PARAMETER CHOICE STRATEGIES FOR MULTI-PENALTY REGULARIZATION

MASSIMO FORNASIER , VALERIYA NAUMOVA , AND SERGEI V. PEREVERZYEV

Abstract. The widespread applicability of the multi-penalty regularization is limited by the fact that theoretically optimal rate of reconstruction for a given problem can be realized by a one-parameter counterpart, provided that relevant information on the problem is available and taken into account in the regularization. In this paper, we explore the situation, where no such information is given, but still accuracy of optimal order can be guaranteed by employing multi-penalty regularization. Our focus is on the analysis and the justification of an a posteriori parameter choice rule for such a regularization scheme. First we present a modified version of the discrepancy principle within the multi-penalty regularization framework. As a consequence we provide a theoretical justification to the multi-penalty regularization scheme equipped with the a posteriori parameter choice rule. We then establish a fast numerical realization of the proposed discrepancy principle based on a model function approximation. Finally, we provide extensive numerical results which confirm and support the theoretical estimates and illustrate the robustness and the superiority of the proposed scheme compared to the "classical" regularization methods.

Key words. multi-penalty regularization, discrepancy principle, model function, order-optimal reconstruction, compensatory properties.

AMS subject classifications. 47A52, 65F22, 65J20.

1. Introduction.

In recent years there has been a fast growing interest in studying multi-penalty regularization for solving inverse and ill-posed problems. In several inspiring applications such as image reconstruction, option pricing [7, 19], reconstruction of the Earth gravity potential [30, 2], multi-penalty regularization schemes have been successfully applied. Moreover, relevant theoretical results appeared in a few papers (see, for instance, [3, 4, 5, 9, 11, 17, 26]). However, apart from recent promising results the widespread applicability of the multi-penalty regularization has remained questionable due to the fact that the theoretically optimal rate of reconstruction for a given problem can be realized by a one-parameter counterpart. One-parameter regularization is well-developed and proves to be particularly simple and effective, though only in the case when relevant information on a solution is provided. Precisely, in this case, the theory of inverse and ill-posed problems provides us with a concrete recipe how to measure the accuracy of the reconstruction.

It is well-known that the reconstruction accuracy depends on a noise model and on the smoothness of the solution. On the one hand, the noise model is provided together with the given data and cannot be changed. On the other hand, the smoothness can be measured in various spaces (see, for instance, [20]), and depending on this, different bounds on the best possible approximation can be derived. Once this space is fixed, the general regularization theory provides us with the best possible error bound that can be achieved if the regularization is performed in the chosen space. Moreover, the theory promises us that this bound can be achieved in the framework of the one-parameter regularization [8, 16].

However, one principal problem is left untouched: Who is going to tell us in which space we need to measure the smoothness of the unknown solution? Apparently, this principal question started to be discussed only recently and have not yet been systematically explored in regularization theory.

Differently from the "classical" works on regularization theory, we continue to investigate in this paper multi-penalty regularization, where one is given the freedom of performing the regularization in several spaces simultaneously with the goal of possibly achieving a better accuracy than given by an a priori fixed one-parameter counterpart. Definitely one of the leading ingredients for the optimal performance of multi-penalty regularization is an appropriate (a posteriori) choice of the multiple regularization parameters. This issue has been essentially studied in the framework of multi-parameter regularization in the papers [3, 11, 17, 12]. However, as will be seen later, the theory developed in [11, 17] is not sufficient for multi-penalty regularization and need to be extended. The conceptually closest work [26] considers a heuristic parameter choice rule strategy, though it does not provide any theoretical justification. In this paper we will combine theoretical and heuristic concepts of the parameter choice strategies to justify the optimality of multi-penalty regularization equipped with an a posteriori parameter choice rule.

The rest of the paper is organized as follows: Section 2 is devoted to studying a multi-parameter discrepancy principle, the so-called *discrepancy domain principle,* as a parameter choice rule for multi-penalty regularization. We show that there exist many combinations of the regularization parameters satisfying the discrepancy principle which corresponds to a reconstruction accuracy of the optimal order. While the estimation of the discrepancy requires the computation of a solution relative to each pair of parameters, which can be a demanding task, in Section 3 we replace the exact discrepancy by a locally approximating surrogate function of the parameters. We show that the proposed model function approximation of the discrepancy domain leads to an efficient iterative algorithm for choosing the regularization parameters. The paper is concluded by extensive numerical experiments involving toy and simulated problems which clearly demonstrate that the proposed scheme allows us to achieve better results than the corresponding one-parameter competitor. Additionally, we demonstrate how the model function approximations can be profitably use to find a domain of the regularization parameters, which lead to the optimal order of the reconstruction accuracy.

2. Preliminaries. As mentioned previously, we are in particular interested in the solution of a linear ill-posed problem

$$Ax = y \tag{2.1}$$

where $A: X \to Y$ is a bounded linear operator between Hilbert spaces X and Y with a 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 (2.1). In general, the observed data is additionally corrupted by noise ξ

$$y_\delta = Ax^\dagger + \xi, \tag{2.2}$$

where $\xi \in Y,\ \|\xi\| \le \delta,\ \delta \in (0,1)$. Due to non-closedness of $\mathcal{R}(A)$, 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].

A well-known technique to stabilize an ill-posed problem is by Tikhonov-Phillips (TP) regularization, i.e., by minimizing the functional

$$TP(\alpha; x) = \|Ax - y_\delta\|_Y^2 + \alpha \|x\|_X^2. \tag{2.3}$$

The minimization has a unique solution

$$x_\alpha^\delta = x_\alpha^\delta(y_\delta) = (\alpha \mathbb{I} + A^* A)^{-1} A^* y_\delta \tag{2.4}$$

with $\alpha > 0$ being the regularization parameter and $\mathbb{I}$ is the identity operator.

Let us shortly recall that the accuracy of the reconstruction depends both on the noise model and on the smoothness of the solution. For regularization methods as (2.3) of the operator equation (2.1) in Hilbert spaces, the smoothness of $x^\dagger$ is usually expressed in the form of the inclusion

$$x^\dagger \in A_\varphi(R) := \{x \in X : \; x = \varphi(A^*A)g, \; \|g\|_X \leq R\}, \tag{2.5}$$

where $\varphi : \; [0, \|A\|^2] \to [0,1], \; \varphi(0) = 0$, is called an index function which is assumed to be continuous, increasing, and such that $\frac{t}{\varphi(t)}$ is nondecreasing. Note that the condition of the type (2.5) is usually called the *source condition* (see, [21, 20], for instance). Then the order of the best possible accuracy, which can be guaranteed within the framework of the given noise model for x_α^δ, is given as

$$\inf_{\alpha} \sup_{x^\dagger \in A_\varphi(R)} \sup_{y_\delta} \|x^\dagger - x_\alpha^\delta(y_\delta)\|_X = O(\varphi(\theta_\varphi^{-1}(\delta))),$$

where $\theta_\varphi(t) = \varphi(t)\sqrt{t}$ and $\varphi^2((\theta_\varphi^2)^{-1}(t))$ is assumed to be concave. This order is attainable at $\alpha = \theta_\varphi^{-1}(\delta)$. Since $\varphi(t) = ct$ is the best index function for which $\frac{t}{\varphi(t)}$ is nondecreasing, the best guaranteed error for Tikhonov-Phillips regularization is known to be $O(\delta^{2/3})$, 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 [27] and changes the form of the penalty term in (2.3), namely instead of the identity operator $\mathbb{I}$ within the penalty term, an unbounded self-adjoint strictly positive operator B on the Hilbert space X is considered. In this case, the regularized solution $x_{\beta,B}^\delta$ is defined as the minimizer of the functional

$$T(\beta; x) = \|Ax - y_\delta\|_Y^2 + \beta\|Bx\|_X^2, \tag{2.6}$$

over the domain $\mathcal{D}(B)$ of the operator B. In the classical work of Tikhonov [27], the operator B is given as a square root of the self-adjoint second order differential operator.

However, the superiority of the latter regularization scheme over (2.3) has been theoretically justified only under the assumption that the operators A and B are related by the so-called *link condition* for which

$$\|B^{-s}x\| \leq \|Ax\| \leq b\|B^{-s}x\|, \qquad \text{for all } x \in X, \tag{2.7}$$

where $s > 0$ and $b \geq 1$ are some constants. Then Natterer [25] has shown that the regularized solutions $x_{\beta,B}^\delta$ converge towards the exact solution with the rate $O(\delta^{\frac{p}{p+s}})$ in the norm of X, if the regularization parameter β is chosen properly and if the exact solution $x^\dagger$ satisfies an analog of the source condition formulated in terms of the operator B as follows

$$x^\dagger \in \mathcal{D}(B^p) = \{x : \; x = B^{-p}g, \; \|g\|_X \leq R\}, \; p > 0. \tag{2.8}$$

It is easily seen that under assumptions (2.7), (2.8), and $p > 2s$ the best guaranteed error for Tikhonov regularization is better than the one guaranteed for Tikhonov-Phillips regularization.

However, the efficiency of Tikhonov regularization is limited by a serious bottleneck. Precisely, the theoretical superiority of $x_{\beta,B}^\delta$ is justified only under the link condition (2.7), which is sometimes hardly verifiable. At the same time, when the link condition is violated Tikhonov regularization may perform poorly as it will be shown in the last section.

Since in reality we do not know the smoothness of $x^\dagger$, it is not clear which of the source conditions should be taken into account in the regularization and which of the one-parameter regularization methods is more suitable for a problem at hand. In the spirit of DeVore [6], we require the blindness of the algorithms with respect to the conditions (of smoothness) on the solution (which is unknown!) for achieving optimal approximation.

In the recent paper [26] a new multi-penalty regularization scheme as an addresser of the above-mentioned problematic issue has been introduced. In this scheme one considers the multi-objective optimization of a functional of the form

$$\Phi(\alpha,\beta;u,v) := \|A(u+v)-y_\delta\|_Y^2 + \alpha\|u\|_X^2 + \beta\|Bv\|_X^2. \tag{2.9}$$

The minimizers $u_{\alpha,\beta}^\delta$ and $v_{\alpha,\beta}^\delta$ of (2.9) have the representation

$$u_{\alpha,\beta}^\delta = (\alpha\mathbb{I}+A^*A)^{-1}(A^*y_\delta - A^*Av_{\alpha,\beta}^\delta), \tag{2.10}$$

$$v_{\alpha,\beta}^\delta = \alpha(\beta B^2 + \alpha A^*A(\alpha\mathbb{I}+A^*A)^{-1})^{-1}(\alpha\mathbb{I}+A^*A)^{-1}A^*y_\delta, \tag{2.11}$$

and a reconstruction of the solution of interest $x_{\alpha,\beta}^\delta$ is calculated then as the sum of these minimizers. For the sake of completeness, we recall the theorem from [26], which shows the compensatory properties of multi-penalty regularization, in the sense that it performs similar to the best of the single-parameter regularization with the corresponding penalizing operator, either $\mathbb{I}$ or B.

Theorem 2.1.

