Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Complexity 22 (2006) 371–381

Journal of COMPLEXITY

www.elsevier.com/locate/jco

The discretized discrepancy principle under general source conditions

Peter Mathé[a,*], Sergei V. Pereverzev[b]

[a] *Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D-10117 Berlin, Germany*

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

Received 22 December 2004; accepted 21 June 2005

Available online 7 March 2006

Abstract

We discuss adaptive strategies for choosing regularization parameters in Tikhonov–Phillips regularization of discretized linear operator equations. Two rules turn out to be based entirely on data from the underlying regularization scheme. Among them, only the discrepancy principle allows us to search for the optimal regularization parameter from the easiest problem. This potential advantage cannot be achieved by the standard projection scheme. We present a modified scheme, in which the discretization level varies with the successive regularization parameters, which has the advantage, mentioned before.

Keywords: Discrepancy principle; Projection method; General source condition

1. Introduction, problem formulation

An operator equation

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

with a compact operator A acting between Hilbert spaces X and Y is essentially ill-posed if the range $R(A)$ of A is not closed in Y. If A is invertible this non-closed range yields the discontinuity of the inverse operator A^{-1}. In general, the minimum-norm solution x_0 of problem (1) does not depend continuously on the right-hand side y. Moreover, the data will not be available exactly in practice, because of measurement errors. One has to be aware of numerical instabilities when

* Corresponding author.

E-mail addresses: mathe@wias-berlin.de (P. Mathé), sergei.pereverzyev@oeaw.ac.at (S.V. Pereverzev).

0885-064X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.

doi:10.1016/j.jco.2005.06.008

noisy observations $y_\delta \in Y$ with

$$\|y - y_\delta\|_Y \leqslant \delta \tag{2}$$

are known instead of y. Hence, in order to approximate the unknown true solution x_0 in a stable way, regularization methods should be applied.

One of the most widely used methods for solving ill-posed problems is Tikhonov–Phillips regularization. Here a regularized approximation $x_{\alpha,\delta}$ is obtained by minimizing the functional

$$J_\alpha(x) = J_\alpha(A, y_\delta; x) = \|Ax - y_\delta\|_Y^2 + \alpha\|x\|_X^2,$$

so that the respective regularized solution is given by

$$x_{\alpha,\delta} = x_{\alpha,\delta}(A) = (\alpha I + A^*A)^{-1}A^*y_\delta. \tag{3}$$

It is well-known that the accuracy of the regularized solution $x_{\alpha,\delta}$ depends critically on the chosen regularization parameter $\alpha > 0$. If α is too small, then the regularization operator $(\alpha I + A^*A)^{-1}A^*$ is too close to the (unbounded) Moore–Penrose generalized inverse operator $(A^*A)^{-1}A^*$, i.e., the procedure for obtaining $x_{\alpha,\delta}$ is not so stable, and if α is too large, then fine features of the true solution x_0 are smoothed out. Therefore, it is important to have a reliable rule for choosing the regularization parameter appropriately.

1.1. Parameter-choice strategies

Of course, it would be nice to choose α a priori, because this would allow us to solve the regularized equation

$$\alpha x + A^*Ax = A^*y_\delta \tag{4}$$

for $x = x_{\alpha,\delta}$, only once. However, the appropriate regularization parameter reflects smoothness properties of the underlying true solution, and such a priori parameter choice can rarely be used in practice.

Therefore rules have been developed, which attempt to determine the regularization parameter α, depending upon the noise level δ and based on the data y_δ. These include the discrepancy principle (DP), originally proposed by Phillips [13] and later analyzed in more detail by Morozov [11] and Marti [8]; a method developed by Gfrerer and Engl [3] and also by Raus [15], which is sometimes called the minimum-bound (MB) method, see [6]; the monotone error rule (ME), as proposed by Tautenhahn and Hämarik [16]. One can find a good account of these rules in [4]. Finally we mention a general adaptation strategy (GAS), as presented by the authors in [9].

In principle, all these rules can be combined with Tikhonov–Phillips regularization (3), but a disadvantage of the MB and ME rules is that they require the knowledge of an additional approximate solution, obtained by some more sophisticated regularization method. For example, the ME rule selects the regularization parameter for Tikhonov–Phillips regularization by constructing an additional approximate solution using iterated Tikhonov regularization, i.e., another regularization method. This is not reasonable, since it doubles computational cost.

