An adaptive discretization for Tikhonov-Phillips regularization with a posteriori parameter selection — Source link

Peter Maaß, Sergei V. Pereverzev, Ronny Ramlau, S. G. Solodky

Institutions: University of Bremen

Published on: 01 Jan 2001 - Numerische Mathematik (Springer-Verlag)

Topics: Regularization (mathematics) and Tikhonov regularization

Universität Potsdam

Peter Maaß, Sergei V. Pereverzev, Ronny Ramlau,
Sergei G. Solodky

An adaptive discretization for Tikhonov-Phillips regularization with a posteriori parameter selection

NLD Preprints ; 48

An adaptive discretization for Tikhonov-Phillips regularization with a posteriori parameter selection

Peter Maaß* Sergei V.Pereverzev Ronny Ramlau
Sergei G.Solodky

Abstract

The aim of this paper is to describe an efficient adaptive strategy for discretizing ill-posed linear operator equations of the first kind: we consider Tikhonov-Phillips regularization

$$x_\alpha^\delta \;=\; (A^*A+\alpha I)^{-1}\,A^*y^\delta$$

with a finite dimensional approximation A_n instead of A. We propose a sparse matrix structure which still leads to optimal convergences rates but requires substantially less scalar products for computing A_n compared with standard methods.

1 Introduction

The aim of this paper is to describe an adaptive strategy for discretizing ill-posed linear operator equations of the first kind

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

We assume that only perturbed data y^δ with $\| y - y^\delta \|$ is available. More precisely, we consider Tikhonov-Phillips regularization

$$x_\alpha^\delta \;=\; (A^*A+\alpha I)^{-1}\,A^*y^\delta \tag{2}$$

*This work has been partially supported by the DFG under grant MA 1657/1–2

where the regularization parameter α is chosen according to Morozov's discrepancy principle, i.e. one determines the largest $\alpha \in \{\alpha_m = q^m\alpha_0 | m = 0, 1, ..\}$ s.t.

$$||Ax_\alpha^\delta - y^\delta|| \ \leq d\delta \ .$$

Any numerical realisation requires to carry out all computations with a finite dimensional approximation A_n instead of A.
The choice of the approximation A_n determines the accuracy as well as the overall complexitiy of the algorithm. The complexity of the algorithm has to be measured in two categories: 1) the number of scalar products required to compute A_n, 2) the number of matrix–vector–multiplications –weighted by the number of non–zero entries of A_n– required to compute (2).
Several authors have investigated approximations of the type

$$A_n \ = \ QAP \ ,$$

where P and Q denote orthogonal projections onto suitable finite dimensional subspaces. These investigations aim at minimizing the dimensions of the subspaces. E.g. [16] treats discretizations of this type for general projection methods and [12] investigates discretizations for Tikhonov regularization. A recent publication [7] also exploits the strutcure of the resulting linear systems for different values of α in order to construct efficient CG-methods for solving (2). However, all these publications link the level of approximation to the data error δ only, i.e. A_n is kept fixed for all α which have to be tested. A first adaptive strategy, which linked the level of approximation to the value of α_m by

$$||A_n - A|| \leq c\sqrt{\alpha}\sqrt{\delta} \ \ ,$$

was suggested in [10]. There the approximations were obtained by wavelet techniques in order to obtain sparse approximations. This allows for an efficient computation of the matrix-vector multiplications needed for determining $x_{\alpha_m}^\delta$.
From a numerical point of view an efficient algorithm for computing the approximation A_n is equally important. This efficiency (or complexity of type 1) is measured by the number of scalar products required to compute A_n. This number is called the required amount of discrete information.

If A_n is obtained from a 2^n–dimensional approximation space

$$V_n \ = \ span\{\varphi_j \mid j = 1, 2, .., 2^n \ \} \ ,$$

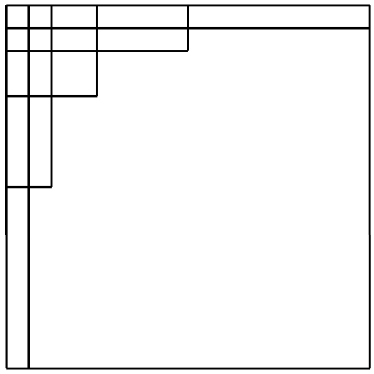


Figure 1: The marked area determines the coefficients of A_n for $n = 2$ which have to be computed in the adaptive discretization scheme. The dimnension of this matrix is $2^{2n} = 16$.

then standard methods require the evaluation of 2^{2n} scalar products $< \varphi_i, A\varphi_j >$ for computing A_n. This is the case even for wavelet–approximations if no further assumptions on the structure of A, e.g. convolution operator or weakly singular integral operators with known degree of singularity [2, 3], can be exploited.
The computation of such scalar products would be cheap if a singular value or a wavelet–vaguelette decomposition of A [5, 4] would be available. However the computation of these decompositions is in general as costly as solving 2^n linear systems of dimension 2^n.
An elegant and efficient approximation was proposed in [14]. The number n has to be even in this approach, i.e. the dimensional index is replaced by $2n$:

$$A_{disc} = A_n = \sum_{j=1}^{2n} (P_j - P_{j-1}) A P_{2n-j} + P_0 A P_{2n} \quad , \tag{3}$$

where P_j denotes the orthogonal projection onto a suitable 2^j – dimensional subspace. This idea amounts to computing only a small fraction of the coefficients of the matrix A_n.