1. *Assume that the link condition (2.7) is satisfied. Then for $\alpha > 1$ there is $\beta = \beta(\alpha,\delta,y_\delta)$ such that $x_{\alpha,\beta}^\delta = u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta$ approximates $x^\dagger$ with the best order of accuracy guaranteed by Tikhonov method.*
2. *Assume that the link condition (2.7) is violated. Then for $\beta > 1$ there is $\alpha = \alpha(\beta,\delta,y_\delta)$ such that $x_{\alpha,\beta}^\delta = u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta$ approximates $x^\dagger$ with the best order of accuracy guaranteed by Tikhonov-Phillips method.*

3. Discrepancy Domain Principle. In order to provide a practical and theoretically justified rule for the choice of optimal parameters as claimed in Theorem 2.1, we need here to extend the well-known theory of regularization for single parameter to multi-penalty regularization. Recall that for the single-parameter regularization schemes, such as (2.3)–(2.4) or (2.6), for example, the discrepancy principle [23]

$$\alpha_{DP} = \max\{\alpha : \|Ax-y_\delta\| \le c\delta\},\ c>1,\ \text{with } x = x_\alpha^\delta \text{ or } x = x_{\beta,B}^\delta,$$

is theoretically justified to be an order-optimal parameter choice rule under the conditions that in (2.5) the function φ is such that $\frac{\sqrt{t}}{\varphi(t)}$ is nondecreasing, or (2.7), (2.8) hold with $p \le s+2$.

Here we consider an extension of the classical discrepancy principle and look for a parameter set (α,β) satisfying the so-called discrepancy domain principle, i.e.,

$$(\alpha,\beta) \in D(\delta) = \{(\alpha,\beta) : \delta < \|A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - y_\delta\| \le c\delta\},\ c>1. \tag{3.1}$$

Here and below we will follow the convention that the symbol c denotes a quantity that does not depend on $\alpha,\beta,\ \delta$ and needs not be the same at different occurrences.

For technical reasons, in the following proofs we shall mainly consider the discrepancy domain principle to be given in terms of the equality, namely

$$\|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}) - y_\delta\| = c\delta,\ c > 1, \tag{3.2}$$

whereas in the numerical experiments we follow (3.1).

3.1. Error bound under satisfied link condition. We begin with the case where the link condition is satisfied. The following theorem shows that multi-penalty regularization equipped with the a posteriori parameter choice rule allows to achieve the best order of accuracy guaranteed by Tikhonov regularization.

THEOREM 3.1. *Let the link condition (2.7) be satisfied with $s > 1$, and $x^\dagger \in \mathcal{D}(B^p)$, $p \in [1, s+2]$. Then for any $(\alpha,\ \beta) \in D(\delta)$ such that $\alpha > 1$, we have an order-optimal error bound*

$$\|x^\dagger - (u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta})\| = O(\delta^{\frac{p}{p+s}}), \tag{3.3}$$

where the coefficient implicit in $O-$symbol depends only on $\|B^p x^\dagger\|$ and $\|A\|$.

Before providing the proof we need to recall the notion of a Hilbert scale $\{X_k\}$ induced by the operator B, where X_k is the completion of $\mathcal{D}(B^k)$ with respect to the Hilbert space norm $\|x\|_k = \|B^k x\|$.

Proof. The difference between the exact solution $x^\dagger$ and its approximation given by multi-penalty regularization can be bounded as

$$\|x^\dagger - (u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta})\| \le \|x^\dagger - v^\delta_{\alpha,\beta}\| + \|u^\delta_{\alpha,\beta}\|. \tag{3.4}$$

Keeping in mind that $y_\delta = Ax^\dagger + \xi$ and $\alpha > 1$ we can deduce from (2.10) that

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

Hence, it is sufficient to estimate $\|x^\dagger - v^\delta_{\alpha,\beta}\|$. Assume that (α, β) is a set of positive parameters satisfying the discrepancy domain principle and $u^\delta_{\alpha,\beta}$, $v^\delta_{\alpha,\beta}$ are the solutions corresponding to (α, β). Taking into account that $u^\delta_{\alpha,\beta}$, $v^\delta_{\alpha,\beta}$ are the minimizers of the functional (2.9), we have

$$\begin{aligned}\|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}) - y_\delta\|^2 + \alpha\|u^\delta_{\alpha,\beta}\|^2 + \beta\|Bv^\delta_{\alpha,\beta}\|^2 &\le \|Ax^\dagger - y_\delta\|^2 + \beta\|Bx^\dagger\|^2 \\ &\le \delta^2 + \beta\|Bx^\dagger\|^2.\end{aligned}$$

Since the discrepancy domain principle is satisfied, with (3.2) as a defining condition, this inequality yields

$$c^2\delta^2 + \alpha\|u^\delta_{\alpha,\beta}\|^2 + \beta\|Bv^\delta_{\alpha,\beta}\|^2 \le \delta^2 + \beta\|Bx^\dagger\|^2.$$

Keeping in mind that $\beta > 0$ and $c \ge 1$, we can conclude that $\|Bv^\delta_{\alpha,\beta}\| \le \|Bx^\dagger\|$ and

$$\begin{aligned}\|Bv^\delta_{\alpha,\beta} - Bx^\dagger\|^2 &= \langle Bv^\delta_{\alpha,\beta}, Bv^\delta_{\alpha,\beta}\rangle - 2\langle Bv^\delta_{\alpha,\beta}, Bx^\dagger\rangle + \langle Bx^\dagger, Bx^\dagger\rangle \\ &\le 2\langle Bx^\dagger, Bx^\dagger\rangle - 2\langle Bv^\delta_{\alpha,\beta}, Bx^\dagger\rangle \\ &= 2\langle Bx^\dagger - Bv^\delta_{\alpha,\beta}, Bx^\dagger\rangle = 2\langle B^2(x^\dagger - v^\delta_{\alpha,\beta}), x^\dagger\rangle \\ &= 2\langle B^{2-p}(x^\dagger - v^\delta_{\alpha,\beta}), B^p x^\dagger\rangle \le 2R\|B^{2-p}(x^\dagger - v^\delta_{\alpha,\beta})\|.\end{aligned}$$

In terms of the Hilbert scales the latter inequality can be rewritten as

$$\|x^\dagger - v^\delta_{\alpha,\beta}\|_1^2 \le 2R\|x^\dagger - v^\delta_{\alpha,\beta}\|_{2-p}.$$

The rest of the proof is based on the interpolation inequality

$$\|x\|_r \le \|x\|_{-s}^{(a-r)/(s+a)}\|x\|_a^{(s+r)/(s+a)} \tag{3.5}$$

which holds for all $r \in [-s, a],\ a + s \neq 0$.

Taking $r = 2 - p$ and $a = 1$ we can continue as follows

$$\|x^\dagger - v^\delta_{\alpha,\beta}\|_1^2 \le 2R\|x^\dagger - v^\delta_{\alpha,\beta}\|_{2-p} \le 2R\|x^\dagger - v^\delta_{\alpha,\beta}\|_1^{(2+s-p)/(s+1)}\|x^\dagger - v^\delta_{\alpha,\beta}\|_{-s}^{(p-1)/(s+1)},$$

which is the same as

$$\|x^\dagger - v^\delta_{\alpha,\beta}\|_1 \le (2R)^{(s+1)/(s+p)}\|x^\dagger - v^\delta_{\alpha,\beta}\|_{-s}^{(p-1)/(s+p)}.$$

Observe now that from (2.10) and (3.2) it follows that for $\alpha > 1$ we have

$$\begin{aligned} c\delta &= \|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}) - y_\delta\| = \|Av^\delta_{\alpha,\beta} - y_\delta + A(\alpha\mathbb{I} + A^*A)^{-1}(A^*y_\delta - A^*Av^\delta_{\alpha,\beta})\| \\ &= \|(\mathbb{I} - A(\alpha\mathbb{I} + A^*A)^{-1}A^*)(Av^\delta_{\alpha,\beta} - y_\delta)\| \ge (1 + \|A\|^2)^{-1}\|Av^\delta_{\alpha,\beta} - y_\delta\|. \end{aligned} \tag{3.6}$$

Using this bound and the link condition (2.7) we get

$$\begin{aligned} \|x^\dagger - v^\delta_{\alpha,\beta}\|_{-s} &\le \|Ax^\dagger - Av^\delta_{\alpha,\beta}\| \le (\|y - y_\delta\| + \|y_\delta - Av^\delta_{\alpha,\beta}\|) \\ &\le \delta(1 + c(1 + \|A\|^2)). \end{aligned}$$

Then we obtain

$$\|x^\dagger - v^\delta_{\alpha,\beta}\|_1 \le (2R)^{(s+1)/(s+p)}\left(1 + c(1 + \|A\|^2)\right)^{(p-1)/(s+p)}\delta^{(p-1)/(s+p)}.$$

Using again interpolation inequality (3.5) with $r = 0$ and $a = 1$, we finally receive the statement of the theorem

$$\begin{aligned} \|x^\dagger - v^\delta_{\alpha,\beta}\|_0 &\le \|x^\dagger - v^\delta_{\alpha,\beta}\|_1^{s/(s+1)}\|x^\dagger - v^\delta_{\alpha,\beta}\|_{-s}^{1/(s+1)} \\ &\le (2R)^{s/(s+p)}\left(1 + c(1 + \|A\|^2)\right)^{p/(s+p)}\delta^{p/(s+p)}. \end{aligned} \tag{3.7}$$

The fact that for $x^\dagger \in \mathcal{D}(B^p)$ the order of accuracy $O(\delta^{\frac{p}{s+p}})$ cannot be in general improved follows, for example, from [24].
□

REMARK 3.2. *Actually, one can choose $\alpha > \alpha^0 > 0$ changing the lower bound in (3.6) and eventually the constant in (3.7). However, the larger α^0 is the better and just to fix a bound we choose $\alpha^0 \equiv 1$.*

3.2. Error bound under violated link condition. Now, assuming that the link condition is violated, we will show that the best order of accuracy guaranteed by Tikhonov-Phillips regularization is achieved by multi-penalty regularization equipped with the a posteriori parameter choice rule.

THEOREM 3.3. *Let (2.5) be satisfied. Then, for any $(\alpha,\ \beta) \in D(\delta)$ such that $\beta > \|B^{-2}\|\max\{\|A\|, \|A\|^2\}$, we have an order-optimal error bound*

$$\|x^\dagger - (u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta})\| = O(\varphi(\theta_\varphi^{-1}(\delta))), \tag{3.8}$$

where $\frac{\sqrt{t}}{\varphi(t)}$ *is a nondecreasing function and* $\theta_\varphi(t) = \varphi(t)\sqrt{t}$.