Both the DP and GAS are free from the drawback of solving additional problems. The intrinsic difference lies in the direction in which we search for the parameter α. To compare these rules we restrict ourselves to the situation in which the regularization parameter α is selected from some finite set, called Δ_M, below. For parameters from Δ_M we compute regularized solutions $x_{\alpha_i,\delta}$, $\alpha_i \in \Delta_M$, and select the regularization according to some principle among $\alpha_i \in \Delta_M$.

Following Tikhonov and Glasko [17], we take this set Δ_M as a geometric sequence $\Delta_M := \{\alpha_i = q^i\alpha_0,\ i = 1, 2, \ldots, M\}$, with $q \in (0, 1)$ and M chosen so that $\alpha_0 q^M \sim \delta^2$, i.e. $M \sim \log(1/\delta)$.

The DP starts with large α_0, and decreases stepwise $\alpha_i = q\alpha_{i-1}$, $i = 1, 2, \ldots$, until $\alpha_* := \alpha_j$ is the first parameter that obeys

$$\|Ax_{\alpha_j,\delta} - y_\delta\|_Y \leqslant \tau\delta, \tag{5}$$

where the design parameter τ is at least two.

The GAS operates in the opposite direction. Here we start with the smallest $\alpha_M \sim \delta^2$ and increase stepwise $\alpha_{i-1} = \alpha_i/q$, $i = M, M-1, \ldots$, until $\alpha_+ := \alpha_k$ with

$$\|x_{\alpha_k,\delta} - x_{\alpha_{k+1},\delta}\| > \frac{\tau\delta}{\sqrt{\alpha_{k+1}}}, \tag{6}$$

for the first time.

Thus, when applying the GAS we start with the hardest regularized problem (4), while the DP allows us to begin with the easiest one, and move to harder problems only if necessary.

However, it is known that the DP does not provide the best order of accuracy for all ill-posed problems (1), for which Tikhonov–Phillips regularization allows us to obtain the best order of reconstruction. Specifically, the best possible error of the Tikhonov–Phillips method is $\mathcal{O}(\delta^{2/3})$, while in combination with the DP one can achieve the accuracy $\mathcal{O}(\delta^{1/2})$, at best. To our knowledge, the GAS, when applied to Tikhonov–Phillips regularization, is the only rule that allows us to reach the best order of accuracy for problems that in principle can be treated in an optimal way by Tikhonov–Phillips regularization. We refer the reader to the discussion in [9] for further details.

In many cases one knows that an order of accuracy better than $\mathcal{O}(\delta^{1/2})$ cannot be reached for the problem of interest; for example, this is true for the so called severely ill-posed problems see [9, Remark 9]. Then the use of the DP within the framework of Tikhonov–Phillips regularization is reasonable.

1.2. The issue of discretization

When claiming that the regularized problem (4) with smaller α is harder than (4) for larger one, we still consider the ideal situation, in which the solution to (3) can be computed. In practice, any numerical realization of the Tikhonov–Phillips scheme requires us to turn to some finite dimensional approximation B of A. As it has been explained above, for each fixed B the realization of DP is easier than GAS, because in the latter case one should treat a linear system with the largest condition number from the very beginning. At the same time, it should be realized that any parameter choice strategy applied to the equation with B will lead to the regularization of this finite dimensional problem. But we are interested in the solution of the original infinite dimensional problem. To achieve this goal one should relate the level of regularization with the level of approximation. We argue that standard discretization schemes do not allow to realize this potential advantage of DP over GAS. It can be achieved within non-standard adaptive discretization scheme presented below. The outline of this study is as follows. We first introduce the *discretized discrepancy* in the way as this is commonly done, see e.g. [7]. Then we analyze the discretized DP, applied to Tikhonov–Phillips regularization when smoothness is measured in terms of *general source conditions*. We end up with some sufficient conditions on the closeness of the approximating B to the original A, which ensure the optimal order of reconstruction. These

conditions allow us to establish a non-standard scheme, based on sets of nested projections, originally introduced by one of the authors in [12]. Based on spars(B), a complexity measure known as sparsity, we finally discuss the superiority of our new scheme over the usual schemes that are analyzed in (e.g.) [14,10].