The idea of this paper is to combine the approaches in [14] and [10], ie. to combine the discretization (3) with an adative strategy for choosing n depending on δ and α. This strategy still yields optimal convergence rates but requires substantially less scalar products for computing the discretized operators, i.e. these methods require a smaller amount of discrete information.

2 Notation and basic assumptions

Denote by $(\cdot,\cdot)$ the inner product for some Hilbert space H and as usual, $\|\varphi\|_H = (\varphi,\varphi)^{1/2}$. For $r \in (0,\infty)$, we let H^r denote a linear subspace of H which is equipped with a norm $\|\varphi\|_{H^r} \geq \|\varphi\|_H$. Moreover, in the spirit of wavelets updating [3] we suppose that there exists a sequence of nested finite dimensional subspaces $V_0 \subset V_1 \subset V_2 \subset \ldots \subset V_\ell \subset \ldots \subset H$ such that

$$\inf_{v_j \in V_j} \|v - v_j\|_H \leq c_r 2^{-rj} \|v\|_{H^r}, \quad v \in H^r, \tag{4}$$

and

$$dim V_j \sim 2^{sj}, \tag{5}$$

where $s \geq 1$ and $a \sim b$ will always mean that a and b can be bounded by constant multiples of each other. The condition (4) can be written in the form

$$\|I - P_j\|_{H^r \to H} \leq c_r 2^{-rj}, \tag{6}$$

with a constant $c_r \geq 1$ and where $P_j : H \to V_j$ is the orthogonal projection onto subspace V_j, i.e.

$$P_j v = \sum_{k=1}^{dim V_j} (v, \varphi_{j,k}) \varphi_{j,k}.$$

Here $\Phi_j = \{\varphi_{j,k} : k = 1, 2, \ldots, dim V_j\}$ denotes an orthonormal basis of V_j, $\Phi_{j-1} \subset \Phi_j$.

Now we define the class of equations (1) which will be considered in the sequel. First of all it will be assumed that the operators A have some smoothness with respect to the family of subspaces H^r, $r \in (0,\infty)$. Namely,

$$A \in \mathcal{H}^r_\gamma = \{A : \|A\|_{H \to H^r} + \|A^*\|_{H \to H^r} \leq \gamma\}, \quad \gamma \geq 1. \tag{7}$$

Here A^* denotes the adjoint operator of $A : H \to H$, i.e. $(f, Ag) = (A^* f, g)$ for any $f, g \in H$.

Now we present a rather simple example to illustrate the assumptions described above. We consider an integral equation

$$Ax(t) := \int_0^1 k(t,\tau) x(\tau) d\tau = y(t). \tag{8}$$

The underlying spaces and projections are chosen as follows. As Hilbert space H we take the space $L_2(0,1)$ with the usual norm and inner product.

As H^r we take the Sobolev space $W_2^r(0,1)$ of functions $f(t)$ having square-summable derivatives $f^{(i)} \in L_2(0,1),\ i = 0,1,\dots,r$, with an appropriate norm. For $r = 1$ as nested finite dimensional subspaces we choose the spaces of piecewise constant functions, such that

$$V_0 = span\{1\}, \quad V_j = span\{\varphi_{j,k},\ \ k = 1,2,\dots,2^j\},$$

$$dim V_j = 2^j.$$

Here $\{\varphi_{j,k}\}$ is the orthonormal basis of Haar-wavelet functions, where $\varphi_{j,1} = 1$ and for $k = 2^{m-1} + i,\ \ m = 1,2,\dots,j, i = 1,2,\dots,2^{m-1}$

$$\varphi_{j,k}(t) = \begin{cases} 2^{(m-1)/2}, & t \in [(i-1)/2^{m-1}, (i-1/2)/2^{m-1}) \\ -2^{(m-1)/2}, & t \in [(i-1/2)/2^{m-1}, i/2^{m-1}) \\ 0, & elsewhere \end{cases}$$

It is well-known that for such spaces V_j and for the orthonormal projection $P_j : L_2(0,1) \to V_j$ we have

$$||I - P_j||_{W_2^1(0,1)\to L_2(0,1)} \leq c_1 2^{-j}.$$

This means that for $H = L_2(0,1),\ \ H^r = W_2^r(0,1)$ and for spaces of Haar-wavelet functions V_j the conditions (4)–(6) hold for $r = 1, s = 1$. As indicated in [3], for example, one can construct wavelets on the interval $[0,1]$ in such a manner that (5), (6) hold for $H = L_2(0,1),\ \ H^r = W_2^r(0,1),\ \ r > 1, s \geq 1$. If the kernel $k(t,\tau)$ of the integral operator of (8) has square-summable partial derivatives $\frac{\partial^i k(t,\tau)}{\partial t^i}, \frac{\partial^i k(t,\tau)}{\partial \tau^i},\ \ i = 0,1,\dots,r$, then it is easy to see that $A \in \mathcal{H}_\gamma^r$ for $H = L_2(0,1),\ \ H^r = W_2^r(0,1)$ and some γ.

We shall study the equation (1) with $A \in \mathcal{H}_\gamma^r$. On the other hand, from the condition (7) one sees that $A \in \mathcal{H}_\gamma^r$ is a compact linear operator acting from H to H and so it is not continuously invertible. In this setting the problem (1) is ill-posed, that is, its minimum norm solution $x^\dagger$ does not depend continuously on the right-hand side y. Small perturbation y_δ of the exact but unknown data y may cause dramatic changes in $x^\dagger$.