Proof. From (2.10), (2.11) we get

$$\beta B^2 v^\delta_{\alpha,\beta} = A^*(y_\delta - A(v^\delta_{\alpha,\beta} + u^\delta_{\alpha,\beta})).$$

Due to our choice of β ($\beta > \|B^{-2}\|\|A\|$), it gives us the following bound

$$\|v^\delta_{\alpha,\beta}\| \le \|y_\delta - A(v^\delta_{\alpha,\beta} + u^\delta_{\alpha,\beta})\| \le c\delta.$$

Moreover, using the notation $x^\delta_\alpha = (\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta$ and (2.10) we can rewrite

$$u^\delta_{\alpha,\beta} = x^\delta_\alpha - (\alpha\mathbb{I} + A^*A)^{-1}A^*Av^\delta_{\alpha,\beta}.$$

In view of the discrepancy domain principle it holds

$$\begin{aligned} c\delta &= \|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}) - y_\delta\| \\ &= \|Ax^\delta_\alpha - y_\delta + Av^\delta_{\alpha,\beta} - A(\alpha\mathbb{I} + A^*A)^{-1}A^*Av^\delta_{\alpha,\beta}\| \\ &\ge \|Ax^\delta_\alpha - y_\delta\| - \|A\|\|v^\delta_{\alpha,\beta}\|, \end{aligned}$$

which is the same as

$$\begin{aligned} \|Ax^\delta_\alpha - y_\delta\| &\le c\delta + \|A\|\|v^\delta_{\alpha,\beta}\| \\ &\le c\delta + \frac{1}{\beta}\|B^{-2}\|\|A\|^2\|y_\delta - A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta})\| \\ &\le \hat{c}\delta. \end{aligned}$$

Thus, if there are sets (α,β) that belong to the discrepancy domain and such that β is sufficiently large ($\beta > \|B^{-2}\|\|A\|^2$)) then Tikhonov-Phillips regularization x^δ_α with the same α as in the pair (α,β) meets the discrepancy principle with a constant $\hat{c} > 1$. Then from [22] it follows that for $x^\dagger = \varphi(A^*A)g$ such that $\frac{\sqrt{t}}{\varphi(t)}$ is a nondecreasing function, we have

$$\|x^\dagger - x^\delta_\alpha\| \le c_1\varphi(\theta^{-1}_\varphi(\delta)),$$

where the constant c_1 does not depend on δ and the order $O(\varphi(\theta^{-1}_\varphi(\delta)))$ cannot be improved in general.

Therefore, for (α,β) meeting the discrepancy principle with a sufficiently large β we have

$$\begin{aligned} \|x^\dagger - u^\delta_{\alpha,\beta} - v^\delta_{\alpha,\beta}\| &\le \|x^\dagger - x^\delta_\alpha + (\alpha\mathbb{I} + A^*A)^{-1}A^*Av^\delta_{\alpha,\beta} - v^\delta_{\alpha,\beta}\| \\ &\le \|x^\dagger - x^\delta_\alpha\| + \alpha\|(\alpha\mathbb{I} + A^*A)^{-1}\|\|v^\delta_{\alpha,\beta}\| \\ &\le c_1\varphi(\theta^{-1}_\varphi(\delta)) + c\delta \le (c_1 + c)\varphi(\theta^{-1}_\varphi(\delta)), \end{aligned}$$

and the order $O(\varphi(\theta^{-1}_\varphi(\delta)))$ cannot be in general improved.
□

REMARK 3.4. *In the context of Theorem 3.1 and Theorem 3.3 one still needs to decide which of the parameters* α, β *should be taken large enough. If such a decision is made then the theorems guarantee the optimal order of accuracy provided that another parameter has been chosen such that* $(\alpha,\beta) \in D(\delta)$. *In Section 5.1 we propose to address the issue about the choice of the larger parameter by combining the discrepancy domain principle with the heuristic quasi-optimality criterion.*

4. Model function approximation for the discrepancy domain principle. In the present section we discuss a numerical realization of the discrepancy domain principle based on the model function approximation [13, 15, 29, 17, 18].

For the standard one-parameter Tikhonov-Phillips method it has been proposed in [13, 15, 29] to implement the discrepancy principle by approximating the discrepancy $\|Ax_\alpha^\delta - y_\delta\|$ locally by means of some simple model function $m(\alpha)$ of the parameter α. Then the underlying concept has been extended to the multi-parameter regularization. We may refer to [17, 18] for more details on this issue.

However, let us shortly mention that the existing approaches to the construction of a model function presuppose that values of the regularization parameters are smaller than 1. Therefore, they cannot be directly used in our case, since, as can be seen from the above theorems, the order-optimality is achieved under the assumption that $(\alpha,\ \beta) \in D(\delta)$ and one of the parameters is sufficiently large. This forces us to consider a special form of the model function which allows for an approximation of the corresponding regions of the parameters, and which is able to satisfy the specifics of the presented multi-penalty scheme. For the subsequent analysis we need the following proposition which can be proven similarly to [15].

PROPOSITION 4.1. *The minimizers $u_{\alpha,\beta}^\delta$, $v_{\alpha,\beta}^\delta$ are differentiable as functions of $\alpha,\ \beta > 0$.*

Now let $F(\alpha, \beta)$ denotes the minimal value of the functional (2.9) for given $\alpha,\ \beta > 0$, i.e., $F(\alpha, \beta) := \Phi(\alpha, \beta; u_{\alpha,\beta}^\delta, v_{\alpha,\beta}^\delta)$. The partial derivatives of the corresponding function are derived in the following lemma.

LEMMA 4.2. *The first order partial derivatives of $F(\alpha, \beta)$ are given by*

$$\frac{\partial F}{\partial \alpha}(\alpha, \beta) = \|u_{\alpha,\beta}^\delta\|^2, \qquad \frac{\partial F}{\partial \beta}(\alpha, \beta) = \|Bv_{\alpha,\beta}^\delta\|^2.$$

Proof. First it is useful to observe that for any function f of the form $f(t) = \|Qx_t - g\|^2$ its derivative can be represented as

$$\frac{df}{dt} = 2\langle Qx_t - g, Qx_t' \rangle,$$

where x_t' is the derivative of x_t as a function of the index t.

Using this observation, we have

$$\begin{aligned}
\frac{\partial F}{\partial \alpha}(\alpha, \beta) &= 2\langle A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - y_\delta, A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta)_\alpha' \rangle + 2\alpha\langle u_{\alpha,\beta}^\delta, (u_{\alpha,\beta}^\delta)_\alpha' \rangle \\
&\quad + \|u_{\alpha,\beta}^\delta\|^2 + 2\beta\langle Bv_{\alpha,\beta}^\delta, B(v_{\alpha,\beta}^\delta)_\alpha' \rangle \\
&= \|u_{\alpha,\beta}^\delta\|^2 + 2\langle A^*A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - A^*y_\delta + \alpha u_{\alpha,\beta}^\delta, (u_{\alpha,\beta}^\delta)_\alpha' \rangle \\
&\quad + 2\langle A^*A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta) - A^*y_\delta + \beta B^2 v_{\alpha,\beta}^\delta, (v_{\alpha,\beta}^\delta)_\alpha' \rangle.
\end{aligned}$$

Observe also that (2.10), (2.11) can be rewritten in the variational form as

$$\left\{ \begin{array}{lcl} \langle A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta), Ag \rangle + \alpha\langle u_{\alpha,\beta}^\delta, g \rangle & = & \langle y_\delta, Ag \rangle \\ \langle A(u_{\alpha,\beta}^\delta + v_{\alpha,\beta}^\delta), Ag \rangle + \beta\langle B^2 v_{\alpha,\beta}^\delta, g \rangle & = & \langle y_\delta, Ag \rangle, \end{array} \right. \tag{4.1}$$

for all $g \in X$.

Then, by taking $g = (u_{\alpha,\beta}^\delta)_\alpha'$ in the first equation and $g = (v_{\alpha,\beta}^\delta)_\alpha'$ in the second one of (4.1) and plugging them in the above representation for $\frac{\partial F}{\partial \alpha}(\alpha, \beta)$, we get that $\frac{\partial F}{\partial \alpha}(\alpha, \beta) = \|u_{\alpha,\beta}^\delta\|^2$.

Analogously, one can derive the representation for $\frac{\partial F}{\partial \beta}(\alpha,\beta)$.

□

By definition (2.9) and Lemma 4.2, we can rewrite (3.2) for $(\alpha,\beta)\in D(\delta)$ as

$$F(\alpha,\beta)-\alpha\frac{\partial F}{\partial \alpha}-\beta\frac{\partial F}{\partial \beta}=c^2\delta^2. \tag{4.2}$$

Note that by taking $g=u^\delta_{\alpha,\beta}$ in the first equation, and $g=v^\delta_{\alpha,\beta}$ in the second equation of (4.1) and summing them up, we obtain

$$\alpha\|u^\delta_{\alpha,\beta}\|^2+\beta\|Bv^\delta_{\alpha,\beta}\|^2+\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2=\langle y_\delta, A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\rangle.$$

Then, using this equality, we can derive

$$\begin{aligned}F(\alpha,\beta)&=\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})-y_\delta\|^2+\alpha\|u^\delta_{\alpha,\beta}\|^2+\beta\|Bv^\delta_{\alpha,\beta}\|^2\\&=\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2-2\langle A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta}),y_\delta\rangle+\|y_\delta\|^2+\alpha\|u^\delta_{\alpha,\beta}\|^2+\beta\|Bv^\delta_{\alpha,\beta}\|^2\\&=\|y_\delta\|^2-\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2-\alpha\|u^\delta_{\alpha,\beta}\|^2-\beta\|Bv^\delta_{\alpha,\beta}\|^2. \qquad (4.3)\end{aligned}$$

In the following, we discuss how one can construct a simple parameterized function, the so-called *model function* $m(\alpha,\beta)$, which preserves the main properties of the function $F(\alpha,\beta)$. To this end, we first derive a new form of the model function which provides an approximation taking into account our specific interest in the parameter choice. Then we present necessary specifications for each of the two cases, when either α or β is sufficiently large.

4.1. Model function construction. In the following, we make the technical assumption that

$$\|y_\delta\|^2-\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2=C_1\delta^2.$$

This assumption can be rigorously justified in the regime of $\delta\in(\delta_0,1)$, where $\delta_0\approx 0,\ \delta_0>0$. In fact, one can find a positive number $C=C(\delta_0)$ such that

$$\begin{aligned}\|y_\delta\|^2-\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2&\le C(\delta_0)(\|y_\delta\|-\|A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|)^2\\&\le C(\delta_0)\|y_\delta-A(u^\delta_{\alpha,\beta}+v^\delta_{\alpha,\beta})\|^2\\&\le C(\delta_0)c\delta^2.\end{aligned}$$

Thus, our technical assumption holds true with $C_1=C(\delta_0)c$. Then we get the approximate formula

$$F(\alpha,\beta)\approx C_1\delta^2-\alpha\frac{\partial F}{\partial \alpha}(\alpha,\beta)-\beta\frac{\partial F}{\partial \beta}(\alpha,\beta).$$

Now we want to derive a surrogate function $m(\alpha,\beta)$ for which the above formula is exact such that we define

$$m(\alpha,\beta)=C_1\delta^2-\alpha\frac{\partial m}{\partial \alpha}(\alpha,\beta)-\beta\frac{\partial m}{\partial \beta}(\alpha,\beta).$$

Solving this partial differential equation we obtain a parametric family of the solutions

$$m(\alpha,\beta)=C_1\delta^2+\frac{C_2}{\alpha}+\frac{C_3}{\beta}+\frac{C_4}{\sqrt{\alpha\beta}}, \tag{4.4}$$

where $C_1,\ C_2,\ C_3,\ C_4$ are constants to be determined. We refer to the function (4.4) as the four-parameter model function.