2. The discretized discrepancy principle for Tikhonov regularization

We shall measure the smoothness of the solution $x = x_0$ relative to the operator A in terms of

$$A_\varphi(R) := \left\{x \in X : \ x = \varphi(A^*A)v \quad \textit{for some } v \in X \text{ with } \|v\| \leqslant R\right\}.$$

Here $\varphi : (0, \|A^*A\|] \to \mathbb{R}^+$ is increasing and $\varphi(0+) = 0$. Such functions are called *index functions*. These give rise to weighted Hilbert spaces X_φ as follows: the operator A^*A admits a (monotonic) Schmidt representation for an orthonormal system $u_1, u_1, \ldots$, given by

$$A^*Ax = \sum_{j=1}^{\infty} s_j \langle x, u_j\rangle u_j, \quad x \in X.$$

Then X_φ is the completion of finite expansions $x = \sum_{j=1}^{n} \langle x, u_j\rangle u_j$ with respect to the scalar product

$$\langle x, y\rangle_\varphi := \sum_{j=1}^{\infty} \frac{\langle x, u_j\rangle\langle y, u_j\rangle}{\varphi^2(s_j)}.$$

In this case we have $A_\varphi(R) = \left\{x \in X \ : \ \|x\|_\varphi \leqslant R\right\}$. For details of this construction and consequences with respect to approximation we refer the reader to [9,10].

As discussed above, regularization will be carried out by solving the approximate equation

$$x_{\alpha,\delta}(B) = (\alpha I + B^*B)^{-1}B^*y_\delta, \tag{7}$$

instead of the original one.

Let $g_\alpha(t) := (\alpha+t)^{-1}$ and $r_\alpha(t) := 1-tg_\alpha(t)$. Using the spectral calculus, the functions g_α and r_α may be extended to non-negative self-adjoint operators in X. In these terms we can see that by construction (7) of $x_{\alpha,\delta}$ the discretized discrepancy can be rewritten as $\|Bg_\alpha(B^*B)B^*y_\delta - y_\delta\| = \|r_\alpha(BB^*)y_\delta\|$. The respective ($B$-discretized) DP reads

$$\alpha_* := \max\left\{\alpha \in \Delta_M, \ \|Bg_\alpha(B^*B)B^*y_\delta - y_\delta\| \leqslant \tau\delta\right\}. \tag{8}$$

Our subsequent analysis will be based on the following assumptions on the closeness of A and B.

Basic assumptions. The operators $B = B(\alpha)$ are chosen in such a way that

$$\|AA^* - BB^*\| \leqslant \delta\sqrt{\alpha}, \tag{9}$$

$$\|A^*A - B^*B\| \leqslant \delta\sqrt{\alpha} \tag{10}$$

and

$$\|(A^* - B^*)A\| \leqslant \delta\sqrt{\alpha}. \tag{11}$$

We recall the following facts, see e.g. [7].

Lemma 1. *For any $\alpha > 0$ we have*

$$\| \left(r_\alpha(AA^*) - r_\alpha(BB^*)\right) y\| \leqslant R\frac{\|AA^* - BB^*\|}{2\sqrt{\alpha}} \tag{12}$$

and

$$\| \left(r_\alpha(A^*A) - r_\alpha(B^*B)\right) x_0\| \leqslant R\frac{\|A^*A - B^*B\|}{\alpha}. \tag{13}$$

With these facts at hand we can derive our first bounds.

Lemma 2. *Under the DP* (8), *and using assumptions* (9) *and* (10) *it holds true that*

$$q\,(\tau - 1 - R/2)\,\delta \leqslant \|r_{\alpha_*}(AA^*)y\| \leqslant (1 + \tau + R/2)\delta. \tag{14}$$

Proof. Using the triangle inequality, we find that

$$\begin{aligned}\|r_\alpha(AA^*)y\| &\leqslant \|r_\alpha(BB^*)y\| + \|r_\alpha(AA^*)y - r_\alpha(BB^*)y\| \\ &\leqslant \|r_\alpha(BB^*)(y - y_\delta)\| + \|r_\alpha(BB^*)y_\delta\| + \|r_\alpha(AA^*)y - r_\alpha(BB^*)y\|.\end{aligned}$$

The first summand is bounded by δ. By (8), the middle summand is bounded by $\tau\delta$ for $\alpha = \alpha_*$. Using (12) the third one is bounded by $(R/2\sqrt{\alpha})\|A^*A - B^*B\|$, such that using (10) we obtain the upper bound.

