}$. We choose $n = 100$ and test the regularization methods in combination with the quasi-optimality criterion, which is implemented with $\alpha_0 = 0.0002$, $\beta_0 = 0.0001$ and $q = 1.25$, $p = 1.3$.

In figure 4, we show the REs produced by three regularization methods. Moreover, table 4 presents statistical information about the performance of the methods.

Again, the compensatory property of the MP regularization is observed even when it is not known *a priori* whether or not the link condition (5) is satisfied.

Note that the mean values of the regularization parameters α and β indicated in tables 3 and 4 can be used in constructing approximations $x^\delta_{\alpha,\beta}, x^\delta_\alpha, x^\delta_{\beta,B}$ to *shaw*(100) and

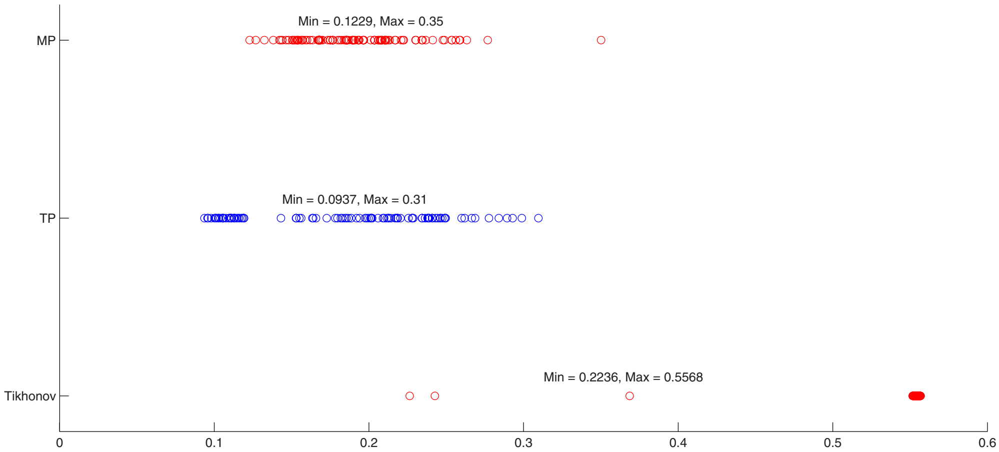

Figure 3. Numerical illustration for the function *shaw*(100). The figure presents REs (circles) for 100 simulations of y_δ with 1% noise.

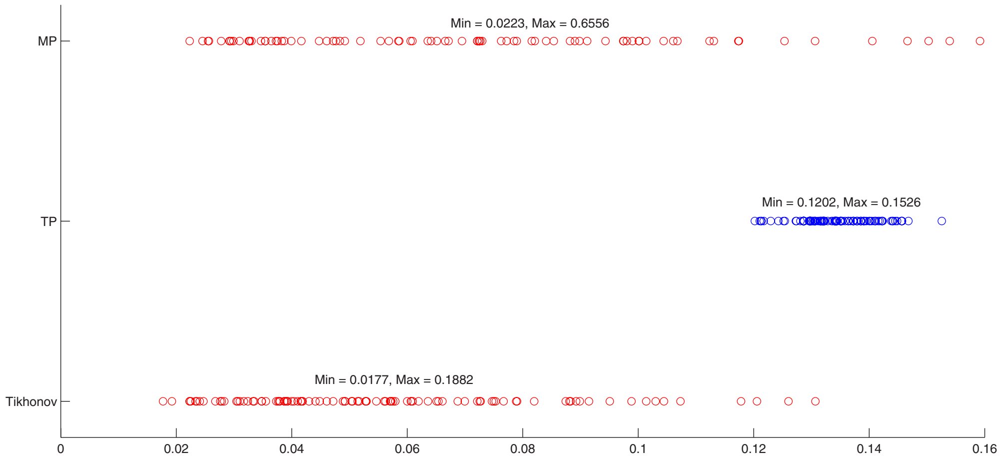

Figure 4. Numerical illustration for the function *ilaplace*(100, 1). The figure presents REs (circles) for 100 simulations of y_δ with 1% noise.

