A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization

Andreas Rieder*
Fachbereich Mathematik, Geb. 38
Universität des Saarlandes
66041 Saarbrücken, Germany

August 1995

Abstract

An additive Schwarz iteration is described for the fast resolution of linear ill-posed problems which are stabilized by Thikhonov regularization. The algorithm and its analysis are presented in a general framework which applies to integral equations of the first kind discretized either by spline functions or Daubechies wavelets. Numerical experiments are reported on to illustrate the theoretical results and to compare both discretization schemes.

Key words: ill-posed problems, Tikhonov regularization, additive Schwarz method, multilevel iteration, wavelets
Subject classification: AMS(MOS) 65R20, 65R30

1 Introduction

Iterative schemes for solving the compact operator equation (1.1) of the first kind,

$$Kf = g, \tag{1.1}$$

are mostly used in the sense of regularization methods due to the ill-posedness of (1.1), that is, they are used to balance the data error and the approximation error, see e. g. Louis [18]. From this point of view the convergence speed and the performance of these iterative methods are of minor importance.

In this paper we discretize equation (1.1) by applying the method of least squares and stabilize it by Tikhonov regularization, see Plato and Vainikko [21]. Since the resulting finite dimensional linear system is already regularized we are primarily interested in achieving an efficient iterative solver in terms of convergence speed, parallel coding, and performance. Therefore, an additive Schwarz relaxation will be the method of our choice.

For the construction of the multilevel solver we will split the test function space into orthogonal subspaces of increasing dimension. The number of subspaces involved is called the splitting level and the subspace with smallest dimension is referred to as coarsest space.

*email-address: andreas@num.uni-sb.de

This approach not only offers all the advantages of multilevel splittings but also yields an asymptotic orthogonality of the splitting spaces with respect to an inner product related to problem (1.1). The latter fact will be essential for the presented convergence analysis where we will rely on well-known convergence results for Schwarz type methods, see e. g. Hackbusch [15], Oswald [20] and Yserentant [27].

A first study of multilevel algorithms in connection with ill-posed problems was done by King in [17] where he proposed a method which is akin to the nested iteration known in multigrid theory, cf. Hackbusch [15]. King applied his multilevel iteration as a regularization technique in the sense mentioned above. This is one of the main differences to our approach. A more detailed comparison of King's algorithm with our algorithm is postponed to Section 5.

The outline of this paper is as follows. In the next section we give a brief account on the adequate discretization and regularization of equation (1.1) by the method of least squares. Also in the next section we introduce the multilevel splitting of the approximation space and prove some of its properties.

Section 3 is devoted to the additive Schwarz iteration. After a motivation we define and analyze the iteration in an abstract framework. We find two qualitative different convergence results: 1. For a fixed splitting level, the convergence is getting faster as the discretization step-size decreases, that is, the dimension of the approximation space increases. 2. In case the coarsest space is fixed, the convergence rate is independent of the discretization step-size and of the splitting level. We complete Section 3 with a representation of the algorithm with respect to wavelet or pre-wavelet splittings of the approximation space.

In the remainder of the paper we apply the proposed iterative scheme to integral equations on $L^2(0,1)$. Here, we present two families of test function spaces which satisfy the hypotheses of our abstract theory. These spaces are spline spaces and the spaces of the Daubechies scaling functions on the interval, see Cohen, Daubechies and Vial [6]. The numerical realization of the method in this setting is considered next. We show that approximate integration, which will be necessary in a general application of the algorithm, does not deteriorate the convergence behaviour. Finally, an analysis of the computational complexity confirms the efficiency of the iteration and the presentation of various numerical experiments support the theoretical results.

Wavelets have already been used for the treatment of inverse problems. For instance, we refer to Donoho [13], Dicken and Maaß [12] and to Xia and Nashed [26].

2 Preliminary considerations

2.1 Discretization, regularization and parameter selection

Let $K : X \to Y$ (X, Y real Hilbert spaces) be a compact non-degenerate operator. Then, it is well known that equation (1.1) is ill-posed, that is, the minimum norm solution $f^\star$ of (1.1) does not depend continuously on the right hand side g.

