

Multi-Penalty Regularization with a Component-Wise Penalization

V. Naumova, S. Pereverzyev

RICAM-Report 2013-03

Multi-Penalty Regularization with a Component-Wise Penalization

Valeriya Naumova and Sergei V Pereverzyev

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.

1. Introduction

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

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

where $A : X \rightarrow Y$ 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 sequel, 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 the equation (1).

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

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

where $\xi \in Y$, $\|\xi\| \leq \delta$, $\delta \in (0, 1)$. 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 [7].

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

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

with $\alpha > 0$ being a regularization parameter. Due to 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 Tikhonov-Phillips regularization suffers from a saturation effect [24, 16]. 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 [26] 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

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

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

$$\|B^{-s}x\| \leq \|Ax\| \leq b\|B^{-s}x\|, \quad (5)$$

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

For more details we refer to the classical paper [20], see also [23, 19, 3] and references therein.

It is clear, that the condition (5) is a serious restriction and, what is even more important, the condition is sometimes hardly verifiable, as it is the case, for instance, when Tikhonov regularization is used for solving nonlinear ill-posed equations [22, 25, 10]. For example, in [25] 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 the condition (5) is violated. To exemplify this kind of difficulties, we refer to the section with numerical experiments and specifically to Figure 2.

Thus, if the 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 Tikhonov-Phillips 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 regularization with a component-wise penalization, in which the following form of the

regularization functional is used

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

This form is inspired by the study [13] 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 the regularization theory.

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

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

Here we may refer to the papers [2, 5, 6, 11, 15]. Our present study is stimulated by the remark made in [15], where in numerical experiments the authors observed the compensatory property of the multi-penalty 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 next section. In the final section with numerical experiments we illustrate the efficiency of the proposed approach equipped with a heuristic parameter choice rule on the number of academic examples.

2. Convergence Rates for the Multi-Penalty Regularization with a Component-Wise Penalization

As it has been already mentioned in Introduction, the multi-penalty regularization could exhibit the compensatory property, at least numerically. In this section we provide theoretical justification of this property for the multi-penalty regularization (6). This will be done by analyzing separately two cases. At first, we consider the case when the link condition (5) is violated. As it follows from the paper [17], in this situation one still can rely on the so-called source condition

$$x^\dagger = \varphi(A^*A)g, \quad \|g\| \leq R, \quad (8)$$

where $\varphi : [0, \|A\|^2] \rightarrow [0, 1]$ 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 the condition (5) is satisfied.

Recall that in the case when the link condition is violated the Tikhonov-Phillips 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 $u_{\alpha,\beta}^\delta, v_{\alpha,\beta}^\delta$ 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

$$\begin{cases} A^*A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - A^*y_\delta + \alpha u_{\alpha,\beta}^\delta = 0 \\ A^*A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - A^*y_\delta + \beta B^2 v_{\alpha,\beta}^\delta = 0, \end{cases}$$

that allows the representation

$$u_{\alpha,\beta}^\delta = (\alpha\mathbb{I} + A^*A)^{-1}(A^*y_\delta - A^*Av_{\alpha,\beta}^\delta), \quad (9)$$

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

where \mathbb{I} is the identity operator.

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

$$\|x^\dagger - (u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta)\| \leq c \left(\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}} \right). \quad (11)$$

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

$$\|x^\dagger - (u_{\alpha_{opt},\beta}^\delta + v_{\alpha_{opt},\beta}^\delta)\| \leq c\varphi(\theta_\varphi^{-1}(\delta)) \quad (12)$$

is obtained.

Proof. Note that the bound (12) is a consequence of (11), and its optimality under the condition (8) is proven in [18]. Therefore, only (11) needs to be proven.

From (9) and (10) it follows that

$$x^\dagger - (u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) = x^\dagger - x_\alpha^\delta + (\alpha\mathbb{I} + A^*A)^{-1}A^*Av_{\alpha,\beta}^\delta - v_{\alpha,\beta}^\delta,$$

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

It is known [18] that

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

Moreover, using spectral calculus and (10) we have

$$\begin{aligned} \|(\alpha\mathbb{I} + A^*A)^{-1}A^*Av_{\alpha,\beta}^\delta - v_{\alpha,\beta}^\delta\| &\leq \alpha\|(\alpha\mathbb{I} + A^*A)^{-1}\| \|v_{\alpha,\beta}^\delta\| \leq \|v_{\alpha,\beta}^\delta\| \\ &\leq \alpha\|(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}\| \|(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta\|. \end{aligned}$$

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

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

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

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

Summing up, we finally arrive at (11). \square

2.2. Error Bound under Satisfied Link Condition

It is well-known [16] that in the Tikhonov-Phillips 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.