We want to realize an approximate solution to (3.2) by an iterative procedure, which, using the model functions of the form (4.4), produces sequences $\{\beta_k(\alpha)\}, k = 1, 2, \dots$, with the fixed $\alpha = \alpha^* > 1$ and $\{\alpha_k(\beta)\}, k = 1, 2, \dots$, with $\beta = \beta^* > 1$.

To this end, we assume that $\alpha = \alpha^*,\ \beta = \beta_k(\alpha^*)$, or $\beta = \beta^*,\ \alpha = \alpha_k(\beta^*)$, have been already found, and the minimizers $u^\delta_{\alpha,\beta},\ v^\delta_{\alpha,\beta}$ are given by the formulae (2.10), (2.11) for these values of the parameters. Then we determine $C_1, \dots, C_4$ in such a way that the corresponding function (4.4) interpolates the function $F(\alpha, \beta)$ and its first order partial derivatives, as well as the mixed derivative at the given point (α, β). Then, it means that the parameters $C_1, \dots, C_4$ should solve the system

$$\tag{4.5} \left\{ \begin{array}{lllll} m(\alpha,\beta) & = & C_1\delta^2 + \frac{C_2}{\alpha} + \frac{C_3}{\beta} + \frac{C_4}{\sqrt{\alpha\beta}} & = & F(\alpha,\beta), \\ \frac{\partial m}{\partial \beta}(\alpha,\beta) & = & -\frac{C_3}{\beta^2} - \frac{1}{2}\frac{C_4}{\sqrt{\alpha\beta^3}} = \frac{\partial F}{\partial \beta}(\alpha,\beta) & = & \|Bv^\delta_{\alpha,\beta}\|^2, \\ \frac{\partial m}{\partial \alpha}(\alpha,\beta) & = & -\frac{C_2}{\alpha^2} - \frac{1}{2}\frac{C_4}{\sqrt{\alpha^3\beta}} = \frac{\partial F}{\partial \alpha}(\alpha,\beta) & = & \|u^\delta_{\alpha,\beta}\|^2, \\ \frac{\partial^2 m}{\partial\alpha\partial\beta}(\alpha,\beta) & = & \frac{1}{4}\frac{C_4}{\sqrt{\alpha^3\beta^3}} = \frac{\partial^2 F}{\partial\alpha\partial\beta}(\alpha,\beta) & = & 2\langle u^\delta_{\alpha,\beta}, \frac{\partial u^\delta_{\alpha,\beta}}{\partial\beta}(\alpha,\beta)\rangle. \end{array} \right.$$

For clarity, below we derive the formula for the calculation of $\frac{\partial u^\delta_{\alpha,\beta}}{\partial\beta}$. By (2.10) we have that

$$\frac{\partial u^\delta_{\alpha,\beta}}{\partial\beta} = -(\alpha\mathbb{I} + A^*A)^{-1}A^*A\frac{\partial v^\delta_{\alpha,\beta}}{\partial\beta}.$$

At first observe that the derivative of functions of the form $z(\beta) = (\beta B^2 + C)^{-1}g$ is given by

$$\frac{dz}{d\beta} = -(\beta B^2 + C)^{-1}B^2(\beta B^2 + C)^{-1}g.$$

Then, using this fact, we can easily deduce that

$$\tag{4.6} \frac{\partial v^\delta_{\alpha,\beta}}{\partial\beta} = -\alpha(\beta B^2+\alpha A^*A(\alpha\mathbb{I}+A^*A)^{-1})^{-1}B^2(\beta B^2+\alpha A^*A(\alpha\mathbb{I}+A^*A)^{-1})^{-1}(\alpha\mathbb{I}+A^*A)^{-1}A^*y_\delta.$$

Combining (4.6) with (4.5) and solving the above system of equations, we finally derive the formulae for the unknown coefficients at any point $(\alpha, \beta) \in D(\delta)$:

$$\tag{4.7} \left\{ \begin{array}{lll} C_1(\alpha,\beta) & = & (\|y_\delta\|^2 - \|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta})\|^2)/\delta^2, \\ C_2(\alpha,\beta) & = & -\alpha^2\|u^\delta_{\alpha,\beta}\|^2 - \frac{1}{2}C_4\sqrt{\frac{\alpha}{\beta}}, \\ C_3(\alpha,\beta) & = & -\beta^2\|Bv^\delta_{\alpha,\beta}\|^2 - \frac{1}{2}C_4\sqrt{\frac{\beta}{\alpha}}, \\ C_4(\alpha,\beta) & = & 8\sqrt{\alpha^5\beta^3}\langle u^\delta_{\alpha,\beta}, (\alpha\mathbb{I} + A^*A)^{-1}A^*A(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}B^2 \times \\ & & \times(\beta B^2 + \alpha A^*A(\alpha\mathbb{I} + A^*A)^{-1})^{-1}(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta\rangle. \end{array} \right.$$

Then the model function (4.4) with the obtained coefficients (4.7) is used to find an updated value of the regularization parameter $\beta = \beta_{k+1} = \beta_{k+1}(\alpha^*),\ \alpha^* > 1$, i.e., by solving in β the equation

$$m(\alpha^*,\beta) - \alpha^*\frac{\partial m}{\partial\alpha}(\alpha^*,\beta) - \beta\frac{\partial m}{\partial\beta}(\alpha^*,\beta) = c^2\delta^2,$$

which corresponds to the model function approximation of the discrepancy principle (4.2). It is easy to see that this equation is equivalent to a quadratic equation, and its solutions $\beta = \beta_{k+1}$, if they exist, are given as

$$\beta_{k+1} = \beta_{k+1}^{1,2} = \frac{16C_3^2}{\left(-\frac{2C_4}{\sqrt{\alpha}} \pm \sqrt{\frac{4C_4^2}{\alpha} - 8C_3(\frac{2C_2}{\alpha} + (C_1 - c^2)\delta^2)}\right)^2}. \tag{4.8}$$

Following a similar path, we can update the value of the regularization parameter $\alpha = \alpha_{k+1} = \alpha_{k+1}(\beta^*)$, $\beta^* > 1$, by solving w.r.t. α the equation of the approximate discrepancy principle

$$m(\alpha, \beta^*) - \alpha \frac{\partial m}{\partial \alpha}(\alpha, \beta^*) - \beta^* \frac{\partial m}{\partial \beta}(\alpha, \beta^*) = c^2\delta^2,$$

whose solutions are given as

$$\alpha_{k+1} = \alpha_{k+1}^{1,2} = \frac{16C_2^2}{\left(-\frac{2C_4}{\sqrt{\beta}} \pm \sqrt{\frac{4C_4^2}{\beta} - 8C_2(\frac{2C_3}{\beta} + (C_1 - c^2)\delta^2)}\right)^2}. \tag{4.9}$$

Finally, we can formulate a parallel iterative algorithm based on the model function approximation in the form of an alternating procedure as follows:

Step 0. Given δ, c, y_δ, A, and α_0, $\beta_0 : (\alpha_0, \beta_0) \notin D(\delta)$, set $k = 0$.

Step 1.
(1) Fix $\alpha^* > 1$ and calculate $x^\delta_{\alpha^*,\beta_k} = u^\delta_{\alpha^*,\beta_k} + v^\delta_{\alpha^*,\beta_k}$; calculate the coefficients $C_1(\alpha^*, \beta_k)$, $C_2(\alpha^*, \beta_k)$, $C_3(\alpha^*, \beta_k)$, and $C_4(\alpha^*, \beta_k)$ in accordance with (4.7), where $\alpha = \alpha^*$, $\beta = \beta_k$; update $\beta = \beta_{k+1}$ in accordance with (4.8), where we choose among $\beta^{1,2}_{k+1}$ the one that belongs to $(0, 1)$.
(2) Calculate $x^\delta_{\alpha^*,\beta_{k+1}}$ by solving (2.10) and (2.11) with $\alpha = \alpha^*$ and $\beta = \beta_{k+1}$.
(3) GO to Step 2 if the stopping criteria $\|Ax^\delta_{\alpha^*,\beta_{k+1}} - y_\delta\| \le c\delta$ is satisfied; otherwise set $k = k + 1$, GOTO (1).

Step 2.
(1) Fix $\beta^* > 1$ and calculate $x^\delta_{\alpha_k,\beta^*} = u^\delta_{\alpha_k,\beta^*} + v^\delta_{\alpha_k,\beta^*}$; calculate the coefficients $C_1(\alpha_k, \beta^*)$, $C_2(\alpha_k, \beta^*)$, $C_3(\alpha_k, \beta^*)$, and $C_4(\alpha_k, \beta^*)$ in accordance with (4.7), where $\alpha = \alpha_k$, $\beta = \beta^*$; update $\alpha = \alpha_{k+1}$ in accordance with (4.9), where we choose among $\alpha^{1,2}_{k+1}$ the one that belongs to $(0, 1)$.
(2) Calculate $x^\delta_{\alpha_{k+1},\beta^*}$ by solving (2.10) and (2.11) with $\alpha = \alpha_{k+1}$ and $\beta = \beta^*$.
(3) STOP if the stopping criteria $\|Ax^\delta_{\alpha_{k+1},\beta^*} - y_\delta\| \le c\delta$ is satisfied; otherwise set $k = k + 1$, GOTO (1).

Remark 4.3. *Note that because the definitions of β_{k+1} (4.8) and α_{k+1} (4.9) involve the solution of the quadratic equations at each iteration, it may not always be straightforward to pick β_{k+1} or α_{k+1} : it may happen that an equation has either two or no solutions. In the first case, when two solutions exist and one of them is positive, one naturally takes the smallest positive root as the new parameter. In the latter case, when no positive solution exists, the following approximation formulae for β_{k+1} and α_{k+1} can be used*

$$\beta_{k+1} = \frac{4\alpha^* C_3^2}{C_4^2},$$

$$\alpha_{k+1} = \frac{4\beta^* C_2^2}{C_4^2},$$

which are the least squares solutions of the equations $G^m(\alpha^*,\beta) = c^2\delta^2$ *and* $G^m(\alpha,\beta^*) = c^2\delta^2$ *respectively.*

In our numerical experiments never occurred that there were no positive solutions to define $\beta_{k+1}, \alpha_{k+1}$.

REMARK 4.4. *Potentially Newton-type reconstruction methods could also be used to reconstruct the discrepancy domain. For instance, the authors in [15] considered a modification of the original Newton method to solve the damped discrepancy principle. However, in our case the application of any Newton-type method requires knowledge of the partial derivatives of* $u^\delta_{\alpha,\beta}$ *and* $v^\delta_{\alpha,\beta}$*, which in turn can be obtained by solving additionally four equations of the form (4.6) at each point. At the same time, employing the model function approach, we are asked for only one partial derivative* $\frac{\partial v^\delta_{\alpha,\beta}}{\partial\beta}$*, which makes each iteration step less expensive to perform.*

4.2. Properties of the model function approximation. In this subsection we are going to show that the algorithm of the four-parameter model function approximation produces decreasing sequences of the regularization parameters $\{\alpha_k(\beta^*)\}$ with $\beta^* > 1$ and $\{\beta_k(\alpha^*)\}$ with $\alpha^* > 1$ provided that in each step the discrepancy is larger than a given threshold.