We now assume that only noisy data $g^\varepsilon \in Y$ are available satisfying $\|g - g^\varepsilon\|_Y \le \varepsilon$ for a known error bound $\varepsilon > 0$. A computable approximation to $f^\star$ is then provided by the unique solution $f_l^{\varepsilon,\alpha}$ of the finite dimensional normal equation

$$(K_l^* K_l + \alpha I) f_l = K_l^* g^\varepsilon, \quad \alpha > 0, \tag{2.1}$$

which is stabilized using Tikhonov regularization (throughout the paper I denotes either the identity operator or the identity matrix of appropriate size). In (2.1), $K_l = KP_l$ where $P_l : X \to V_l$ is the orthogonal projection onto the subspace $V_l \subset X$. In the sequel we will assume that the sequence $\{V_l\}_l$ of finite dimensional approximation spaces is expanding, i. e. $V_l \subset V_{l+1}$, and that the union $\cup_l V_l$ is dense in X. Under these assumptions the quantity

$$\gamma_l := \|K - K_l\| = \|K(I - P_l)\|, \tag{2.2}$$

which will be crucial for the further analysis, satisfies

$$\gamma_{l+1} \le \gamma_l \quad \text{and} \quad \gamma_l \to 0 \text{ as } l \to \infty \quad \text{iff} \quad K \text{ is compact}, \tag{2.3}$$

see e. g. Groetsch [14]. The operator norm in (2.2) is defined by $\|K\| = \sup\{\|Ku\|_Y \mid u \in X,\, \|u\|_X = 1\}$.

An a-priori ($\alpha = \alpha(l, \varepsilon)$) as well as an a-posteriori ($\alpha = \alpha(l, \varepsilon, g^\varepsilon)$) choice for the regularization parameter α in (2.1) is established by Plato and Vainikko [21] leading to the convergence of $f_l^{\varepsilon,\alpha}$ to the minimum norm solution $f^\star$ as ε tends to zero and l goes to infinity. Moreover, the resulting convergence rate is optimal in ε.

It is the goal of the paper to provide an efficient multilevel solver for equation (2.1) with one of the above mentioned parameter selection strategies *under the general assumption of a fixed noise level* ε. In this framework α is bounded below by a positive constant $\alpha_0(\varepsilon)$ uniformly in the discretization level l which gives that

$$\gamma_l \le \sqrt{\alpha_0(\varepsilon)} < \sqrt{\alpha} \quad \text{for } l \text{ sufficiently large.} \tag{2.4}$$

The above inequality guarantees a high performance of our multilevel solver as will be discussed in the Sections 4 and 5.

Remark: From an abstract point of view the normal equation (2.1) is a symmetric operator equation of the second kind. Therefore, our algorithm applies to such a class of problems in general. Nevertheless we present our algorithm in the above context of ill-posed problems since Tikhonov regularization automatically leads to linear problems (2.1) with symmetric and positive definite matrices.

2.2 Multilevel splitting of the approximation spaces

The basis of all multilevel algorithms is the decomposition of the approximation space into subspaces. To this end we define the space W_l as the X–orthogonal complement of V_l with respect to the larger space V_{l+1}: $V_{l+1} = V_l \oplus W_l$ where $\oplus$ denotes the X–orthogonal sum. Consequently, we have the orthogonal multilevel splitting

$$V_l = V_{l_{\min}} \oplus \bigoplus_{j=l_{\min}}^{l-1} W_j, \quad l_{\min} \le l - 1, \tag{2.5}$$

which also can be expressed in terms of projection operators

$$P_l = P_{l_{\min}} + \sum_{j=l_{\min}}^{l-1} Q_j \tag{2.6}$$

where Q_j is the orthogonal projection from X onto W_j.

Compact operators vanish asymptotically on the complement spaces W_l.

Lemma 2.1 *Let V_l and W_l be the spaces defined above and let $K : X \to Y$ be a compact linear operator. Then,*

$$\|K\,Q_l\| \;\le\; \gamma_l \;\to\; 0 \quad as \quad l \to \infty$$

where γ_l is defined in (2.2)*.*

Proof: The orthogonality of V_l and W_l gives $P_l Q_l = 0$. Therefore, $\|K\,Q_l\| = \|K\,(I - P_l)\,Q_l\| \le \|K\,(I - P_l)\| = \gamma_l$. ∎

The regularized normal equation (2.1) can be reformulated as a variational problem

$$\text{find } f_l^{\varepsilon,\alpha} \in V_l : \quad a(f_l^{\varepsilon,\alpha}, v_l) \;=\; \langle K_l^* g^\varepsilon, v_l\rangle_X \quad \text{for all } v_l \in V_l \tag{2.7}$$

where the bilinear form $a : X \times X \to I\!R$,

$$a(u,v) \;:=\; \langle Ku, Kv\rangle_Y \;+\; \alpha\,\langle u, v\rangle_X, \tag{2.8}$$

is symmetric and positive definite. The operators $\mathcal{A}_l = K_l^* K_l + \alpha P_l$ and $\mathcal{B}_l = Q_l K^* K Q_l + \alpha Q_l$ are associated to a via $a(u_l, v_l) = \langle \mathcal{A}_l u_l, v_l\rangle_X$ for all $u_l, v_l \in V_l$ and $a(w_l, z_l) = \langle \mathcal{B}_l w_l, z_l\rangle_X$ for all $w_l, z_l \in W_l$, respectively.