To prove the lower estimate let $\alpha := \alpha_*/q \in \Delta_M$. Using the triangle inequality and (8) we find that

$$\begin{aligned}\|r_\alpha(AA^*)y\| &\geqslant \|r_\alpha(BB^*)y\| - \| \left(r_\alpha(BB^*) - r_\alpha(AA^*)\right) y\| \\ &\geqslant \|r_\alpha(BB^*)y_\delta\| - \|r_\alpha(BB^*)(y - y_\delta)\| - \| \left(r_\alpha(BB^*) - r_\alpha(AA^*)\right) y\| \\ &\geqslant (\tau - 1)\delta - R\frac{\|AA^* - BB^*\|}{2\sqrt{\alpha}}.\end{aligned}$$

By (9), this implies $\|r_\alpha(AA^*)y\| \geqslant (\tau - 1 - R/2)\delta$. Next, an easy calculation shows that for Tikhonov regularization $\|r_{\alpha_*}(AA^*)y\| \geqslant q\|r_\alpha(AA^*)y\|$, from which the proof can be completed. □

With these preparations we can provide the first impact of the DP; it provides a lower bound on the parameter α_*. The important function in this context is $\Theta(t) = \sqrt{t}\varphi(t)$, $t > 0$.

Proposition 3. *Suppose the true solution x_0 belongs to $A_\varphi(R)$ for some index function φ for which φ^2 is concave. If α_* is chosen as in* (8) *then*

$$\Theta(\alpha_*) \geqslant q\,(\tau - 1 - R/2)\,\delta/R.$$

Proof. If $x_0 \in A_\varphi(R)$, then there exists $v \in X$ satisfying $\|v\| \leqslant R$ with $x_0 = \varphi(A^*A)v$. If φ^2 is concave, then $\varphi^2(t)/t$ is decreasing and so is $\sqrt{t}\varphi(t)/t$. Then from Remark 5 and Proposition 3 of [9, Remark 5 & Prop. 3] it follows that

$$\begin{aligned}\|r_{\alpha_*}(AA^*)y\| &= \|r_{\alpha_*}(AA^*)A\varphi(A^*A)v\| \\ &= \|r_{\alpha_*}(AA^*)(A^*A)^{1/2}\varphi(A^*A)v\| \leqslant R\Theta(\alpha_*).\end{aligned}$$

Using Lemma 2 the proof can be completed. □

Next we shall see that we can bound the difference between the exact and the regularized solutions for pure data.

Corollary 4. *Let $x = x_0 \in A_\varphi(R)$. Then*

$$\|r_{\alpha_*}(A^*A)x\| \leqslant (1+\tau+R/2)R\varphi(\Theta^{-1}(\delta/R)). \tag{15}$$

Proof. For $x \in A_\varphi(R)$, we have $\|r_\alpha(A^*A)x\|_\varphi \leqslant R$. By Lemma 2, we obtain

$$\|r_{\alpha_*}(A^*A)x\|_{1/\sqrt{t}} = \|Ar_{\alpha_*}(A^*A)x\| = \|r_{\alpha_*}(AA^*)y\| \leqslant (1+\tau+R/2)\delta.$$

Because φ^2 is concave, the same holds true for $t \mapsto \varphi^2((\Theta^2)^{-1}(t))$, see e.g. [9], and we can apply the *interpolation inequality*

$$\varphi^{-1}\left(\frac{\|r_\alpha(A^*A)x\|_{\varphi/\varphi}}{\|r_\alpha(A^*A)x\|_\varphi}\right) \leqslant \Theta^{-1}\left(\frac{\|r_\alpha(A^*A)x\|_{\varphi/\Theta}}{\|r_\alpha(A^*A)x\|_\varphi}\right)$$

of [10, Theorem 4]. Using the concavity of $t \mapsto \varphi(\Theta^{-1}(t))$ this leads to the estimate (15). □

Now we are ready to state and prove the main result of this section.