In each updating step the discrepancy function

$$G(\alpha,\beta) = \|A(u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}) - y^\delta\|^2$$

is approximated by the function

$$G^m(\alpha,\beta) = m(\alpha,\beta) - \alpha\frac{\partial m}{\partial\alpha}(\alpha,\beta) - \beta\frac{\partial m}{\partial\beta}(\alpha,\beta).$$

By definition

$$G(\alpha,\beta) = F(\alpha,\beta) - \alpha\frac{\partial F}{\partial\alpha} - \beta\frac{\partial F}{\partial\beta}$$

and for any $k = 0, 1, \dots,$ we have

$$G(\alpha^*,\beta_k) = G^m(\alpha^*,\beta_k), \quad G(\alpha_k,\beta^*) = G^m(\alpha_k,\beta^*).$$

Now we show that the sequences of the parameters produced by the model function approximation are decreasing with k.

THEOREM 4.5. *Assume that for* (α^*,β_k) *we have* $\|A(u^\delta_{\alpha^*,\beta_k} + v^\delta_{\alpha^*,\beta_k}) - y^\delta\| > c\delta$*. If* $\beta = \beta_{k+1}$ *is given by the formula (4.8) as the positive solution of the equation* $G^m(\alpha^*,\beta) = c^2\delta^2$ *corresponding to the model function approximation of the discrepancy principle, then* $\beta_{k+1} < \beta_k$.

Proof. Observe that $g(\beta) := G^m(\alpha^*,\beta)$ is an increasing function of β since

$$\begin{aligned}
\frac{dg(\beta)}{d\beta} &= \frac{\partial m}{\partial\beta}(\alpha^*,\beta) - \alpha^*\frac{\partial^2 m}{\partial\alpha\partial\beta}(\alpha^*,\beta) - \frac{\partial m}{\partial\beta}(\alpha^*,\beta) - \beta\frac{\partial^2 m}{\partial\beta^2}(\alpha^*,\beta)\\
&= -\frac{1}{4}\frac{C_4}{\sqrt{\alpha^*\beta^3}} - \beta\left(2\frac{C_3}{\beta^3} + \frac{3}{4}\frac{C_4}{\sqrt{\alpha^*\beta^5}}\right)\\
&= -\frac{1}{4}\frac{C_4}{\sqrt{\alpha^*\beta^3}} - \frac{1}{\beta^2}\left(-2\beta^2\|Bv^\delta_{\alpha^*,\beta}\|^2 - C_4\sqrt{\frac{\beta}{\alpha^*}}\right) - \frac{3}{4}\frac{C_4}{\sqrt{\alpha^*\beta^3}} = 2\|Bv^\delta_{\alpha^*,\beta}\|^2 > 0.
\end{aligned}$$

Since β_{k+1} satisfies $g(\beta_{k+1}) = G^m(\alpha^*, \beta_{k+1}) = c^2\delta^2$, from $g(\beta_k) = G^m(\alpha^*, \beta_k) = G(\alpha^*, \beta_k) > c^2\delta^2$ and the monotonicity of $g(\beta)$, we have

$$\beta_{k+1} < \beta_k.$$

□

A similar theorem is also valid for the case of $\alpha_{k+1}(\beta^*),\ \beta^* > 1$.

THEOREM 4.6. *Assume that for* (α_k, β^*) *we have* $\|A(u^\delta_{\alpha_k,\beta^*} + v^\delta_{\alpha_k,\beta^*}) - y^\delta\| > c\delta$. *If* $\alpha = \alpha_{k+1}$ *is the minimal positive solution of the equation* $G^m(\alpha, \beta^*) = c^2\delta^2$*, then* $\alpha_{k+1} < \alpha_k$.

REMARK 4.7. *From the above theorems it follows that the discrepancy domain can be approximately reconstructed by taking grids of the parameters*

$$\begin{aligned} &\alpha \in Q^\alpha_N = \{\alpha = \alpha_i = \alpha_0 q^i,\ i = 0, 1, 2, \ldots, N\},\ q > 1, \\ &\beta \in P^\beta_M = \{\beta = \beta_j = \beta_0 p^j,\ j = 0, 1, 2, \ldots, M\},\ p > 1. \end{aligned} \tag{4.10}$$

Then one constructs two sequences $\{\alpha_k(\beta^*)\}$ *for each* $\beta^* = \beta_j > 1,\ \beta^* \in P^\beta_M$ *and* $\{\beta_k(\alpha^*)\}$ *for each* $\alpha^* = \alpha_i > 1,\ \alpha^* \in Q^\alpha_N$. *The constructed points* $(\alpha_k(\beta^*), \beta^*)$ *and* $(\alpha^*, \beta_k(\alpha^*))$ *will converge and, as numerical experiments show, lie in the discrepancy domain.*

5. Numerical realization and illustrations. In this section we provide a computational confirmation that, as predicted by Theorem 3.1 and Theorem 3.3, multi-penalty regularization may perform better than the corresponding one-parameter counterpart. The regularization parameters for both one-parameter regularization schemes (2.3) and (2.6) are chosen in accordance with the discrepancy principle, whereas as the parameter choice rule for multi-penalty regularization we consider the newly introduced discrepancy domain principle in combination with the quasi-optimality criterion. The latter was originally proposed in [28] and has been recently advocated in [14]. Moreover, we show that the discrepancy domain can be accurately and fast approximated using the model function approach, introduced in the previous section.

Finally, motivated by some positive results in [17], we also consider the three-penalty regularization scheme as a possible step towards generalization of the given multi-penalty regularization approach.

5.1. Discrepancy Domain Principle and Quasi-Optimality Criterion. We start the discussion of the numerical results with the realization of the discrepancy domain principle in combination with the quasi-optimality criterion. Recall that from Theorem 3.1 and Theorem 3.3 it follows that the application of the multi-penalty regularization leads to an order-optimal solution $x^\delta_{\alpha,\beta} = u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}$ that is achieved either for $(\alpha, \beta) \in D(\delta),\ \alpha > 1$ or $(\alpha, \beta) \in D(\delta),\ \beta > 1$.

To find the pairs (α, β) which allow us to achieve this order-optimal reconstruction we consider the two grids of the parameters (4.10) and search for those (α, β) which satisfy the discrepancy domain principle (3.1).

Then, one would ideally find the pairs $(\alpha_{i_*}, \beta_J),\ (\alpha_I, \beta_{j_*}) \in D(\delta)$ such that $\alpha_I, \beta_J > 1$ and

$$\begin{aligned} \|x^\dagger - x^\delta_{\alpha_{i_*},\beta_J}\| &= \min\{\|x^\dagger - x^\delta_{\alpha_i,\beta_j}\| :\ (\alpha_i, \beta_j) \in D(\delta),\ \beta_j > 1\}, \\ \|x^\dagger - x^\delta_{\alpha_I,\beta_{j_*}}\| &= \min\{\|x^\dagger - x^\delta_{\alpha_i,\beta_j}\| :\ (\alpha_i, \beta_j) \in D(\delta),\ \alpha_i > 1\}, \end{aligned}$$

and then choose the one that allows the most accurate reconstruction, i.e.,

$$(\alpha_+, \beta_+) = \begin{cases} (\alpha_{i_*}, \beta_J), & \text{if} \quad \|x^\dagger - x^\delta_{\alpha_{i_*},\beta_J}\| \le \|x^\dagger - x^\delta_{\alpha_I,\beta_{j_*}}\|, \\ (\alpha_I, \beta_{j_*}), & \text{if} \quad \|x^\dagger - x^\delta_{\alpha_{i_*},\beta_J}\| > \|x^\dagger - x^\delta_{\alpha_I,\beta_{j_*}}\|. \end{cases}$$

Unfortunately, since $x^\dagger$ is unknown, one can not determine such an ideal pair of the parameters exactly. However, due to the belief [8] that in the quasi-optimality criterion [28] for the one-parameter regularization schemes the values

$$\|x^\delta_{\alpha_i} - x^\delta_{\alpha_{i-1}}\| \quad \text{or} \quad \|x^\delta_{\beta_j,B} - x^\delta_{\beta_{j-1},B}\|$$

serve as "surrogates" of

$$\|x^\dagger - x^\delta_{\alpha_i}\| \quad \text{or} \quad \|x^\dagger - x^\delta_{\beta_j,B}\|,$$

one can select $\alpha = \alpha_{i_*} \in Q^\alpha_N,\ \beta = \beta_{j_*} \in P^\beta_M$ such that

$$\|x^\delta_{\alpha_{i_*}} - x^\delta_{\alpha_{i_*-1}}\| = \min\{\|x^\delta_{\alpha_i} - x^\delta_{\alpha_{i-1}}\|,\ i = 1, 2, \ldots, N\}$$

or

$$\|x^\delta_{\beta_{j_*},B} - x^\delta_{\beta_{j_*-1},B}\| = \min\{\|x^\delta_{\beta_j,B} - x^\delta_{\beta_{j-1},B}\|,\ j = 1, 2, \ldots, M\}.$$

Thus, we employ the idea of the quasi-optimality criterion to use the above-mentioned "surrogates" as the performance indicators and choose in the discrepancy domain the pairs (α_{i_*}, β_J) and (α_I, β_{j_*}) that allow the minimum distance between the regularized solutions corresponding to two successive values of the regularization parameters:

$$\begin{aligned} \|x^\delta_{\alpha_{i_*},\beta_J} - x^\delta_{\alpha_{i_*-1},\beta_J}\| &= \min\{\|x^\delta_{\alpha_i,\beta_j} - x^\delta_{\alpha_{i-1},\beta_j}\| :\ (\alpha_i, \beta_j) \in D(\delta),\ \beta_j > 1\}, \\ \|x^\delta_{\alpha_I,\beta_{j_*}} - x^\delta_{\alpha_I,\beta_{j_*-1}}\| &= \min\{\|x^\delta_{\alpha_i,\beta_j} - x^\delta_{\alpha_i,\beta_{j-1}}\| :\ (\alpha_i, \beta_j) \in D(\delta),\ \alpha_i > 1\}. \end{aligned}$$

Then, similar to the ideal case we choose the pair of the parameters as follows

$$(\alpha_+, \beta_+) = \begin{cases} (\alpha_{i_*}, \beta_J), \ \text{if } \|x^\delta_{\alpha_{i_*},\beta_J} - x^\delta_{\alpha_{i_*-1},\beta_J}\| \le \|x^\delta_{\alpha_I,\beta_{j_*}} - x^\delta_{\alpha_I,\beta_{j_*-1}}\|, \\ (\alpha_I, \beta_{j_*}), \ \text{if } \|x^\delta_{\alpha_{i_*},\beta_J} - x^\delta_{\alpha_{i_*-1},\beta_J}\| > \|x^\delta_{\alpha_I,\beta_{j_*}} - x^\delta_{\alpha_I,\beta_{j_*-1}}\|. \end{cases}$$

As one can see below, such combination of the discrepancy domain principle and the quasi-optimality criterion is a quite flexible tool for choosing the regularization parameters. Moreover, it is easily implementable and its computational efficiency is justified with the use of the model function approximation approach, presented in the previous section.