Table 3. Numerical illustration for the function *shaw*(100). Statistical performance measures for the regularized approximations $x^\delta_{\alpha,\beta}, x^\delta_\alpha, x^\delta_{\beta,B}$ and 100 simulations of y_δ with 1% noise.

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^\delta_{\alpha,\beta}$	0.1919	0.188	0.0374	$\alpha = 0.00011, \beta = 13.89$
x^δ_α	0.1843	0.1957	0.0605	0.0014
$x^\delta_{\beta,B}$	0.5458	0.5538	0.0484	0.0019

ilaplace(100, 1) for any data y_δ with the same noise intensity δ as in the above simulations. By the way, such a choice of the regularization parameters is in the spirit of the approach [6]. In figures 5 and 6, we display the shapes of the solutions to *shaw*(100) (figure 5) and *ilaplace*(100, 1) (figure 6), as well as their regularized approximations $x^\delta_{\alpha,\beta}, x^\delta_\alpha, x^\delta_{\beta,B}$ corresponding to the mean values of α and β from tables 3 and 4. These approximations have

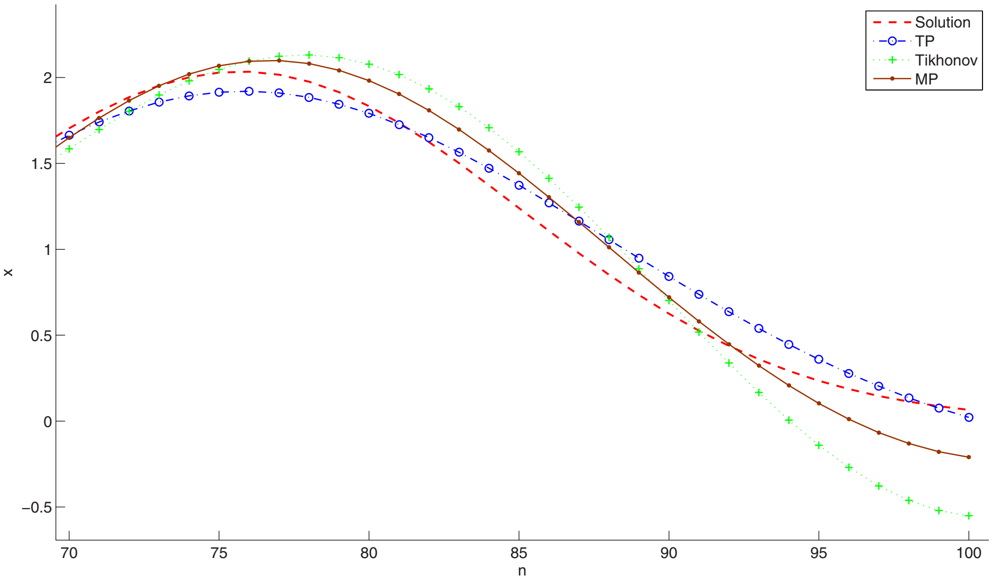

Figure 5. The representative fragments of the solution to *shaw*(100) and of its approximations given by the TP method, the Tikhonov method and the MP regularization corresponding to the mean values of the regularization parameters.

Table 4. Numerical illustration for the function *ilaplace*(100, 1). Statistical performance measures for the regularized approximations $x^{\delta}_{\alpha,\beta}$, x^{δ}_{α}, $x^{\delta}_{\beta,B}$ and 100 simulations of y_{δ} with 1% noise.

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^{\delta}_{\alpha,\beta}$	0.1068	0.0767	0.103	$\alpha = 1.2037$, $\beta = 0.0066$
x^{δ}_{α}	0.1342	0.134	0.006	0.0014
$x^{\delta}_{\beta,B}$	0.0575	0.0515	0.029	0.1138

been constructed for new sets of simulated data y_{δ}. As can be seen in the figures, the MP regularization again performs similar to the best single-penalty regularization.