Theorem 5. *Suppose $x = x_0 \in A_\varphi(R)$ for an index function φ with concave square, and that α_* is chosen according to the DP. There is a constant $C < \infty$ for which*

$$\|x - x_{\alpha_*,\delta}\| \leqslant CR\varphi(\Theta^{-1}(\delta/R)). \tag{16}$$

Proof. We start with the obvious error decomposition

$$\begin{aligned}\|x - x_{\alpha_*,\delta}\| &\leqslant \|x - g_{\alpha_*}(B^*B)B^*Ax\| + \delta\|g_{\alpha_*}(B^*B)B^*\| \\ &\leqslant \|\left(\mathrm{I} - g_{\alpha_*}(B^*B)B^*A\right)x\| + \frac{\delta}{\sqrt{\alpha_*}}.\end{aligned}$$

We bound the first summand on the right as

$$\begin{aligned}&\|\left(\mathrm{I} - g_{\alpha_*}(B^*B)B^*A\right)x\| \\ &\quad\leqslant \|\left(\mathrm{I} - g_{\alpha_*}(A^*A)A^*A\right)x\| + \|\left(g_{\alpha_*}(B^*B)B^* - g_{\alpha_*}(A^*A)A^*\right)y\|.\end{aligned}$$

The first summand can be rewritten as $\|r_{\alpha_*}(A^*A)x\|$ and was estimated in Corollary 4. The second summand can be treated similar to [7, estim. (28)]. Recall that the inequalities (9)–(11) hold for $B = B(\alpha)$. For $\alpha = \alpha_*$ we obtain

$$\begin{aligned}&\|\left(g_{\alpha_*}(B^*B)B^* - g_{\alpha_*}(A^*A)A^*\right)y\| \\ &\quad\leqslant \|\left(\alpha_*\mathrm{I} + B^*B\right)^{-1}(B^*B - A^*A)\left(\alpha_*\mathrm{I} + A^*A\right)^{-1}A^*Ax)\| \\ &\qquad + \|\left(\alpha_*\mathrm{I} + B^*B\right)^{-1}(A^* - B^*)y\| \\ &\quad\leqslant {\alpha_*}^{-1}\|A^*A - B^*B\|\|x\| + {\alpha_*}^{-1}\|(A^* - B^*)A\|\|x\| \leqslant 2R\frac{\delta}{\sqrt{\alpha_*}}.\end{aligned}$$

Thus from Proposition 3 and Corollary 4, we deduce that for some large enough $C < \infty$

$$\|x - x_{\alpha_*,\delta}\| \leqslant (1 + \tau + R/2)R\varphi(\Theta^{-1}(\delta/R)) + (2R+1)\frac{\delta}{\sqrt{\alpha_*}}$$
$$\leqslant CR\varphi(\Theta^{-1}(\delta/R)).$$

The last estimate holds true because $\varphi(\Theta^{-1}(2t)) \leqslant 2\varphi(\Theta^{-1}(t))$, which follows from concavity. □

From [9, Corollary 1] one knows that $\delta \mapsto \varphi(\Theta^{-1}(\delta/R))$ is the best possible order of accuracy in the case under consideration. Thus, Theorem 5 tells that under assumptions (9)–(11), the discretized DP provides us with the order-optimal choice of regularization parameter. Let us stress that for index functions $\varphi(t) = t^\nu$, the requirement that φ^2 is concave restricts the best possible order of accuracy to $\sqrt{\delta}$, because necessarily $0 < \nu \leqslant \frac{1}{2}$, and $\varphi(\Theta^{-1}(t)) = t^{\nu/(\nu+1/2)} \geqslant \sqrt{t}$, $t \in [0, 1]$. In view of this limitation there is no reason to apply regularization schemes, which are more sophisticated than Tikhonov–Phillips regularization, because even the latter already allows to achieve that accuracy.

Remark 6. It is well known (see, e.g. [9, Theorem 4]) that Tikhonov–Phillips regularization cannot provide an optimal order of accuracy for $\alpha > \delta^{2/3}$. Therefore it is reasonable to choose α_0 for Δ_M as $\alpha_0 = \delta^{2/3}$. Then for $\alpha \in \Delta_M$, we have $\alpha^2 < \delta\sqrt{\alpha}$, and hence condition (10) is weaker than the requirement $\|A^*A - B^*B\| \sim \alpha^2$, a condition presented by Groetsch in [5], i.e., the discretization according to (10) can be more efficient.

3. Complexity issues

3.1. Approximation by projections

Several authors have investigated schemes where the approximation B to A is of the form $B = Q_k A P_l$ and where Q_k and P_l are orthogonal projections onto suitable finite dimensional subspaces $R(Q_k) \subset Y$ and $R(P_l) \subset X$, respectively. These are called *projection schemes*.