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

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

$$a_k \asymp b_k^s, \quad k = 1, 2, \dots \quad (14)$$

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

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

From [17] 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 sequel, 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

$$Bx^\dagger = \psi(C_\alpha^* C_\alpha) g_\alpha, \quad \|g_\alpha\| \leq R. \quad (15)$$

The essence of Assumption 1 is that in (15) the function 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 [17] ensures that for $Bx^\dagger \in X$ there are $g \in X$ and an index function ψ such that

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

Without loss of generality, we may assume that ψ meets Δ_2 -condition. Then for $\alpha \in (1, \mathcal{A})$ we have

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

Consider now

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

It is clear that $\|g_\alpha\| \leq \|g\|/c$ 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 by product we have

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

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

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

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

$$\|x^\dagger - (u_{\alpha,\beta_{\text{opt}}}^\delta + v_{\alpha,\beta_{\text{opt}}}^\delta)\| \leq c\psi(\theta_\psi^{-1}(\delta))(\theta_\psi^{-1}(\delta))^{\frac{1}{2(1+s)}}. \quad (18)$$

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

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

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

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

Note that

$$\|x^\dagger - v_{\alpha,\beta}^\delta\| \leq \|x^\dagger - v_{\alpha,\beta}^0\| + \|v_{\alpha,\beta}^0 - v_{\alpha,\beta}^\delta\|, \quad (20)$$

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

Now we are going to use the well-known interpolation inequality [14] of the form

$$\|x\| \leq \|B^{-s} x\|^{\frac{1}{1+s}} \|Bx\|^{\frac{s}{1+s}}, \quad (21)$$

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

We will also use the fact (see, for example, [18]) 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

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

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

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

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

$$\begin{aligned} \|B^{-s}(x^\dagger - v_{\alpha,\beta}^0)\| &\leq \|AB^{-1}(Bx^\dagger - Bv_{\alpha,\beta}^0)\| \\ &= \left\| (\alpha \mathbb{I} + A A^*)^{1/2} C_\alpha \left(\mathbb{I} - \left(\frac{\beta}{\alpha} \mathbb{I} + C_\alpha^* C_\alpha \right)^{-1} C_\alpha^* C_\alpha \right) Bx^\dagger \right\| \\ &\leq R(\alpha + \|A\|^2)^{1/2} \left\| \left(\mathbb{I} - \left(\frac{\beta}{\alpha} \mathbb{I} + C_\alpha C_\alpha^* \right)^{-1} C_\alpha C_\alpha^* \right) \psi(C_\alpha C_\alpha^*) (C_\alpha C_\alpha^*)^{1/2} \right\| \\ &\leq R(\alpha + \|A\|^2)^{1/2} \sup_t \left| \frac{\frac{\beta}{\alpha}}{\frac{\beta}{\alpha} + t} \psi(t) \sqrt{t} \right| \leq c(\alpha + \|A\|^2)^{1/2} \psi \left(\frac{\beta}{\alpha} \right) \sqrt{\frac{\beta}{\alpha}}, \end{aligned}$$

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

Thus, we arrive at the bound

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

Applying the same argument to

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

we also have

$$\begin{aligned} \|B(v_{\alpha,\beta}^0 - v_{\alpha,\beta}^\delta)\| &\leq \left\| \left(\frac{\beta}{\alpha} \mathbb{I} + C_\alpha^* C_\alpha \right)^{-1} (C_\alpha^* C_\alpha)^{1/2} \right\| \|(\alpha \mathbb{I} + AA^*)^{-1/2} \xi\| \leq c \frac{\delta}{\sqrt{\beta}}, \\ \|B^{-s}(v_{\alpha,\beta}^0 - v_{\alpha,\beta}^\delta)\| &\leq \delta \left\| (\alpha \mathbb{I} + AA^*)^{1/2} \left(\frac{\beta}{\alpha} \mathbb{I} + C_\alpha C_\alpha^* \right)^{-1} C_\alpha C_\alpha^* \right\| \leq c\delta. \end{aligned}$$

Then the interpolation inequality (21) gives us

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

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

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

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

□

Remark 1. 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

$$\psi(\theta_\psi^{-1}(\delta))(\theta_\psi^{-1}(\delta))^{\frac{1}{2(s+1)}} = \delta^{\frac{2p(s+1)+1}{(2p+1)(s+1)}}. \quad (23)$$

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

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

On the other hand, it is known [20] that under the link condition (5) the solution $x^\dagger \in \text{Range}(B^{-\mu})$ 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 multi-penalty 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 one, 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 [27] and has been recently advocated in [12].