The presented MP regularization equipped with the quasi-optimality criterion can be used for a more flexible numerical treatment of an ill-posed problem, when, on the one hand, an additional penalizing operator B is used to relax the saturation effect, and, on the other hand, a link condition (5) is not granted *a priori*, as is the case for noisy operators A, for example. Moreover, the considered MP regularization may also be relevant in the situation when some parts of data are more accurately known than others. Such a situation occurs, for example, in geomathematics [9], and deserves further study.

Remark 3. It is clear that the quasi-optimality criterion is only one possible parameter choice rule and it might not guarantee the optimal choice of the parameters. Indeed, in the experiments it was observed that a proper choice of the sets Q^{α}_{N} and P^{β}_{M} is crucial for obtaining good performance of the methods. We believe that a deeper study of this issue is important. Future possible work in this direction is to consider a choice of the sets (25) and (26) using meta-learning [4, 24], which proved to be an efficient method for dealing with problems of similar type.

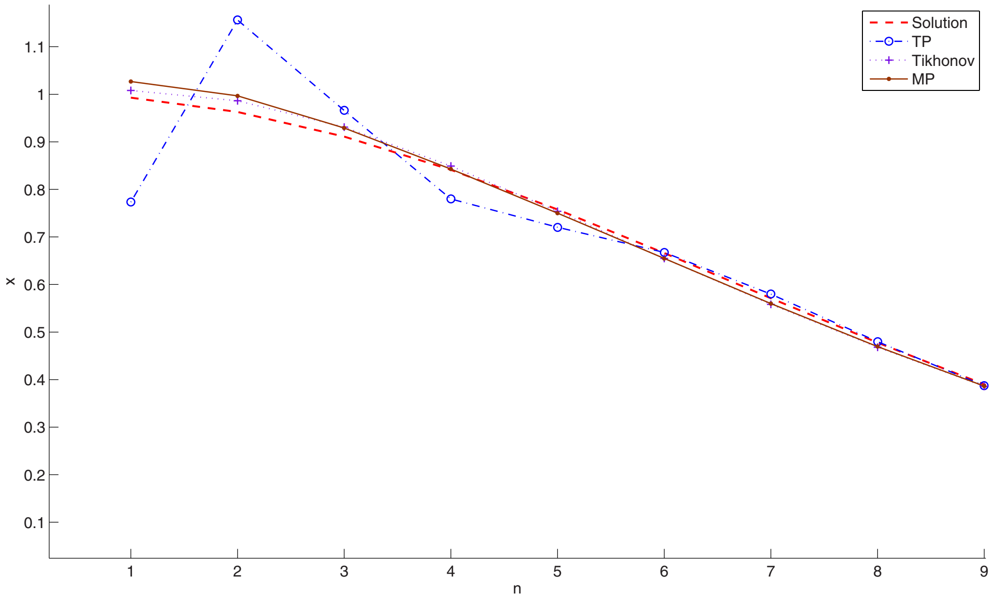

Figure 6. The representative fragments of the solution to *ilaplace*(100, 1) and of its approximations given by the TP method, the Tikhonov method and the MP regularization corresponding to the mean values of the regularization parameters.

Acknowledgments

The authors are supported by the Austrian Fonds Zur Förderung der Wissenschaftlichen Forschung (FWF), grant P25424. Moreover, the authors would like to acknowledge the stimulating comments of the anonymous reviewers.