In the spirit of multi-resolution analysis, see e.g. [1], we will assume that the spaces $R(Q_k)$ and $R(P_l)$ belong to some families $\{R(Q_j)\}_{j=0}^\infty$ and $\{R(P_j)\}_{j=0}^\infty$ of nested finite dimensional subspaces, i.e., we have the inclusions

$$R(Q_j) \subset R(Q_{j+1}) \subset Y, \quad R(P_j) \subset R(P_{j+1}) \subset X$$

and for some $s \geqslant 1$ we have

$$\dim R(Q_j) \sim \dim R(P_j) \sim 2^{sj}, \tag{17}$$

where $a \sim b$ always means that a and b can be bounded by constant multiples of each other.

As it has been observed in [14] and [10], the error bounds of a standard projection scheme depend on the bounds on $\|A(I - P_l)\|$ and on $\|(I - Q_k)A\|$ and it is natural to assume that there exist $r, c_r, d_r > 0$ such that for $k, l = 0, 1, \ldots,$ we have

$$\|A(I - P_l)\|_{X\to Y} \leqslant c_r 2^{-rl} \quad \text{and} \quad \|(I - Q_k)A\|_{X\to Y} \leqslant d_r 2^{-rk}. \tag{18}$$

Example 7. An example to illustrate the assumptions above is given by an integral operator

$$Ax(t) = \int_\Gamma a(t,\tau)x(\tau)\,\mathrm{d}\tau, \quad t \in \Gamma,$$

with kernel $a(t,\tau)$ belonging to the Sobolev space $W^r(\Gamma \times \Gamma)$, where Γ is a (smooth or at least piecewise smooth) manifold. In this case it is easy to see that both operators, A and

$$A^*x(t) = \int_\Gamma a(\tau,t)x(\tau)\,\mathrm{d}\tau$$

act from $L_2(\Gamma)$ into the Sobolev space $W^r(\Gamma)$. Then nested finite dimensional subspaces meeting (17), (18) can be constructed using wavelets as presented in [2].

3.2. Sparsity as a measure of complexity

We shall measure the complexity by the amount of discrete information required to compute B, or (more specifically) by the number of matrix–vector multiplications required to compute the solution to (7), based on the operator B. Both the quantities mentioned above are determined by the number of non-zero entries in the matrix representing the finite dimensional operator B in corresponding bases. This number is known as the *sparsity* of B, see [7]. We will denote the sparsity of B by spars(B) and we use this as a complexity measure.

3.3. Sparsity of the standard projection scheme

The DP (8) with $x_{\alpha_j,\delta} = x_{\alpha_j,\delta}(B)$ and $B = Q_k A P_l$ instead of A has been studied in detail in [14]. The authors have shown that a discretized version of the DP provides an order optimal a posteriori parameter choice for Tikhonov–Phillips regularization (7), in which $B = Q_k A P_l$, whenever

$$\|A(I - P_l)\|_{X\to Y} \sim \|(I - Q_k)A\|_{X\to Y} \sim \delta. \tag{19}$$

Under assumptions (18) this can be guaranteed by letting $k = l$ and $2^k = 2^l \sim \delta^{-1/r}$. In view of (17) such a choice in general leads to the finite dimensional operator $B = Q_k A P_l$ having the sparsity

$$\mathrm{spars}(Q_k A P_l) \sim \delta^{-2s/r}. \tag{20}$$

Rule (6), applied to the same approximation $x_{\alpha_j,\delta} = x_{\alpha_j,\delta}(Q_k A P_l)$, has been studied in [10]. Again, it was shown that under (19), the GAS will provide an order optimal a posteriori parameter choice for all problems (1), which in principle can be treated by Tikhonov–Phillips regularization optimally. Moreover, in [10] an adaptive α-dependent discretization strategy has been proposed. For $\alpha > \alpha_M \sim \delta^2$ one can use finite dimensional operators $B = Q_k A P_l$, sparser than (20).

Summarizing the above discussion, we see that these standard projection schemes do not allow us to utilize the main advantage of the DP, which is that one starts with the easiest regularized problem. In both cases (the DP, as well as the GAS) we start with some operator $B = Q_k A P_l$ of the same sparsity (20).

3.4. Sparsity of nested discretization

It is the goal of the present section to analyze a modified projection scheme under the DP that will really allow us to start with the easiest problem (in the sense of sparsity of B, as well as in the sense of condition number of the matrix associated with $\alpha I + B^*B$), and that will increase the complexity only if necessary for increasing the accuracy.

The analysis from the previous section does not indicate how to choose the finite dimensional operators $B = B(\alpha)$ meeting the basic assumptions (9)—(11). Next we are going to present such a recipe.