Later, we will rely on the following strong Cauchy inequality which basically says that the spaces V_l and W_l are not only X–orthogonal but also asymptotically orthogonal with respect to the inner product on X induced by the bilinear form a (2.8). The corresponding norm $\|\cdot\|_a^2 = a(\cdot\,,\cdot)$ is called energy norm on X.

Theorem 2.2 *Let V_l and W_m be defined as above and let $m \ge l$. The strong Cauchy inequality*

$$|\,a(v_l, w_m)\,| \;\le\; \min\{1, \gamma_m/\sqrt{\alpha}\} \;\; \|\,v_l\,\|_a\,\|\,w_m\,\|_a$$

holds true for all $v_l \in V_l$ and for all $w_m \in W_m$.

Proof: Since v_l and w_m are orthogonal in X we have that $a(v_l, w_m) = \langle K v_l, K w_m\rangle_Y$. Further,

$$\begin{aligned}
|a(v_l, w_m)| &= |\langle K_l \mathcal{A}_l^{-1/2}\mathcal{A}_l^{1/2} v_l\,,\; K Q_m \mathcal{B}_m^{-1/2}\mathcal{B}_m^{1/2} w_m\rangle_Y\,| \\
&\le \|K_l \mathcal{A}_l^{-1/2}\|\;\|\mathcal{A}_l^{1/2} v_l\|_X\;\|K Q_m \mathcal{B}_m^{-1/2}\|\;\|\mathcal{B}_m^{1/2} w_m\|_X \\
&\le \|K_l \mathcal{A}_l^{-1/2}\|\;\|\,v_l\,\|_a\;\|K Q_m\|\;\|\mathcal{B}_m^{-1/2}\|\;\|\,w_m\,\|_a .
\end{aligned}$$

Using arguments from spectral theory it is easy to verify that $\|K_l\,\mathcal{A}_l^{-1/2}\,\| \le 1$ and $\|\mathcal{B}_m^{-1/2}\| \le \alpha^{-1/2}$. Thus, the strong Cauchy inequality is proved by $\|K Q_m\| \le \gamma_m$ (Lemma 2.1). ∎

Corollary 2.3 *Let $j < l$. Then,*

$$|\,a(w_j, w_l)\,| \;\le\; \min\{1, \gamma_l/\sqrt{\alpha}\} \;\; \|\,w_j\,\|_a\,\|\,w_l\,\|_a$$

for all $w_j \in W_j$ and for all $w_l \in W_l$.

Proof: Because $W_j \subset V_l$ for $j < l$ the statement follows readily from Theorem 2.2. ∎

3 The additive Schwarz iteration

3.1 Abstract formulation and convergence analysis

The general philosophy behind any multilevel iteration with respect to a given splitting is to replace the original large scale problem by auxiliary problems on the subspaces which can be solved cheaply. If the subproblems are chosen in an adequate way, their combination should yield a reasonable approximation to the original problem, see e. g. Oswald [20] and Rüde [22].

We introduce some notation to clarify this multilevel concept in our setting. Recalling the statement of Lemma 2.1, the auxiliary bilinear forms $b_j : W_j \times W_j \to \mathbb{R}$ defined by

$$b_j(w_j, u_j) \;:=\; \alpha \, \langle w_j, u_j \rangle_X, \tag{3.1}$$

approximate a (2.8) reasonably well on W_j, at least for j large. Furthermore, let $\mathcal{T}_j : V_l \to W_j$ be given by the variational problem ($l_{\min} \leq j \leq l-1$)

$$b_j(\mathcal{T}_j v_l, w_j) \;=\; a(v_l, w_j) \quad \text{for } v_l \in V_l \text{ and for all } w_j \in W_j. \tag{3.2}$$