5.2. Numerical Illustrations and Comparisons: Operators with Known Singular Value Expansion. Similar to [1, 26] in our first numerical experiment we consider compact operators A and B^{-1} that are related as

$$A = \sum_i a_i \langle \psi_i, \cdot \rangle \varphi_i, \qquad B^{-1} = \sum_i b_i^{-1} \langle \psi_i, \cdot \rangle \psi_i, \tag{5.1}$$

where $\{a_i\}, \{b_i\}$ denote sets of eigenvalues of the self-adjoint operators $(A^*A)^{1/2}$ and B correspondingly.

Note that the knowledge of the singular value expansion of the operators allows us to verify easily whether the link condition (2.7) is violated or not.

In the first experiments the operators A and B are given as diagonal matrices of the size n. The matrix corresponding to the operator A has diagonal elements $a_i = i^{-r},\ i = 1, 2, \ldots, n,\ n = 50,\ r = 3$. Further, we assume that the source condition (2.5) is satisfied with $\varphi(t) = t^p,\ p = 4$, and the solution $x^\dagger$ is given in the form of the $n-$dimensional vector

$$x^\dagger = (A^* A)^4 g, \tag{5.2}$$

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 a random vector with uniformly distributed components. Both vectors g and ϵ are generated 100 times, so that we have 100 problems of the form (2.1) 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 (5.2) Tikhonov-Phillips regularization may suffer from the saturation. On the other hand, this effect can be relaxed by using Tikhonov regularization with a proper choice of the regularization operator B for which the condition (2.7) is satisfied. At first, we choose the self-adjoint operator B such that the corresponding diagonal matrix has the elements $b_{ii} = b_i = i,\ i = 1, 2, \ldots, n$. For the considered A, the chosen operator B satisfies (2.7) with $s = 3$. In the experiments, we use the discrepancy domain principle in combination with the quasi-optimality criterion for multi-penalty regularization and the classical discrepancy principle for the one-parameter counterparts. In all cases grids of the parameters are given by (4.10) with the initial parameters $\alpha_0 = \beta_0 = 10^{-4},\ q = p = 1.25$ and $N = M = 60$.

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 5.1, where each circle represents a relative error in solving the problems with one of 100 simulated data, for each of the three regularization methods: multi-penalty regularization (MP), Tikhonov-Phillips regularization (TP), and Tikhonov regularization (Tikhonov).

Moreover, in Table 5.1 the statistical measures such as the mean, the median, and the standard deviation of the relative error, as well as the mean values of the regularization parameters are given for each of the methods.

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

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

then from Figure 5.2 and Table 5.2 we can see that the saturation cannot be relaxed by Tikhonov method due to the fact that for the considered B the link condition

Table 5.1

Numerical illustration (first experiment). 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
$x^\delta_{\alpha,\beta}$	0.0069	0.0060	0.0036	$\alpha = 56.05, \beta = 0.002$
x^δ_α	0.0117	0.0114	0.0024	0.009
$x^\delta_{\beta,B}$	0.0086	0.086	0.0009	0.009

Table 5.2

Numerical illustration (second experiment). 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
$x^\delta_{\alpha,\beta}$	0.0117	0.0115	0.0029	$\alpha = 0.0064, \beta = 56.05$
x^δ_α	0.0117	0.0114	0.0022	0.009
$x^\delta_{\beta,B}$	0.0529	0.0535	0.0083	0.0116

(2.7) is violated ($\|B^{-s}\| \geq (n-1)^s \geq 1 \geq \|A\|$). At the same time, similar as it was observed in [26], we can see that multi-penalty regularization equipped with the discrepancy domain principle and the quasi-optimality criterion shows performances at the level of the best single-penalty regularization.

Note that for the operators (5.1) and the solutions (5.2) we can verify the convergence rates indicated in Theorem 3.1 and Theorem 3.3. In the first experiment, when $b_i = i$, the solutions belong to the set (2.8) with $p = 24$, which is above the saturation level $p = s + 2 = 5$ calculated according to Theorem 3.1. Therefore, in the considered situation the theorem guarantees us the accuracy of order $O(\delta^{\frac{s+2}{2s+2}}) = O(\delta^{5/8})$. In Table 5.3 we present the mean values of the numbers $C = \frac{\|x^\dagger - x^\delta_{\alpha,\beta}\|}{\delta^{5/8}}$ over 100 simulated solutions $x^\dagger$ of the form (5.2) for different noise levels δ and parameters α, β chosen according to the discrepancy domain principle. As it can be seen from the table the numbers exhibit a rather stable behaviour supporting the statement of Theorem 3.1. In the second experiment, the link condition (2.7) is violated and in view of the saturation of the discrepancy principle and Tikhonov-Phillips regularization itself, we can expect only the accuracy of order $O(\delta^{1/2})$. The behaviour of the mean values of the numbers $C = \frac{\|x^\dagger - x^\delta_{\alpha,\beta}\|}{\delta^{1/2}}$ for different δ is reflected in Table 5.4 and also can be accepted as an evidence supporting the theory.

5.3. Numerical Illustrations and Comparisons: First Kind Fredholm Integral Equations. In this subsection we are going to demonstrate that the compensatory property of multi-penalty regularization or even improvement in the performance can be observed in a more general case, when the singular value expansion of the operators is not known. In such a case it may be hard to check the link condition (2.7). On the other hand, as Figure 5.2 shows, Tikhonov regularization (2.6) alone may perform poorly if a link condition is not granted. From such perspective, multi-penalty scheme (2.9)-(2.11) can be seen as a tool to make the regularization more flexible and reliable.

Remark 5.1. *One may think that in the case of the failing link condition the need for a compensatory property arises from restricting to the Tikhonov type regularization, and the use of multi-penalty regularization can be avoided by switching*

TABLE 5.3
Numerical illustration (first experiment). Estimation of the constants from the error bound (3.8) for different noise levels

Noise level	$10^{-1.5}$	$10^{-1.8}$	$10^{-2.1}$	$10^{-2.4}$	$10^{-2.7}$	10^{-3}
Constant	0.2321	0.1860	0.1616	0.1347	0.1206	0.1190

TABLE 5.4
Numerical illustration (second experiment). Estimation of the constants from the error bound (3.3) for different noise levels

Noise level	$10^{-1.5}$	$10^{-1.8}$	$10^{-2.1}$	$10^{-2.4}$	$10^{-2.7}$	10^{-3}
Constant	0.2171	0.3067	0.4332	0.6119	0.8644	1.2210

to usual single-parameter iterative regularization schemes, such as iterated Tikhonov regularization [8, p. 123].

Note that such iterative schemes are also oriented towards smoothness expressed in the form (2.5), but in contrast to Tikhonov-Phillips regularization, the iterative schemes can guarantee an accuracy of order $O(\varphi(\theta_\psi^{-1}(\delta)))$ *even in the case when the function* $\frac{t}{\varphi(t)}$ *is decreasing (see, e.g. [16, p.82]).*

But the point is that for a given problem (2.1) a source condition (2.5) may not be an adequate form of measuring the smoothness of $x^\dagger$. *Then an iterative regularization scheme may be outperformed by multi-penalty regularization (2.9)-(2.11) even if it uses a penalizing operator* B *for which the link condition is not granted.*

We illustrate this on Figure 5.4, where iterated Tikhonov regularization of order 3 (IterTP) performs better than Tikhonov-Phillips regularization, but both of them are outperformed by multi-penalty regularization (2.9)-(2.11) and Tikhonov regularization (2.6) that are based on the penalizing operator for which we do not know whether the link condition is satisfied or not.

We would like to stress that the use of iterated Tikhonov regularization of order higher than 3 does not change the picture. The authors are grateful to the anonymous referee whose comment inspires this remark.

Similar to [17, 26] we generate the test problems of the form (2.1) 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

$$\int_a^b k(s,t)f(t)dt = g(s), \ s \in [a,b], \tag{5.3}$$

with a known solution $f(t)$. As in the 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 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 $B = (D^T D)^{1/2}$, where

$$D = \begin{pmatrix} 1 & -1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & & & 1 & -1 \end{pmatrix} \tag{5.4}$$

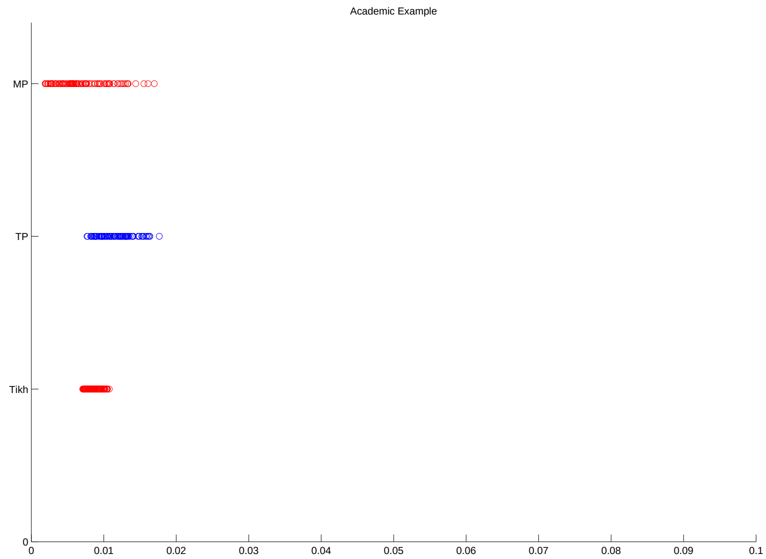

FIG. 5.1. *Numerical illustration (first experiment). The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise*

TABLE 5.5
Numerical illustration for the function shaw(100). 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
$x^\delta_{\alpha,\beta}$	0.0950	0.0924	0.0140	$\alpha = 2.7784, \beta = 0.0007$
x^δ_α	0.1736	0.1732	0.0029	0.0011
$x^\delta_{\beta,B}$	0.1774	0.1768	0.0063	0.0146

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

We perform the experiment with the function $shaw(n)$ that is a discretization of the equation (5.3) with $a = -\pi/2$ and $b = \pi/2$. The kernel and the solution are given as

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

The corresponding equation (5.3) is discretized by a simple quadrature with n equidistant points. Similar to [17, 26] we take $n = 100$. The results are displayed in Figure 5.3 and Table 5.5.

In the next 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 (5.3) with $a = 0$, $b = \infty$, $k(s,t) = e^{-st}$, $f(t) = e^{-t/2}$, $g(s) = (s + 1/2)^{-1}$.