The usual discussion of the order of accuracy of solution techniques for (1) is done under the assumptions that the minimum norm solution $x^\dagger$ lies in the range of $(A^*A)^\nu, \nu > 0$, that is

$$x^\dagger = (A^*A)^\nu v, \quad \|v\|_H \leq \rho, \tag{9}$$

and the perturbed data y_δ satisfies $\|y - y_\delta\|_H \leq \delta$ with an a priori known noise level $\delta > 0$. From [17], [18] it follows that under these assumptions for any solution technique the best possible order of accuracy in the power scale is $\delta^{2\nu/(2\nu+1)}$. Therefore in the sequel we shall consider the class $\Phi^{r,\nu}_{\gamma,\rho}$ of equations (1) with $A \in \mathcal{H}^r_\gamma$ and

$$y \in AM_{\nu,\rho}(A) = \{f : \quad f = Au, \quad u \in M_{\nu,\rho}(A)\},$$

where $M_{\nu,\rho}(A) = \{u : \ u = (A^*A)^\nu v, \ \|v\|_H \leq \rho\}$. It is easy to see that in this case the solution $x^\dagger$ of (1) has the form (9).

3 Morozov's discrepancy principle for the standard projection methods with a predetermined level of discretization

Traditionally the discretization of problem (1) is done by a Galerkin method. This means that instead of (1) we consider now the equation

$$A_{disc}x = P_m y_\delta, \tag{10}$$

where $A_{disc} = P_m A P_n$. But if (1) is ill-posed, i.e., the solution x is not a continuous function of the data y, regularization techniques are required for solving (10). In this paper we consider Tikhonov-Phillips regularization. In this method a solution of (10), and hence of (1), is approximated by a solution $x_{\alpha,m,n}$ of the equation

$$\alpha x + A^*_{disc}A_{disc}x = A^*_{disc}y_\delta. \tag{11}$$

Note that finding on element $x_{\alpha,m,n}$ reduces to solving a system of $\min\{dimV_m, dimV_n\}$ linear algebraic equations.

One of the most widely used strategies for choosing the regularization parameter α is Morozov's discrepancy principle [11]. Following [16], we shall consider this discrepancy principle in a form tailored to the discretized version of Tikhonov-Phillips regularization and $A \in \mathcal{H}^r_\gamma$: Let $1 < d_1 \leq d_2$. If

$\|P_m y_\delta\|_H \leq d_1\delta$, then take $x = 0$ as the approximation. If $\|P_m y_\delta\|_H > d_1\delta$, then choose $\alpha \geq \alpha_{\min} = (\gamma c_r 2^{-nr})^2$ such that

$$d_1\delta \leq \|P_m y_\delta - A_{disc} x_{\alpha,m,n}\|_H \leq d_2\delta. \tag{12}$$

If there is no $\alpha \geq \alpha_{\min}$ such that (12) holds, then choose $\alpha = \alpha_{\min}$. The following theorem allows us to estimate the efficiency of the traditional approach to discretization (10)–(12).

Theorem 1 [16]. *Let the parameter α be chosen according to the discrepancy principle (12). If equation (1) belongs to the class $\Phi^{r,\nu}_{\gamma,\rho}, 0 < \nu \leq 1/2$, then*

$$\|x^\dagger - x_{\alpha,m,n}\|_H \leq d_\nu \left\{\delta^{\frac{2\nu}{2\nu+1}} + 2^{-2\nu nr} + 2^{-2\nu mr}\right\},$$

where d_ν is independent of δ, n, m.

Let us consider the following situation. We know that equation (1) belongs to the class $\Phi^{r,\nu}_{\gamma,\rho}$ for some $\nu \in (0, 1/2]$, but we don't know the exact value of ν. From Theorem 1 it follows that – within the framework of the standard projection methods (10)–(12) and with a predetermined level of discretization – we can guarantee the optimal order of accuracy $\delta^{2\nu/(2\nu+1)}$ in the case when $2^n \geq \delta^{-1/(2\nu+1)r}, 2^m \geq \delta^{-1/(2\nu+1)r}$ for all $\nu \in (0, 1/2]$. As with [16], the minimal m and n satisfying above conditions for all $\nu \in (0, 1/2]$ have the following orders: $2^n \sim 2^m \sim \delta^{-1/r}$.

Denote by $card(IP)$ the number of inner products of the form

$$(\varphi_{m,k}, A\varphi_{n,k}) \quad \text{and} \quad (\varphi_{m,k}, y_\delta), \tag{13}$$

required to construct an approximate solution $x_{\alpha,m,n}$ realizing the optimal order of accuracy for all $\nu \in (0, 1/2]$. Then by virtue of (5)

$$card(IP) = dimV_m(dimV_n + 1) \sim 2^{s(m+n)} \sim \delta^{-2s/r}. \tag{14}$$

4 An adaptive discretization scheme

So far we have discussed to which extend A may be replaced by a discretized operator A_{disc}, where A_{disc} is kept fixed for all possible values of the regularization parameter α. However since we choose α by testing different values of the regularization parameter we would like to link the amount of discrete information $card(IP)$ to α. This will allow us to use coarser discretizations for large values of α and to obtain the optimal order of accuracy in the power scale $\delta^{2\nu/(2\nu+1)}$ using an amount of discrete information of the form (13) such that

$$card(IP) \sim (\delta\sqrt{\alpha})^{-s/r} \log^{1+s/r}(\delta\sqrt{\alpha})^{-1}.$$