References

[1] Bauer F and Lukas M 2011 Comparing parameter choice methods for regularization of ill-posed problems *Math. Comput. Simul.* **81** 1795–841
[2] Belge M, Kilmer M and Miller E 2002 Efficient determination of multiple regularization parameters in a generalized L-curve framework *Inverse Problems* **18** 1161–83
[3] Böttcher A, Hofmann B, Tautenhahn U and Yamamoto M 2006 Convergence rates for Tikhonov regularization from different kinds of smoothness conditions *Appl. Anal.* **85** 555–78
[4] Brazdil P, Giraud-Carrier C, Soares C and Vilalta R 2009 *Metalearning: Applications to Data Mining* (Berlin: Springer)
[5] Brezinski C, Redivo-Zaglia M, Rodriguez G and Seatzu S 2003 Multiparameter regularization techniques for ill-conditioned linear systems *Numer. Math.* **94** 203–28
[6] Calvetti C and Somersalo E 2005 Priorconditioners for linear systems *Inverse Problems* **21** 1397–418
[7] Chen Z, Lu Y, Xu Y and Yang H 2008 Multi-parameter Tikhonov regularization for linear ill-posed operator equations *J. Comput. Math.* **26** 37–55
[8] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems (Mathematics and Its Applications* vol 375*)* (Dordrecht: Kluwer)
[9] Freeden W 1999 *Multiscale Modeling of Spaceborne Geodata* (Leipzig: Teubner)
[10] Hansen P 1994 Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems *Numer. Algorithms* **6** 1–35
[11] Hansen P 1998 *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion* (Philadelphia, PA: SIAM)

[12] Hohage T and Pricop M 2008 Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise *Inverse Problems Imaging* **2** 271–90
[13] Ito K, Jin B and Takeuchi T 2011 Multi-parameter Tikhonov regularization *Methods Appl. Anal.* **18** 31–46
[14] Kindermann S and Neubauer A 2008 On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization *Inverse Problems Imaging* **2** 291–9
[15] Koltchinskii V and Yuan M 2010 Sparsity in multiple kernel learning *Ann. Stat.* **38** 3660–95
[16] Krein S and Petunin Y 1966 Scales of Banach spaces *Russ. Math. Surv.* **14** 85–159
[17] Lu S and Pereverzev S V 2011 Multi-parameter regularization and its numerical realization *Numer. Math.* **118** 1–31
[18] Mathé P 2004 Saturation of regularization methods for linear ill-posed problems in Hilbert spaces *SIAM J. Numer. Anal.* **42** 968–73
[19] Mathé P and Hofmann B 2008 How general are general source conditions? *Inverse Problems* **24** 015009
[20] Mathé P and Pereverzev S V 2003 Geometry of linear ill-posed problems in variable Hilbert scales *Inverse Problems* **19** 789–803
[21] Meyer Y 2002 *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations (AMS University Lecture Series* vol 22*)* (Providence, RI: American Mathematical Society)
[22] Nair T, Pereverzev S V and Tautenhahn U 2005 Regularization in Hilbert scales under general smoothing conditions *Inverse Problems* **21** 1851–69
[23] Natterer F 1984 Error bounds for Tikhonov regularization in Hilbert scales *Appl. Anal.* **18** 29–37
[24] Naumova V, Pereverzyev S V and Sivananthan S 2012 A meta-learning approach to the regularized learning—case study: blood glucose prediction *Neural Netw.* **33** 181–93
[25] Neubauer A 1988 An *a posteriori* parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates *SIAM J. Numer. Anal.* **25** 1313–26
[26] Neubauer A 1992 Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales *Appl. Anal.* **46** 59–72
[27] Neubauer A 1997 On converse and saturation results for Tikhonov regularization of linear ill-posed problems *SIAM J. Numer. Anal.* **34** 517–27
[28] Tautenhahn U 1998 On a general regularization scheme for nonlinear ill-posed problems: II. Regularization in Hilbert scales *Inverse Problems* **14** 1607–16
[29] Tikhonov A N 1963 On the regularization of ill-posed problems *Dokl. Akad. Nauk SSSR* **153** 49–52
[30] Tikhonov A N and Glasko V B 1965 Use of the regularization methods in non-linear problems *USSR Comput. Math. Phys.* **5** 93–107
[31] Vese L and Osher S 2003 Modeling textures with total variation minimization and oscillating patterns in image processing *J. Sci. Comput.* **19** 553–72
[32] Vese L and Osher S 2004 Image denoising and decomposition with total variation minimization and oscillatory functions *J. Math. Imaging Vis.* **20** 7–18