Let us consider the discretization

$$B = B_n = \sum_{j=1}^{2n} (Q_j - Q_{j-1}) A P_{2n-j} + Q_0 A P_{2n}. \tag{21}$$

This was introduced in [12, § 17] for discretizing operator equations of the second kind (well-posed problems). In [7] this discretization was also applied for linear ill-posed problems (1) under the assumptions that $x_0 \in A_\varphi(R)$ for $\varphi(t) = t^\nu$, $\quad 0 < \nu \leqslant \frac{1}{2}$. Our analysis uses the following result that can be proved similar to [7, Lemma 1].

Lemma 8. *Under assumptions* (17) *and* (18) *let* $b_r := 2^{r+3} d_r c_r$.

If $n = n(\alpha, \delta)$ *is the minimal integer number satisfying the inequality*

$$n(\alpha, \delta) 2^{-2n(\alpha,\delta)r} \leqslant b_r^{-1} \delta \sqrt{\alpha}, \tag{22}$$

then the basic assumptions (9)–(11) *are fulfilled.*

Moreover, $\operatorname{spars}(B_n) \sim 2^{2ns} n$.

As a consequence we obtain

Corollary 9. *Let* $n := n(\alpha, \delta)$ *be chosen as in* (22). *For* α_* *chosen according to the DP, we have*

$$\operatorname{spars}(B_{n(\alpha_*,\delta)}) \leqslant c(\delta \sqrt{\Theta^{-1}(\delta)})^{-s/r} \log^{1+s/r}(1/(\delta \sqrt{\Theta^{-1}(\delta)})). \tag{23}$$

Proof. For such a choice of $n(\alpha, \delta)$, we obtain

$$\operatorname{spars}(B_{n(\alpha,\delta)}) \sim (\delta \sqrt{\alpha})^{-s/r} \log^{1+s/r}(1/\delta \sqrt{\alpha}).$$

Recall that the projections are nested as $R(Q_j) \subset R(Q_{j+1})$ and $R(P_j) \subset R(P_{j+1})$. Then for $\alpha_i < \alpha_j$ (with $\alpha_i, \alpha_j \in \Delta_M$)

$$R(B_{n(\alpha_j,\delta)}) \subset R(B_{n(\alpha_i,\delta)}) \quad \text{and} \quad R(B^*_{n(\alpha_j,\delta)}) \subset R(B^*_{n(\alpha_i,\delta)}).$$

This implies that the matrix representing the operator $B_{n(\alpha_j,\delta)}$ in appropriate bases of the subspaces $R(Q_{2n(\alpha_j,\delta)})$, $R(P_{2n(\alpha_j,\delta)})$ is part of the matrix representing $B_{n(\alpha_i,\delta)}$. Using this sequence of operators $B_{n(\alpha_i,\delta)}$, $\alpha_i \in \Delta_M$, within the DP (8), one starts with the matrix representing $B_{n(\alpha_0,\delta)}$ to finally end up with the matrix representing $B_{n(\alpha_*,\delta)}$. This resulting matrix contains all previous ones. Moreover, Proposition 3 and Lemma 8 allow us to complete the proof. □

4. Conclusion

It is illuminating to compare the number of non-zero entries (23) of the final matrix obtained by the nested discretization with the corresponding number, obtained when using a projection scheme with an α-dependent choice of discretization. The latter scheme was discussed in [10]. From the analysis there, and within the present setup, see (17) and (18), the number of non-zero entries at the end is of the order $\left(\Theta^{-1}(\delta)\right)^{-s/r}$. This number can be seen to be substantially smaller than the one given in (23). However, this small matrix appears at the end after a sequence of computations, where already at the beginning a problem has to solved, which is more involved than the final thus hardest one under nested discretization, which has sparsity (23). Thus, for both regularization parameter choice rules under discussion the standard discretization strategy, known also as a Galerkin scheme, leads to the finite dimensional operator having the sparsity (20). The only difference is that the combination of Galerkin scheme with the DP keeps the same matrix during the whole search of the final regularization parameter, whereas in the case of GAS the most populated matrix is used only at the beginning of the search. At the same time, Corollary 9 suggests α-dependent nested discretization scheme (21), which allows to implement the DP in such a way that the corresponding matrix becomes more populated only if necessary, but even at the end of the search the final matrix will be much more sparse than a Galerkin matrix.