Let us consider the discretization scheme within the framework of which

$$\begin{aligned} A_{disc} = A_n = \sum_{j=1}^{2n} (P_j - P_{j-1}) A P_{2n-j} + P_0 A P_{2n} = \\ = \sum_{j=1}^{2n} P_{2n-j} A (P_j - P_{j-1}) + P_{2n} A P_0. \end{aligned} \tag{15}$$

Note that this scheme was used earlier in [15], p.295, for discretizing second kind operator equations (well-posed problems). This discretization uses a discretization space of dimension 2^{2n} but computes only a small fraction of the scalar products required to compute the standard discretization $P_{2n}AP_{2n}$. To be more precise, this approximation incorporates the full discretization P_mAP_m only for $m = n$ and adds some coefficients describing the mixing of high and low frequency components by the action of A.

In the sequel we need the lemma

Lemma 1 *For* $A \in \mathcal{H}_\gamma^r$

$$\|A^*A - A_n^*A_n\|_{H\to H} \le c_{r,\gamma} n 2^{-2nr}, \quad \|AA^* - A_nA_n^*\|_{H\to H} \le c_{r,\gamma} n 2^{-2nr},$$

$$\|(A^* - A_n^*)A\|_{H\to H} \le c_{r,\gamma} n 2^{-2nr},$$

where $c_{r,\gamma} = 2^{r+3} c_r^2 \gamma^2$.

Proof. We prove only the first estimate. Other estimates are established in a similar manner.

It is easy to see that

$$\begin{aligned} \|A^*A - A_n^*A_n\|_{H\to H} \le \|A^*(I - P_{2n})^2 A\|_{H\to H} + \\ + \|A^*P_{2n}A - A_n^*A_n\|_{H\to H}. \end{aligned} \tag{16}$$

Keeping in mind that

$$A_n^* A_n = \sum_{j=1}^{2n} P_{2n-j} A^* (P_j - P_{j-1}) A P_{2n-j} + P_{2n} A^* P_0 A P_{2n}, \tag{17}$$

we have

$$\begin{gathered} \|A^* P_{2n} A - A_n^* A_n\|_{H\to H} \le \|A^* P_0 A - P_{2n} A^* P_0 A P_{2n}\|_{H\to H} + \\ + \sum_{j=1}^{2n} \|G_j\|_{H\to H}, \end{gathered} \tag{18}$$

where

$$G_j = A^*(P_j - P_{j-1})A - P_{2n-j} A^* (P_j - P_{j-1}) A P_{2n-j}.$$

It is easy to calculate that

$$\|G_j\|_{H\to H} \le \|(I - P_{2n-j}) A^* (P_j - P_{j-1}) A\|_{H\to H} +$$

$$+ \|P_{2n-j} A^* (P_j - P_{j-1}) A (I - P_{2n-j})\|_{H\to H} \le$$

$$\le 2\|I - P_{2n-j}\|_{H^r\to H} \|A^*\|_{H\to H^r} \|P_j - P_{j-1}\|_{H^r\to H} \|A\|_{H\to H^r} \le$$

$$\le 2c_r^2 \gamma^2 2^{-(2n-j)r} \left(2^{-jr} + 2^{-(j-1)r}\right) = 2c_r^2 \gamma^2 (1 + 2^r) 2^{-2nr},$$

$$\|A^* P_0 A - P_{2n} A^* P_0 A P_{2n}\|_{H\to H} \le 2c_r \gamma^2 2^{-2nr},$$

$$\|A^* (I - P_{2n})^2 A\|_{H\to H} \le c_r^2 \gamma^2 2^{-4nr}.$$

Then by virtue of (16), (18) we obtain the assertion of the lemma.

Let us study the approximation properties of Tikhonov-Phillips algorithm with a parameter selection according to discrepancy principle which has the following form:

1. given data: $A \in \mathcal{H}_\gamma^r, y_\delta, \delta$;
2. initialization: $\alpha_0, \ 0 < q < 1, \ d > 2$;
3. iterate

(a) $\alpha = \alpha_m = q^m \alpha_0$,

(b) determine a discretization level $n = n(\alpha, \delta)$ such that

$$n2^{-2nr} = 4c^{-1}_{r,\gamma,\rho}\delta\sqrt{\alpha_m}, \tag{19}$$

(c) compute the inner products

$$(\varphi_{2n,i}, y_\delta), \quad i = 1, 2, \ldots, dim V_{2n}, \tag{20}$$

(d) compute the inner products

$$(\varphi_{j,k}, A\varphi_{2n-j,\ell}), \qquad \begin{array}{c} j = 1, 2, \ldots, 2n, \\ k = 1, 2, \ldots, dim V_j, \\ \ell = 1, 2, \ldots, dim V_{2n-j} \end{array} \tag{21}$$

required to construct A_n,

(e) compute $x^\delta_{\alpha_m,n} = (\alpha_m I + A^*_n A_n)^{-1} A^*_n y_\delta$ by solving

$$\alpha_m x + A^*_n A_n x = A^*_n y_\delta \tag{22}$$

until

$$\|A_n x^\delta_{\alpha_m,n} - y_\delta\|_H \leq d\delta. \tag{23}$$

As we will see in the following, this variant of Tikhonov-Phillips algorithm insures the best possible order of accuracy in the power scale.

Denote by $\{u_i, v_i, \sigma_i\}$ the singular value decomposition for a compact operator A, where $u_i, v_i \in H$ are the singular vectors and $\sigma_i > 0$ are the singular values.