For both problems the regularization parameters are chosen in accordance with

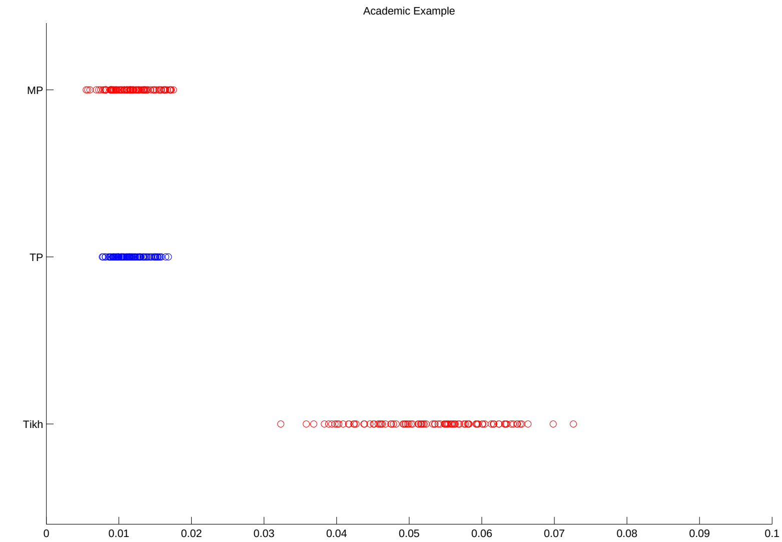

Fig. 5.2. *Numerical illustration (second experiment). The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise*

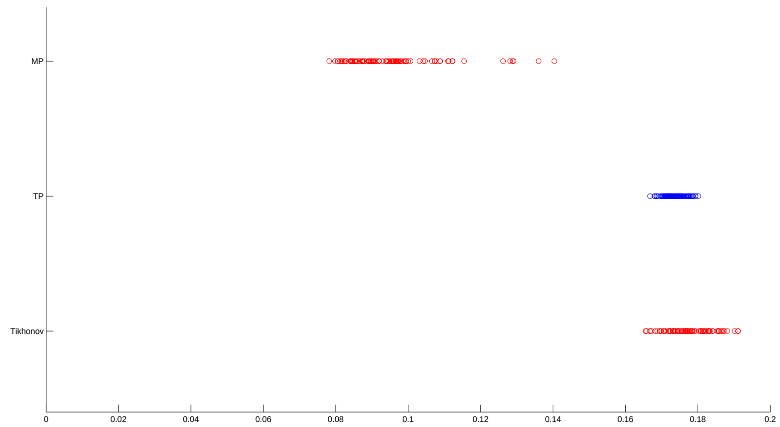

Fig. 5.3. *Numerical illustration for the function shaw*(100). *The figure presents relative errors (circles) for 100 simulations of y_δ with 1% noise*

the procedure described above, in which starting values of the parameter grids are given as $\alpha_0 = \beta_0 = 10^{-4},\ q = p = 1.3,\ N = M = 40$.

In Figure 5.4 we show the relative errors produced by the three regularization methods. Moreover, Table 5.6 presents a statistical information about the performance of the methods. Additionally, in Figure 5.4 one can see the relative errors produced by iterated Tikhonov regularization of order 3, as it has been discussed in Remark 5.1.

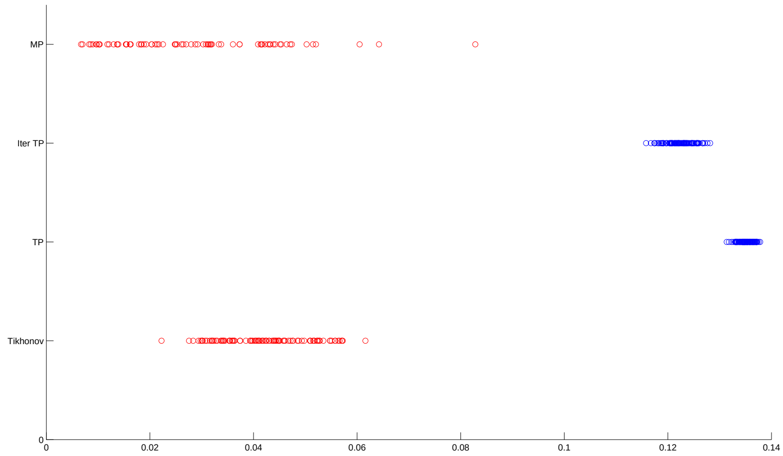

FIG. 5.4. *Numerical illustration for the function ilaplace*$(100, 1)$. *The figure presents relative errors (circles) for 100 simulations of* y_δ *with 1% noise*

TABLE 5.6
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
$x^\delta_{\alpha,\beta}$	0.0250	0.0209	0.0105	$\alpha = 2.7784, \beta = 0.0042$
x^δ_α	0.1323	0.1320	0.0005	0.0011
$x^\delta_{\beta,B}$	0.0421	0.0418	0.0027	0.0706

Again superior performances of multi-penalty regularization are observed even when it is not known a priori whether or not the link condition (2.7) is satisfied.

5.4. Three-penalty regularization. Motivated by some positive results with a three-parameter regularization [17], in this subsection we also consider the three-penalty regularization with a component wise penalization, in which the following form of the functional is considered

$$\Phi_1(\alpha,\beta,\gamma;u,v,z) := \|A(u+v+z) - y_\delta\|_Y^2 + \alpha\|u\|_X^2 + \beta\|Bv\|_X^2 + \gamma\|Cz\|_X^2. \tag{5.5}$$

Then the regularized approximation $x^\delta_{\alpha,\beta,\gamma}$ is given as $x^\delta_{\alpha,\beta,\gamma} = u^\delta_{\alpha,\beta,\gamma} + v^\delta_{\alpha,\beta,\gamma} + z^\delta_{\alpha,\beta,\gamma}$, where $u^\delta_{\alpha,\beta,\gamma}$, $v^\delta_{\alpha,\beta,\gamma}$, and $z^\delta_{\alpha,\beta,\gamma}$ are the minimizers of the functional (5.5). In order to write down the formulae for the minimizers explicitly, we, at first, introduce two linear operators

$$M = (\alpha\mathbb{I} + A^*A)^{-1}A^*A,$$
$$K = B^2(\beta B^2 + \alpha M)^{-1}.$$

Then the minimizers $u^\delta_{\alpha,\beta,\gamma}$, $v^\delta_{\alpha,\beta,\gamma}$, and $z^\delta_{\alpha,\beta,\gamma}$ of the functional $\Phi_1(\alpha,\beta,\gamma;u,v,z)$ have the form

TABLE 5.7

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

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^\delta_{\alpha,\beta,\gamma}$	0.0904	0.0872	0.0100	$\alpha = 2.39,\ \beta = 0.0001,\ \gamma = 1.4063$
x^δ_α	0.1732	0.1734	0.0019	0.0011
$x^\delta_{\gamma,C}$	0.2072	0.2070	0.0029	0.7850
$x^\delta_{\beta,B}$	0.1770	0.1766	0.0037	0.0146

$$
\begin{aligned}
u^\delta_{\alpha,\beta,\gamma} &= (\alpha\mathbb{I} + A^*A)^{-1}(A^*y_\delta - A^*A(v^\delta_{\alpha,\beta,\gamma} + z^\delta_{\alpha,\beta,\gamma})),\\
v^\delta_{\alpha,\beta,\gamma} &= \alpha(\beta B^2 + \alpha M)^{-1}(\alpha\mathbb{I} + A^*A)^{-1}(A^*y_\delta - z^\delta_{\alpha,\beta,\gamma}),\\
z^\delta_{\alpha,\beta,\gamma} &= (KM + \frac{\gamma}{\alpha\beta}C^2)^{-1}K(\alpha\mathbb{I} + A^*A)^{-1}A^*y_\delta.
\end{aligned}
$$

In our numerical experiments the operators $A,\ B$ are the same as before and the second penalizing operator $C = (\hat{D}^T\hat{D})^{1/2}$ is $n\times n-$matrix, where $(n-2)\times n-$matrix

$$
\hat{D} = \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & & 1 & -2 \end{pmatrix}
$$

is the discrete approximation to the second derivative operator on the regular grid with n points.

As in the previous subsection we consider the problems $shaw(100)$ and $ilaplace(100,1)$ with 1% of noise in data to compare performances of the three-parameter regularization $x^\delta_{\alpha,\beta,\gamma}$ and the standard single-parameter ones $x^\delta_\alpha,\ x^\delta_{\gamma,C}$, and $x^\delta_{\beta,B}$.

Recall that for the one-parameter schemes we employ as a parameter choice rule the discrepancy principle and its corresponding modification for multi-penalty method with starting parameters $\alpha_0 = \beta_0 = \gamma_0 = 10^{-4}$ and step-size $q = p = l = 1.7$ for $\alpha \in Q^\alpha_N,\ \beta \in P^\beta_M$ and

$$
\gamma \in S^\gamma_W = \{\gamma = \gamma_k = \gamma_0 l^k,\ k = 0,1,2,\dots,W\},\ l > 1.
$$

The results are displayed in Figures 5.5 and 5.6 in which the notation is similar to ones in the figures above. Note that the relative error corresponding to Tikhonov regularization with the penalizing operator C is denoted as "Tikhonov 2" in the respective figures. For the sake of completeness, we also indicate in Tables 5.7 and 5.8 the mean, the median, and the standard deviation of the relative errors, as well as the mean values of the regularization parameters for all four regularization methods under the consideration.

REMARK 5.2. *In the considered experiments three-penalty regularization performs worse than the two-penalty one. It can be explained by the fact that* $x^\delta_{\alpha,\beta,\gamma}$ *involves the penalizing operator* C *that produces the poorest regularization effect when it is used alone in* $x^\delta_{\gamma,C}$*. The multi-penalty scheme is still able to compensate this poor regularization, but, in this case, this compensation appears at a lower performance level.*

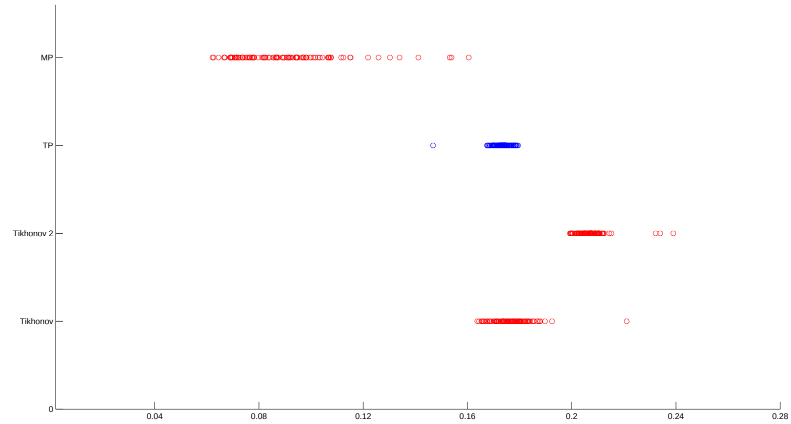

FIG. 5.5. *Numerical illustration for the function shaw*(100). *The figure presents relative errors (circles) for 100 simulations of* y_δ *with 1% noise*