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

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

such that

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

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

$$P_M^\beta = \{\beta = \beta_j = \beta_0 p^j, j = 0, 1, 2, \dots, M\}, p > 1, \quad (26)$$

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

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

Then for the multi-penalty regularization the quasi-optimality criterion can be implemented as follows. At first, for every $\beta = \beta_j \in P_M^\beta$ we choose $\alpha = \alpha_l = \alpha(\beta_j)$ from the set (25) such that

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

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

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

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

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 experiments, the operators A and B^{-1} are given as diagonal matrices of the size n . The matrix corresponding to the operator A has diagonal elements $a_k = k^{-r}$, $k = 1, 2, \dots, n$, $n = 50$, $r = 3$. 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

$$x^\dagger = (A^*A)^2g, \quad (27)$$

where g is a random vector which 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% of data noise.

In accordance with the theory, under the source condition (27) the Tikhonov-Phillips regularization (3) can suffer from the 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 the condition (5) is satisfied. At first, we choose the self-adjoint operator B such that the corresponding diagonal matrix has the elements $b_{kk} = b_k = k$, $k = 1, 2, \dots, n$. 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 $x = x_{\alpha,\beta}^\delta$, $x = x_\alpha^\delta$, and $x = x_{\beta,B}^\delta$.

The results are displayed in Figure 1, where each circle exhibits a relative error in solving the problems with one of 100 simulated data, for each of three regularization methods: the multi-penalty regularization (MP), the Tikhonov-Phillips regularization (TP), and the Tikhonov regularization (Tikhonov). Moreover, in Table 1 the statistical measures such as mean values, median values, standard deviation of the relative error, as well as mean values of the regularization parameters are given for each of the methods.

The numerical results confirm the theoretical conclusion that saturation of the method (3) can be potentially relaxed by the use of the method (4). Moreover, in the

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

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x_{\alpha,\beta}^\delta$	0.0071	0.0063	0.0036	$\alpha = 6.02, \beta = 0.002$
x_α^δ	0.0117	0.0117	0.0027	0.007
$x_{\beta,B}^\delta$	0.0072	0.0061	0.0035	0.002

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

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x_{\alpha,\beta}^\delta$	0.0118	0.0119	0.0029	$\alpha = 0.0066, \beta = 12.1$
x_α^δ	0.0118	0.0119	0.0028	0.0066
$x_{\beta,B}^\delta$	0.0319	0.0322	0.0028	0.0247

considered case the multi-penalty method (9), (10) performs similar to the method (4), as it has been predicted by Theorem 2.

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

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

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 ($\|B^{-s}\| \geq (n-1)^s \geq 1 \geq \|A\|$). At the same time, in both cases we can observe that the multi-penalty 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 [15] we generate the test problems of the form (1) by using the functions $shaw(n)$ and $ilaplace(n, 1)$ from the Matlab regularization toolbox [9]. These functions occur as the results of a discretization of the first kind Fredholm integral equation of the form

$$\int_a^b k(s, t) f(t) dt = g(s), \quad s \in [a, b], \quad (28)$$

with a known solution $f(t)$. As in two previous experiments, the operator A and the solution x^\dagger are given as $n \times n$ -matrix and n -dimensional vector respectively. The noisy

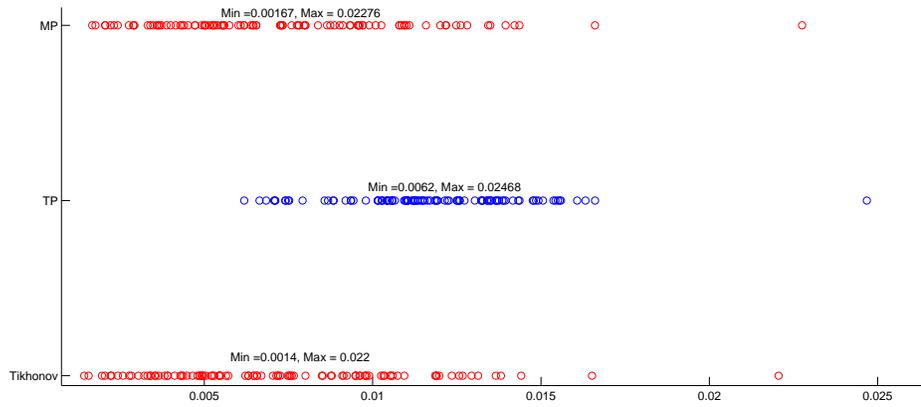


Figure 1. Numerical illustration (first experiment). The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise

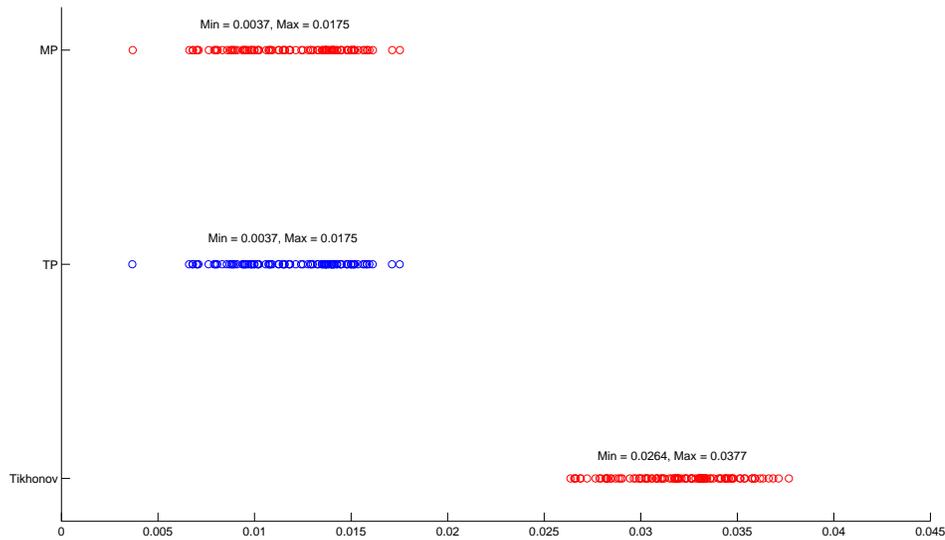


Figure 2. Numerical illustration (second experiment). The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise

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 $n \times n$ -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 the equation (28) with $a = -\pi/2$ and $b = \pi/2$. The kernel and the solution are given as

$$k(s, t) = (\cos(s) + \cos(t))^2 \left(\frac{\sin(u)}{u} \right)^2, \quad u = \pi(\sin(s) + \sin(t)),$$

$$f(t) = 2e^{-6(t-0.8)^2} + e^{-2(t+0.5)^2}.$$

The corresponding equation (28) is discretized by a simple quadrature with n equidistant points. Similar to [15] 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.

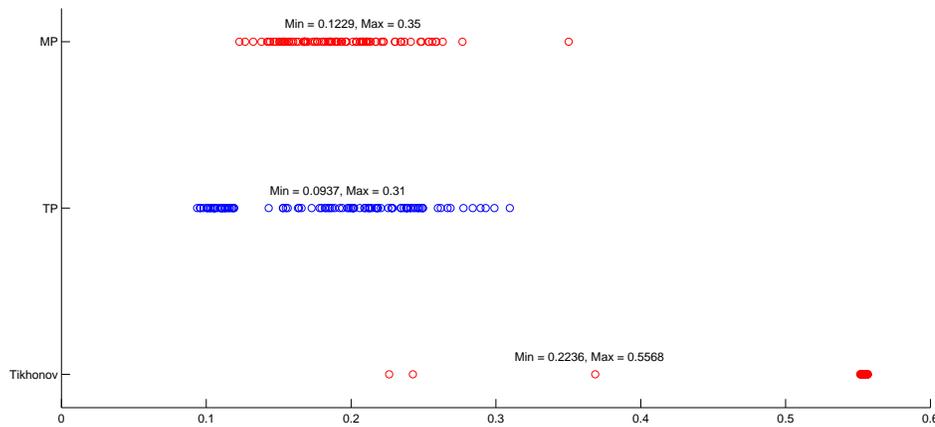


Figure 3. Numerical illustration for the function $shaw(100)$. The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise

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 the equation (28) with $a = 0$, $b = \infty$, $k(s, t) = e^{-st}$, $f(t) = e^{-t/2}$, $g(s) = (s + 1/2)^{-1}$. 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 relative errors produced by three regularization methods. Moreover, Table 4 presents statistical information about the performance of the methods.