Lemma 2 *Assume that $A \in \mathcal{H}^r_\gamma$ and that $x^\dagger$ obeys (9). Then for $\nu \in (0, 1/2]$*

$$\|x^\dagger - x^\delta_{\alpha,n}\|_H \leq \frac{\delta}{2\sqrt{\alpha}} + \alpha^\nu c_{\nu,\alpha}(v) + \frac{c_{r,\gamma,\rho} n 2^{-2nr}}{\alpha},$$

where $c_{r,\gamma,\rho} = 2c_{r,\gamma}\gamma\rho$,

$$c^2_{\nu,\alpha}(v) = \sum_{i \geq 0} \left\{ \frac{\alpha^{1-\nu}\sigma_i^{2\nu}}{(\sigma_i^2 + \alpha)}(v, u_i) \right\}^2 \leq \left\{(1-\nu)^{1-\nu}\nu^\nu\rho\right\}^2.$$

Proof. We follow the proof of Lemma 2.5 in [12]. First of all we note that

$$\|x^\dagger - x_{\alpha,n}^\delta\|_H \leq \|x^\dagger - (\alpha I + A^*A)^{-1}A^*y\|_H +$$
$$+\|(\alpha I + A_n^*A_n)^{-1}A_n^*(y - y_\delta)\|_H + \tag{24}$$
$$+\|(\alpha I + A^*A)^{-1}A^*y - (\alpha I + A_n^*A_n)^{-1}A_n^*y\|_H.$$

Equation (9) and inserting the singular value decomposition yields

$$\|x^\dagger - (\alpha I + A^*A)^{-1}A^*y\|_H =$$
$$= \|(\alpha I + A^*A)^{-1}(\alpha x^\dagger + A^*Ax^\dagger - A^*y)\|_H = \tag{25}$$
$$= \alpha\|(\alpha I + A^*A)^{-1}x^\dagger\|_H \leq c_{\nu,\alpha}(v)\alpha^\nu.$$

Moreover, from standard estimates using the singular value decomposition of compact operator T one knows:

$$\|(\alpha I + T^*T)^{-1}\|_{H\to H} \leq \alpha^{-1}, \quad \|(\alpha I + T^*T)^{-1}T^*\|_{H\to H} \leq \tfrac{1}{2\sqrt{\alpha}},$$
$$\|(\alpha I + T^*T)^{-1}T^*T\|_{H\to H} \leq 1, \tag{26}$$

where $T = A, A^*$ or $T = A_n, A_n^*$. Then

$$\|(\alpha I + A_n^*A_n)^{-1}A_n^*(y - y_\delta)\|_H \leq \frac{\delta}{2\sqrt{\alpha}}. \tag{27}$$

On the other hand, from (9), (26) and Lemma 1 we know that

$$\|(\alpha I + A^*A)^{-1}A^*y - (\alpha I + A_n^*A_n)^{-1}A_n^*y\|_H \leq$$
$$\leq \|(\alpha I + A_n^*A_n)^{-1}(A_n^*A_n - A^*A)(\alpha I + A^*A)^{-1}A^*y\|_H +$$
$$+\|(\alpha I + A_n^*A_n)^{-1}(A^* - A_n^*)y\|_H \leq$$
$$\leq \alpha^{-1}\|A_n^*A_n - A^*A\|_{H\to H}\|(\alpha I + A^*A)^{-1}A^*Ax^\dagger\|_H + \tag{28}$$
$$+\alpha^{-1}\|(A^* - A_n^*)Ax^\dagger\|_H \leq 2\alpha^{-1}c_{r,\gamma}n2^{-2nr}\|x^\dagger\|_H \leq$$
$$\leq 2c_{r,\gamma}\gamma^{2\nu}\rho\alpha^{-1}n2^{-2nr} \leq c_{r,\gamma,\rho}\tfrac{n2^{-2nr}}{\alpha}.$$

The assertion of the lemma follows from (24)–(27).

We now continue to analyse the convergence properties of the proposed adaptive scheme by following the standard lines of proof; i.e. we first show that the proposed stopping criterion with noisy data yields a regularization parameter α which would have been also obtained by a related discrepancy principle with perfect data.

Lemma 3 *If $\alpha = \alpha_N$ and $n = n(\alpha_N, \delta)$ are chosen within the framework of algorithm (19)–(23) for $A \in \mathcal{H}_\gamma^r$ then there exist $d_1, d_2 > 0$ such that*

$$d_1\delta \le \|Ax_\alpha - y\|_H \le d_2\delta,$$

where $x_\alpha = (\alpha I + A^*A)^{-1}A^*y$.

Proof. We follow the proof of Lemma 7 and Lemma 10 in [10].