Specifically, let us assume that the index function φ measuring the solution smoothness obeys $\varphi(t) < t^{\nu}$ for some $\nu > 0$, i.e., the underlying problem is moderately ill-posed. This yields for any $\mu > 0$ the estimate

$$\log^{-\mu}(1/\delta) \gg (\delta^2 \log^{2\mu}(1/\delta))^{\nu} > \varphi(\delta^2 \log^{2\mu}(1/\delta)),$$

or equivalently, $\delta \gg \Theta(\delta^2 \log^{2\mu}(1/\delta))$.[1] This in turn yields

$$\delta^{-2s/r} \gg c(\delta\sqrt{\Theta^{-1}(\delta)})^{-s/r} \log^{1+s/r}(1/(\delta\sqrt{\Theta^{-1}(\delta)})) > \mathrm{spars}(B_{n(\alpha_*,\delta)}).$$

Comparing this with (20) we conclude that for moderately ill-posed problems the combination of the DP with finite dimensional approximation (21) is much more efficient than the DP combined with the standard projection scheme. To the best of our knowledge the discretization (21) is the only one which allows to utilize the main advantage of the DP, which is that we start with the easiest regularized problem.

Acknowledgment

This study started when the first named author visited the Johann Radon Institute, Linz. He wants to express his gratitude for the excellent conditions and generous support.

References

[1] W. Dahmen, Wavelet and multiscale methods for operator equations, in: Acta numerica, 1997, Acta Numerica, vol. 6, Cambridge University Press, Cambridge, 1997, pp. 55–228.
[2] W. Dahmen, R. Schneider, Wavelets on manifolds. I. Construction and domain decomposition, SIAM J. Math. Anal. 31 (1) (1999) 184–230.
[3] H.W. Engl, H. Gfrerer, A posteriori parameter choice for general regularization methods for solving linear ill-posed problems, Appl. Numer. Math. 4 (5) (1988) 395–417.

[1] The notation $a(t) \gg b(t)$ means that $a(t)/b(t) \to \infty$ for $t \to 0$.

[4] H.W. Engl, M. Hanke, A. Neubauer, Regularization of inverse problems, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers, Dordrecht, 1996.
[5] C.W. Groetsch, Convergence analysis of a regularized degenerate kernel method for Fredholm integral equations of the first kind, Integral Equations Operator Theory 13 (1) (1990) 67–75.
[6] M.A. Lukas, Comparisons of parameter choice methods for regularization with discrete noisy data, Inverse Problems 14 (1) (1998) 161–184.
[7] P. Maaß, S.V. Pereverzev, R. Ramlau, S.G. Solodky, An adaptive discretization for Tikhonov–Phillips regularization with a posteriori parameter selection, Numer. Math. 87 (3) (2001) 485–502.
[8] J.T. Marti, An algorithm for computing minimum norm solutions of Fredholm integral equations of the first kind, SIAM J. Numer. Anal. 15 (6) (1978) 1071–1076.
[9] P. Mathé, S.V. Pereverzev, Geometry of linear ill-posed problems in variable Hilbert scales, Inverse Problems 19 (3) (2003) 789–803.
[10] P. Mathé, S.V. Pereverzev, Discretization strategy for linear ill-posed problems in variable Hilbert scales, Inverse Problems 19 (6) (2003) 1263–1277.
[11] V.A. Morozov, On the solution of functional equations by the method of regularization, Soviet Math. Dokl. 7 (1966) 414–417.
[12] S.V. Pereverzev, Optimization of methods for approximate solution of operator equations, Computational Mathematics and Analysis Series, Nova Science Publishers Inc., Commack, NY, 1996.
[13] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comput. Mach. 9 (1962) 84–97.
[14] R. Plato, G. Vainikko, On the regularization of projection methods for solving ill-posed problems, Numer. Math. 57 (1) (1990) 63–79.
[15] T. Raus, The principle of the residual in the solution of ill-posed problems, Tartu Riikl. Ül. Toimetised (672) (1984) 16–26.
[16] U. Tautenhahn, U. Hämarik, The use of monotonicity for choosing the regularization parameter in ill-posed problems, Inverse Problems 15 (6) (1999) 1487–1505.
[17] A.N. Tikhonov, V.B. Glasko, An approximate solution of Fredholm integral equations of the first kind, Ž. Vyčisl. Mat. i Mat. Fiz. 4 (1964) 564–571.