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

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



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

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

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

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

The presented multi-penalty 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 side, an additional penalizing operator B is used to relax the saturation effect, and, on the other side, a link condition (5) is not granted a priori, as it is the case for noisy operators A , for example. Moreover, the considered multi-penalty regularization may be also relevant in the situation when some parts of data are more accurately known than others. Such a situation occurs, for example, in Geomathematics

[8], and deserves further study.

Remark 2. *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_N^α and P_M^β is crucial for obtaining good performance of the methods. We believe that a deeper study of this issue is important. One of the possible future work in this direction is to consider a choice of the sets (25), (26) using meta-learning [4, 21] that proved to be an efficient method dealing with problems of the similar type.*

Acknowledgments

The authors are supported by the Austrian Fonds Zur Förderung der Wissenschaftlichen Forschung (FWF), Grant P25424.

References

- [1] F. Bauer and M. Lukas. Comparing parameter choice methods for regularization of ill-posed problems. *Mathematics and Computers in Simulation*, 81(9):1795–1841, 2011.
- [2] M. Belge, M. Kilmer, and E. Miller. Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18:1161–1183, 2002.
- [3] A. Böttcher, B. Hofmann, U. Tautenhahn, and M. Yamamoto. Convergence rates for Tikhonov regularization from different kinds of smoothness conditions. *Appl. Anal.*, 85:555–578, 2006.
- [4] P. Brazdil, C. Giraud-Carrier, C. Soares, and R. Vilalta. *Metalearning: Applications to Data Mining*. Springer-Verlag, Berlin Heidelberg, 2009.
- [5] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Multiparameter regularization techniques for ill-conditioned linear systems. *Numerische Mathematik*, 94:203–228, 2003.
- [6] Z. Chen, Y. Lu, Y. Xu, and H. Yang. Multi-parameter Tikhonov regularization for linear ill-posed operator equations. *J. Comput. Math.*, 26(1):37–55, 2008.
- [7] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.
- [8] W. Freeden. *Multiscale modeling of spaceborne geodata*. B. G. Teubner, Leipzig, 1999.
- [9] P. Hansen. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms*, 6:1–35, 1994.
- [10] T. Hohage and M. Pricop. Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise. *Inverse Problems and Imaging*, 2:271–290, 2008.
- [11] K. Ito, B. Jin, and T. Takeuchi. Multi-parameter Tikhonov regularization. *Methods and Applications of Analysis*, 18(1):31–46, 2011.
- [12] S. Kindermann and A. Neubauer. On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization. *Inverse Problems and Imaging*, 2(2):291–299, 2008.
- [13] V. Koltchinskii and M. Yuan. Sparsity in multiple kernel learning. *Ann. Statist.*, 38(6):3660–3695, 2010.
- [14] S. Krein and Y. Petunin. Scales of Banach spaces. *Russian Math. Surv.*, 14:85–159, 1966.
- [15] S. Lu and S. V. Pereverzev. Multi-parameter regularization and its numerical realization. *Numerische Mathematik*, 118(1):1–31, 2011.
- [16] P. Mathé. Saturation of regularization methods for linear ill-posed problems in Hilbert spaces. *SIAM J. Numer. Anal.*, 42(3):968–973, 2004.

- [17] P. Mathé and B. Hofmann. How general are general source conditions? *Inverse Problems*, 24:015009, pp. 1–5, 2008.
- [18] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19:789–803, 2003.
- [19] T. Nair, S. V. Pereverzev, and U. Tautenhahn. Regularization in Hilbert scales under general smoothing conditions. *Inverse Problems*, 21:1851–1869, 2005.
- [20] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.*, 18:29–37, 1984.
- [21] V. Naumova, S. V. Pereverzev, and S. Sivananthan. A meta-learning approach to the regularized learning – case study: Blood glucose prediction. *Neural Networks*, 33:181–193, 2012.
- [22] A. Neubauer. An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM J. Numer. Anal.*, 25:1313–1326, 1988.
- [23] A. Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Appl. Anal.*, 46:59–72, 1992.
- [24] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM J. Numer. Anal.*, 34(2):517–527, 1997.
- [25] U. Tautenhahn. On a general regularization scheme for nonlinear ill-posed problems: II. Regularization in Hilbert scales. *Inverse Problems*, 14:1607–1616, 1998.
- [26] A. N. Tikhonov. *On the regularization of ill-posed problems*, volume 153. Dokl. Akad. Nauk SSSR, 1963.
- [27] A. N. Tikhonov and V. B. Glasko. *Use of the regularization methods in non-linear problems*, volume 5. USSR Comput. Math. Phys., 1965.