We put $R_\alpha(T) = (\alpha I + T^*T)^{-1}T^*$ and $x_\alpha^{\delta,n} = R_\alpha(A_n)y^\delta$. Then

$$Ax_\alpha - y = AR_\alpha(A)y - y =$$

$$(AR_\alpha(A) - A_nR_\alpha(A_n))y + (I - A_nR_\alpha(A_n))(y - y_\delta) + (A_nx_{\alpha,n}^\delta - y_\delta). \tag{29}$$

Keeping in mind that

$$T(\alpha I + T^*T)^{-1} = (\alpha I + TT^*)^{-1}T,$$

$$T^*(\alpha I + TT^*)^{-1} = (\alpha I + T^*T)^{-1}T^*,$$

we have

$$(AR_\alpha(A) - A_nR_\alpha(A_n))y = \left(AA^*(\alpha I + AA^*)^{-1} - A_nA_n^*(\alpha I + A_nA_n^*)^{-1}\right)y$$

$$= (AA^* - A_nA_n^*)(\alpha I + AA^*)^{-1}y - A_nA_n^*(\alpha I + A_nA_n^*)^{-1}(AA^* - A_nA_n^*)(\alpha I + AA^*)^{-1}y$$

$$= (\alpha I + A_nA_n^*)^{-1}\left\{(\alpha I + A_nA_n^*) - A_nA_n^*\right\}(AA^* - A_nA_n^*)(\alpha I + AA^*)^{-1}y$$

$$= \alpha(\alpha I + A_nA_n^*)^{-1}(AA^* - A_nA_n^*)(\alpha I + AA^*)^{-1}y.$$

Using this formula, (26) and Lemma 1, we obtain the following estimate for any $\alpha = \alpha_m$ and $n = n(\alpha_m, \delta)$

$$\|(AR_\alpha(A) - A_nR_\alpha(A_n))y\|_H =$$

$$= \alpha\|(\alpha I + A_nA_n^*)^{-1}(AA^* - A_nA_n^*)(\alpha I + AA^*)^{-1}Ax^\dagger\|_H \le$$

$$\leq \frac{c_{r,\gamma} n 2^{-2nr} \gamma^{2\nu} \rho}{2\sqrt{\alpha}} \leq \frac{c_{r,\gamma,\rho} n 2^{-2nr}}{4\sqrt{\alpha}} = \delta.$$

Moreover, it is easy to see that

$$\|(I - A_n R_\alpha(A_n))(y - y_\delta)\|_H \leq \|(I - (\alpha I + A_n A_n^*)^{-1} A_n A_n^*\|_{H\to H} \|y - y_\delta\|_H \leq$$

$$\leq \alpha\delta \|(\alpha I + A_n A_n^*)^{-1}\|_{H\to H} \leq \delta. \tag{30}$$

If $\alpha = \alpha_N$ satisfies (23) then combining (29)–(30) we have

$$\|Ax_\alpha - y\|_H \leq (d+2)\delta.$$

On the other hand, the same steps as in the proof of Lemma 10 [10] lead to the inequality

$$\|Ax_{\alpha_N} - y\|_H = \|Ax_{q\alpha_{N-1}} - y\|_H \geq q\|Ax_{\alpha_{N-1}} - y\|_H, \tag{31}$$

where q is the denominator of geometric progression $\{\alpha_m : \alpha_m = q^m \alpha_0, m = 1, 2, \ldots\}$, which contains α_N. But for $\alpha = \alpha_{N-1}$ and $n = n(\alpha_{N-1}, \delta)$ from the definition of algorithm (19)–(23) it follows

$$\|A_n x^\delta_{\alpha_{N-1},n} - y_\delta\|_H > d\delta.$$

Then combining similarly (29)–(30) for $\alpha = \alpha_{N-1}$, $n = n(\alpha_{N-1}, \delta)$, by inverse triangle inequality we have

$$\|Ax_{\alpha_{N-1}} - y\|_H \geq \|A_n x^\delta_{\alpha_{N-1},n} - y_\delta\|_H - 2\delta > (d-2)\delta. \tag{32}$$

Now by virtue of (31), (32) we find

$$\|Ax_{\alpha_N} - y\|_H > q(d-2)\delta.$$

Thus, we obtain the assertion of the lemma for $d_1 = q(d-2)$, $d_2 = d+2$.

5 Complexity of the adaptive discretization scheme

Theorem 2 *The optimal order of accuracy in the power scale $\delta^{2\nu/(2\nu+1)}$ on the classes $\Phi^{r,\nu}_{\gamma,\rho}, 0 < \nu \leq 1/2$, is realized by the algorithm (19)–(23).*

Proof. From Lemma 2 it follows that for any $\alpha = \alpha_m$ and $n = n(\alpha_m, \delta)$

$$\|x^\dagger - x^\delta_{\alpha,n}\|_H \leq \frac{\delta}{2\sqrt{\alpha}} + \alpha^\nu c_{\nu,\alpha}(v) + \frac{4\delta}{\sqrt{\alpha}} = \frac{9\delta}{2\sqrt{\alpha}} + \alpha^\nu c_{\nu,\alpha}(v). \tag{33}$$

Moreover, one should note that inserting the singular value decomposition shows (see,e.g., [13]) that

$$\|Ax_\alpha - y\|^2_H = \alpha^{2\nu+1} d^2_{\alpha,\nu}(v), \tag{34}$$

where $d_{\alpha,\nu}(v)$ itself is bounded for $0 < \nu \leq 1/2$ and

$$c_{\nu,\alpha}(v)\,[d_{\alpha,\nu}(v)]^{-2\nu/(2\nu+1)} \leq c. \tag{35}$$

Now if $\alpha = \alpha_N$ satisfies (23) and n is chosen according to (19) for $\alpha_m = \alpha_N$ then from Lemma 3 and (34), (35) one sees that

$$\frac{\delta}{\sqrt{\alpha}} = \delta\left(\frac{d_{\alpha,\nu}(v)}{\|Ax_\alpha - y\|_H}\right)^{1/(2\nu+1)} \leq \delta\left(\frac{d_{\alpha,\nu}(v)}{d_1\delta}\right)^{1/(2\nu+1)} \leq c\delta^{2\nu/(2\nu+1)}, \tag{36}$$

$$\alpha^\nu c_{\nu,\alpha}(v) = c_{\nu,\alpha}(v)\left(\frac{\|Ax_\alpha - y\|_H}{d_{\alpha,\nu}(v)}\right)^{2\nu/(2\nu+1)} \leq$$

$$\leq c_{\nu,\alpha}(v)\,[d_{\alpha,\nu}(v)]^{-2\nu/(2\nu+1)}\,(d_2\delta)^{2\nu/(2\nu+1)} \leq c\delta^{2\nu/(2\nu+1)}. \tag{37}$$

The assertion of the theorem follows from (33), (36), (37) .