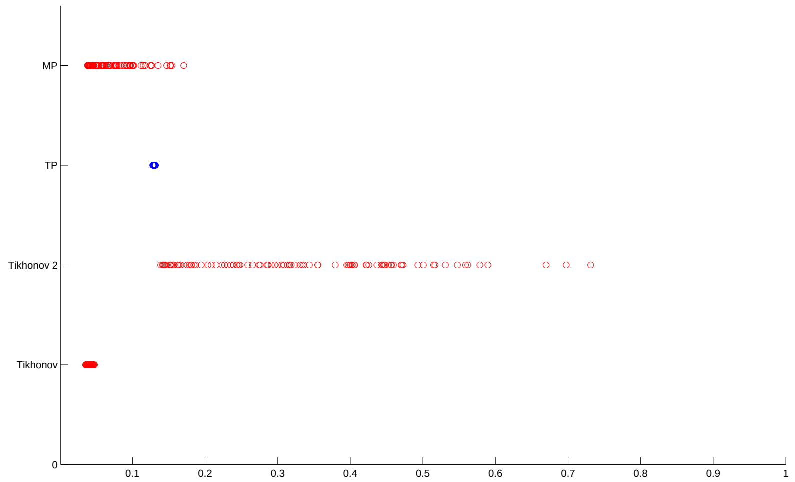

FIG. 5.6. *Numerical illustration for the function ilaplace*(100, 1). *The figure presents relative errors (circles) for 100 simulations of* y_δ *with 1% noise*

TABLE 5.8

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

	Mean RE	Median RE	Standard deviation RE	Mean parameter
$x^\delta_{\alpha,\beta,\gamma}$	0.0680	0.0577	0.0320	$\alpha = 2.81,\ \beta = 0.0001,\ \gamma = 1.65$
x^δ_α	0.1298	0.1298	0.0007	0.0008
$x^\delta_{\gamma,C}$	0.3261	0.3086	0.1427	0.0005
$x^\delta_{\beta,B}$	0.0406	0.0404	0.0028	0.0583

5.5. Four-parametric model function. In this subsection we demonstrate the adequacy of the four-parametric model function approximation using it as a tool for reconstructing the discrepancy domain. Recall that in view of Theorem 3.1 and Theorem 3.3 this domain $D(\delta)$ is of interest, because any point $(\alpha, \beta) \in D(\delta)$, where one of the parameters is sufficiently large, provides us with a solution $x^\delta_{\alpha,\beta} = u^\delta_{\alpha,\beta} + v^\delta_{\alpha,\beta}$, realizing an accuracy of optimal order.

In the previous subsection we have used the reconstruction of the discrepancy domain by a straightforward approach which consists in the direct calculation of the discrepancy $\|Ax^\delta_{\alpha_i,\beta_j} - y_\delta\|$ for all grid points

$$(\alpha, \beta) \in K = \{(\alpha_i, \beta_j):\ \alpha_i = \alpha_0 q^i \in Q^\alpha_N,\ \beta_j = \beta_0 p^j \in P^\beta_M,\\ i,\ j = 1, 2\ldots, 40,\ \alpha_0 = \beta_0 = 0.0001,\ q = p = 1.3\}.$$

Recall that we are interested in the parts of the domain where one of the parameters α or β is sufficiently large, i.e., greater than 1. Thus, we will reconstruct these parts of the domain separately by means of the model function approach. To this end, we distinguish two cases:

1. If $\alpha > 1$, then for each $\alpha_i \in Q^\alpha_N$ and $\alpha_i > 1$ the parameter $\beta_{k+1} = \beta_{k+1}(\alpha_i)$ is found by iterations (4.8) with an initialization $\alpha = \alpha_i,\ \beta_0 = 0.5$. The stopping criterion for the iteration (4.8) consists in checking that $\|Ax^\delta_{\alpha_i,\beta_{k+1}} - y_\delta\| \le c\delta$.
2. If $\beta > 1$, then for each $\beta_j \in P^\beta_M$ and $\beta_j > 1$, $\alpha_{k+1}(\beta_j)$ is found by (4.9) with an initial guess $\alpha_0 = 0.5$ and $\beta = \beta_j$. We terminate the iteration (4.9) when $\|Ax^\delta_{\alpha_{k+1},\beta_j} - y_\delta\| \le c\delta$.

Then in accordance with Theorem 3.1 and Theorem 3.3 the domain below the points $(\alpha_i, \beta_{k+1}(\alpha_i))$ for all $\alpha_i \ge 1$ and $(\alpha_{k+1}(\beta_j), \beta_j)$ with $\beta_j \ge 1$ can be seen as an approximate reconstruction of the discrepancy domain.

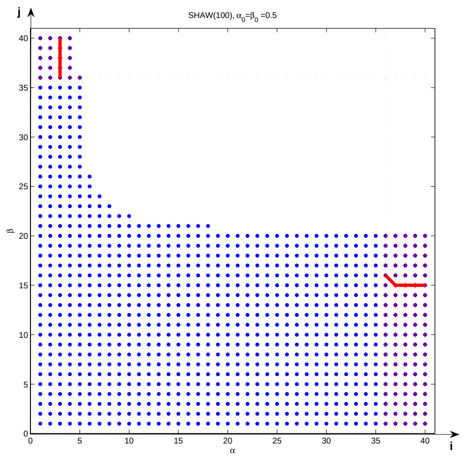

Fig. 5.7. *Discrepancy domain and its reconstruction by means of the model function for the function shaw*(100).

Figure 5.7 displays the reconstruction of the discrepancy domain for the problem *shaw*(100) with 1 % noise in data. Each dot corresponds to the indices (i, j) of the parameters α_i and β_j respectively such that $(\alpha_i, \beta_j) \in D(\delta)$, crosses indicate the indices of the parameters in which we are interested, namely $\alpha_i > 1,\ 35 < i \leq 40$ or $\beta_j > 1,\ 35 < j \leq 40$. The red lines represent the approximation of the boundary of the domains $D(\delta) \cap \{\alpha > 1\},\ D(\delta) \cap \{\beta > 1\}$ found by the model function approach. As we can see the domain reconstructed by means of the model function almost coincides with the one found by the direct calculation of the discrepancy for all grid points.

Acknowledgments. Massimo Fornasier acknowledges the support of the START Project "Sparse Approximation and Optimization in High-Dimensions." Valeriya Naumova and Sergei Pereverzyev are partially supported by the Austrian Fonds zur Frderung der Wissenschaftlichen Forschung (FWF), grant P25424 "Data-driven and problem-oriented choice of the regularization space."

REFERENCES

[1] F. Bauer and M. Lukas, *Comparing parameter choice methods for regularization of ill-posed problems*, Mathematics and Computers in Simulation, 81 (2011), pp. 1795–1841.

[2] F. Bauer and S. V. Pereverzev, *An utilization of a rough approximation of a noise covariance within the framework of multi-parameter regularization*, Int. J. Tomogr. Stat., 4 (2006), pp. 1–12.

[3] M. Belge, M. Kilmer, and E. Miller, *Efficient determination of multiple regularization parameters in a generalized L-curve framework*, Inverse Problems, 18 (2002), pp. 1161–1183.

[4] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu, *Multiparameter regularization techniques for ill-conditioned linear systems*, Numerische Mathematik, 94 (2003), pp. 203–228.

[5] Z. Chen, Y. Lu, Y. Xu, and H. Yang, *Multi-parameter Tikhonov regularization for linear ill-posed operator equations*, J. Comput. Math., 26 (2008), pp. 37–55.

[6] R. DeVore, *Optimal computation*, in International Congress of Mathematicians. Vol. I, Eur. Math. Soc., Zürich, 2007, pp. 187–215.

[7] D. Düvelmeyer and B. Hofmann, *A multi-parameter regularization approach for estimating parameters in jump diffusion processes*, J. Inverse Ill-Posed Probl., 14 (2006), pp. 861–880.

[8] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.

[9] M. Grasmair, *Multi-parameter Tikhonov regularisation in topological spaces*, 2011. ArXiv:1109.0364v1.

[10] P. Hansen, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[11] K. Ito, B. Jin, and T. Takeuchi, *Multi-parameter Tikhonov regularization*, Methods and Applications of Analysis, 18 (2011), pp. 31–46.

[12] K. Ito, B. Jin, and T. Takeuchi, *Multi-Parameter Tikhonov Regularization – An Augmented Approach*, ArXiv e-prints, (2013).

[13] K. Ito and K. Kunisch, *On the choice of the regularization parameter in nonlinear inverse problems*, SIAM J. Optim., 2 (1992), pp. 376–404.

[14] S. Kindermann and A. Neubauer, *On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization*, Inverse Problems and Imaging, 2 (2008), pp. 291–299.

[15] K. Kunisch and J. Zou, *Iterative choices of regularization parameters in linear inverse problems*, Inverse Problems, 14 (1998), pp. 1247–1264.

[16] S. Lu and S. Pereverzev, *Regularization Theory for ill-posed Problems, Selected topics*, DeGruyter, Berlin, Boston, 2013.

[17] S. Lu and S. V. Pereverzev, *Multi-parameter regularization and its numerical realization*, Numerische Mathematik, 118 (2011), pp. 1–31.

[18] S. Lu, S. V. Pereverzev, and U. Tautenhahn, *Dual regularized total least squares and multi-parameter regularization*, Comput. Methods Appl. Math., 8 (2008), pp. 253–262.

[19] Y. Lu, L. Shen, and Y. Xu, *Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors*, IEEE Transactions on Circuits and Systems I, 54 (2007), pp. 1788–1799.

[20] P. Mathé and B. Hofmann, *How general are general source conditions?*, Inverse Problems, 24, 015009 (2008), pp. 1–5.

[21] P. Mathé and S. V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), pp. 789–803.

[22] ———, *The discretized discrepancy principle under general source conditions*, Journal of Complexity, 22 (2006), pp. 371–381.

[23] V.A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.

[24] T. Nair, S. V. Pereverzev, and U. Tautenhahn, *Regularization in Hilbert scales under general smoothing conditions*, Inverse Problems, 21 (2005), pp. 1851–1869.

[25] F. Natterer, *Error bounds for Tikhonov regularization in Hilbert scales*, Appl. Anal., 18 (1984), pp. 29–37.

[26] V. Naumova and S. V. Pereverzev, *Multi-penalty regularization with a component-wise penalization*, Inverse Problems, 29, 075002 (2013), pp. 1–16.

[27] A. N. Tikhonov, *On the regularization of ill-posed problems*, vol. 153, Dokl. Akad. Nauk SSSR, 1963.

[28] A. N. Tikhonov and V. B. Glasko, *Use of the regularization methods in non-linear problems*,

vol. 5, USSR Comput. Math. Phys., 1965.

[29] J. Xie and J. Zou, *An improved model function method for choosing regularization parameters in linear inverse problems*, Inverse Problems, 18 (2002), pp. 631–643.

[30] P. Xu, Y. Fukuda, and Y.M. Liu, *Multiple parameter regularization: Numerical solutions and applications to the determination of geopotential from precise satellite orbits*, Journal of Geodesy, 80 (2006), pp. 17–27.