Let us denote by $card(Eq)$ the number of linear algebraic equations in the system corresponding to (22). Using the representation

$$A_n = \sum_{j=1}^n (P_j - P_{j-1})AP_{2n-j} + \sum_{j=1}^n P_{2n-j}A(P_j - P_{j-1}) +$$

$$+ P_{2n}AP_0 + P_0AP_{2n} - P_nAP_n$$

and (5) we get the estimate

$$card(Eq) = rank A_n^* = rank A_n \leq rank\left(\sum_{j=1}^n (P_j - P_{j-1})AP_{2n-j}\right) +$$
$$+ rank\left(\sum_{j=1}^n P_{2n-j}A(P_j - P_{j-1})\right) \leq 2 rank P_n \sim dim V_n \sim 2^{sn}. \tag{38}$$

Now we estimate the number $card(IP)$ of inner products of the form (20), (21) required to construct an approximate solution $x^{\delta}_{\alpha_m,n}$. From (5), (20), (21) it follows

$$card(IP) \le \sum_{j=0}^{2n} dimV_j \cdot dimV_{2n-j} + dimV_{2n} \sim \\ \sim \sum_{j=0}^{2n} 2^{sj} 2^{(2n-j)s} + 2^{2ns} \sim n2^{2ns}. \tag{39}$$

To illustrate the advantages of proposed adaptive discretization scheme we assume that α_N satisfying (23) has the order $\delta^{2-2\lambda}$ for some $\lambda \in (0, 1/2)$. It is sufficiently natural assumption because (see, e.g. [6]) the regularization parameter α is normally chosen in dependence of δ such that

$$\lim_{\delta \to 0} \frac{\delta^2}{\alpha} = 0, \quad \lim_{\delta \to 0} \alpha = 0.$$

Keeping in mind (19) for $\alpha_m = \alpha_N \sim \delta^{2-2\lambda}$ and

$$n2^{-2nr} \sim \delta^{2-\lambda}$$

we obtain the following estimates

$$card(Eq) \le c2^{ns} \sim \delta^{-(2-\lambda)s/(2r)} \log^{s/(2r)} \tfrac{1}{\delta},$$

$$card(IP) \le cn2^{2ns} \sim \delta^{-(2-\lambda)s/r} \log^{1+s/r} \tfrac{1}{\delta}.$$

Then the total number of inner products of the form (13) required to achieve the optimal order of accuracy in the power scale within the framework of algorithm (19)–(23) is no more than

$$card(IP) \cdot N \le c \log \frac{1}{\delta} n2^{2ns} \sim \delta^{-(2-\lambda)s/r} \log^{2+s/r} \frac{1}{\delta}.$$

Moreover, from (36) it follows that if α_N satisfies (23) then $(\delta\sqrt{\alpha_N})^{-1} \le c\delta^{-2(\nu+1)/(2\nu+1)}$. Therefore

$$Card(IP) \cdot N \le c(\delta\sqrt{\alpha_N})^{-s/r} \log^{2+s/r} (\delta\sqrt{\alpha_N})^{-1} \le c\delta^{-\frac{2s(\nu+1)}{r(2\nu+1)}} \log^{2+s/r} \frac{1}{\delta}.$$

When these relations are compared with (14) it is apparent that for the classes $\Phi^{r,\nu}_{\gamma,\rho}, 0 < \nu \le 1/2$, the adaptive discretization scheme (19)–(23) is more efficient than traditional non-adaptive approach to discretization (10)–(12).

6 Numerical example

A standard example for inverse ill–posed problems is given by (see [9])

$$Ax = y$$

with the compact operator

$$A : L^2([0,1]) \longrightarrow L^2([0,1]), \quad x \longmapsto \int_0^1 k(s,t)\; x(t)dt \;,$$

where

$$k(s,t) \;=\; \begin{cases} t(1-s) & \text{for } t \leq s \\ s(1-t) & \text{for } s \leq t \end{cases} .$$

As the solution x we take $x(s) = s(1-s)$ which yields

$$y(s) \;=\; Ax(s) \;=\; \frac{1}{12}(s^4 - 2s^3 + s) \;\; .$$

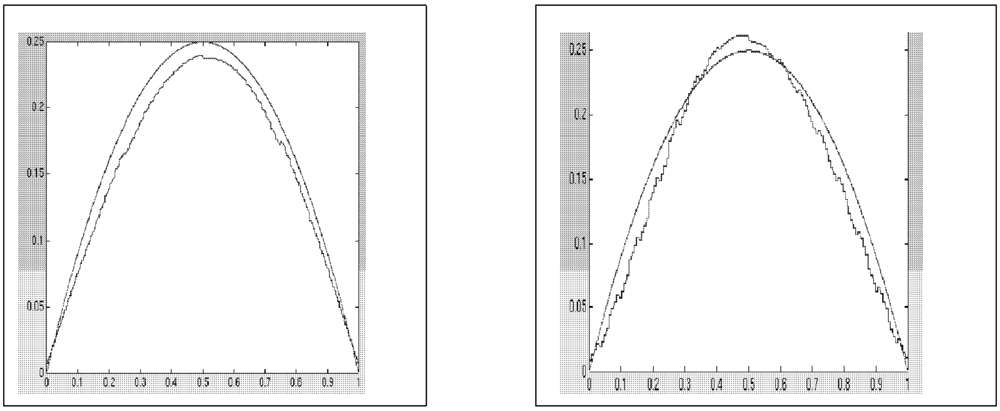
This operator lies in the considered class $A \in \mathcal{H}_\gamma^r$ with $r = 2$. We now compare the results of the full discretized operator $A_n = P_n A P_n$ with those obtained with our algorithm (19)–(23). The discrepancy principle was used in both cases with $d = 2$ and the sequence of regularization parameters $\alpha_m = q^m \alpha_0$ was obtained with $\alpha_0 = 1$ and $q = 0.8$. Moreover the constant $4c_{r,\gamma,\rho}^{-1}$ in equation (19) was set to 1 for simplicity.

The discretization spaces V_j were obtained as in Chapter 2 with the Haar-wavelet basis, i.e. $s = 2$ gives the desired level of approximation. The proposed adaptive scheme shows its advantage for small noise levels, i.e. for comparatively large dimensions n. Figure 2 displays the reconstructions with both methods for a noise level of $0,5\%$, the reconstruction error $||x^\dagger - x_{\alpha,n}^\delta||$ was 0.0145 for the adaptive scheme and 0.0165 for the full discretization. However the number of scalar products required to construct the discretized matrix A_n was 764 compared to 16384. Table 1 displays the results for different noise levels, the value of "dimension" refers to finest level of discretization, which was used for the final value of α.

noise level	method	α	dimension	number of scalar products
5%	adaptive	1	2^4	36
	full	1	2^2	16
1%	adaptive	0.0024	2^6	164
	full	0.0059	2^6	4096
0.5%	adaptive	0.0010	2^8	764
	full	0.0038	2^7	16384

Table 1: The adaptive discretization scheme shows its advantage for small noise levels, the computation of A_n requires substantially less scalar products, despite the smaller value of the α, which was determined by Morozov's discrepancy principle.

Figure 1: Left: result obtained with the adaptive strategy and noise level 0.5%, right: result obtained with full discretization and noise level 0.5%



References

[1] Bakushinskii, A.B. (1965), On some numerical method for solving integral equations of the first kind, *USSR Comput.Maths.Math.Phys.* **5**, 744–749.

[2] G. Beylkin, R. Coifman, V. Rokhlin, (1991) *Fast Wavelet Transform and Numerical Algorithms I*, Comm. Pure Appl. Math., 44, pp. 141-183.

[3] Dahmen, W., Kunoth, A.,and Schneider, R. (1996), Operator equations, multiscale concepts and complexity, In Renegar, J., Shub, M., Smale, S. (eds) *Lectures in Applied Mathematics,* **32**, Amer.Math.Soc. 225–261.

[4] Dicken, D., Maass, P. (1996), Wavelet–Galerkin Methods for Ill–posed Problems, *J. Inv. and Ill-posed Probl.,* Vol. 4, No.3., pp. 203-222.

[5] Donoho, D. L. (1992), Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition *Technical Report, Stanfrod University, Department of Statistics.*

[6] Engl, H. W. (1981), Necessary and sufficient conditions for convergence of regularization methods for solving linear operator equations of the first kind, *Numer.Funct.Anal.Optim.* **3**, 201–222.

[7] Frommer, A., Maass, P. (1998), Fast CG–based methods for Tikhonov regularization, accepted for publication *SIAM J. Sci. Comp.*.

[8] Ivanov, V.K., Vasin, V.V., Tanana, V.P. (1978) "Theory of linear ill-posed problems and its applications," Nauka, Moscow. [Russian]

[9] Louis, A. K., 1992, Inverse und schlecht gestellte Probleme, Teubner, Stuttgart.

[10] Maaß, P.,and Rieder, A. (1997), Wavelet-Accelerated Tikhonov-Phillips regularization with applications, submitted for publication, Preprint University of Potsdam.

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

[12] Neubauer, A. (1988), An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error, *Apl.Num.Math.* **4**, 507–519.

[13] Neubauer, A. (1988), An a posteriori parameter selection choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates, *SIAM J.Numer.Anal.* **25**, 1313–1326.

[14] Pereverzev, S.V., Optimization of Projection Methods for Solving Ill–Posed Problems , *Computing* **55** (1995)

[15] Pereverzev, S.V. (1996), *Optimization of Methods for Approximate Solution of Operator Equations*, Nova Science Publishers, New York.

[16] Plato, R., and Vainikko, G. (1990), On the Regularization of Projection Methods for solving Ill-posed Problems, *Numer.Math.* **57**, 63–79.

[17] Schock, E. (1984), On the asymptotic order of accuracy of Tikhonov regularization, *J.Optim.Theory Appl.* **44**, 95–104.

[18] Vainikko, G.M. (1982), The discrepancy principle for a class of regularization methods, *USSR Comput.Maths.Math.Phys.* **22**, 1–19.