In operator notation we may write $\mathcal{T}_j = \alpha^{-1} Q_j \mathcal{A}_l$. On the coarsest approximation space $V_{l_{\min}}$ we keep the original bilinear form a. Here, we get $\mathcal{R}_{l_{\min}} : V_l \to V_{l_{\min}}$ by

$$a(\mathcal{R}_{l_{\min}} v_l, v_{l_{\min}}) \;=\; a(v_l, v_{l_{\min}}) \quad \text{for } v_l \in V_l \text{ and for all } v_{l_{\min}} \in V_{l_{\min}} \tag{3.3}$$

or $\mathcal{R}_{l_{\min}} = \mathcal{A}_{l_{\min}}^{-1} P_{l_{\min}} \mathcal{A}_l$ in operator notation. Now, we sum up the $\mathcal{T}_j$'s and $\mathcal{R}_{l_{\min}}$,

$$\mathcal{R}_{l_{\min}} + \sum_{j=l_{\min}}^{l-1} \mathcal{T}_j \;=\; \underbrace{\Big(\mathcal{A}_{l_{\min}}^{-1} P_{l_{\min}} \;+\; \alpha^{-1} \sum_{j=l_{\min}}^{l-1} Q_j \Big) \, \mathcal{A}_l}_{=:\; \mathcal{C}_{l,l_{\min}}},$$

and consider $\mathcal{C}_{l,l_{\min}}$ as an approximate inverse of $\mathcal{A}_l$ with respect to the splitting (2.5).

Starting with an initial guess $u_l^0 \in V_l$ we define the additive Schwarz iteration for the approximate solution of (2.1) resp. (2.7) by

$$u_l^{\mu+1} \;=\; u_l^{\mu} \;-\; \omega \, \mathcal{C}_{l,l_{\min}} \left(\mathcal{A}_l u_l^{\mu} \;-\; K_l^* g^{\varepsilon} \right), \quad \mu = 0, 1, 2, \ldots, \tag{3.4}$$

with a damping parameter $\omega \in \mathbb{R}$. If $\mathcal{C}_{l,l_{\min}}$ is close to $\mathcal{A}_l^{-1}$ we may expect fast convergence of $\{u_l^{\mu}\}_{\mu \geq 0}$ to $f_l^{\varepsilon,\alpha}$. For a convergence analysis we introduce a new norm on V_l relative to the splitting (2.5):

$$|||\, v_l \,|||^2 \;:=\; \|\, P_{l_{\min}} v_l \,\|_a^2 \;+\; \alpha \sum_{j=l_{\min}}^{l-1} \|\, Q_j v_l \,\|_X^2 .$$

Theorem 3.1 *Suppose that*

$$\Gamma_{\mathrm{u}} \, |||\, v_l \,|||^2 \;\leq\; \|\, v_l \,\|_a^2 \;\leq\; \Gamma_{\mathrm{o}} \, |||\, v_l \,|||^2 \quad \textit{for all } v_l \in V_l \tag{3.5}$$

holds true for some numbers $0 < \Gamma_{\mathrm{u}} \leq \Gamma_{\mathrm{o}}$. *Then,*

$$\|\, f_l^{\varepsilon,\alpha} - u_l^{\mu} \,\|_a \;\leq\; \rho_\omega^{\mu} \, \|\, f_l^{\varepsilon,\alpha} - u_l^0 \,\|_a, \quad \mu = 1, 2, \ldots,$$

where $\rho_\omega \;=\; \max\big\{ \, |1 - \omega\,\Gamma_{\mathrm{u}}|, \, |1 - \omega\,\Gamma_{\mathrm{o}}| \, \big\}$. *The convergence rate* ρ_ω *attains its minimum* $\rho_{\omega_{\mathrm{opt}}} = (\Gamma_{\mathrm{o}} - \Gamma_{\mathrm{u}}) \,/\, (\Gamma_{\mathrm{o}} + \Gamma_{\mathrm{u}}) < 1$ *for* $\omega_{\mathrm{opt}} = 2/(\Gamma_{\mathrm{u}} + \Gamma_{\mathrm{o}})$.

Proof: See e. g. Oswald [20], Theorem 18 on page 79. See also Hackbusch [15] and Yserentant [27]. ■

Providing γ_l (2.2) tends to zero sufficiently fast we can estimate the bounds Γ_{u} and Γ_{o} reasonably well.

Theorem 3.2 *Let η_l be an upper bound of γ_l ($\gamma_l \le \eta_l$) satisfying*

$$\eta_l \le \eta_{l-1} \quad \text{and} \quad \sum_{j=l_{\min}}^{l-1} \eta_j \le C_\eta\, \eta_{l_{\min}} \tag{3.6}$$

with a positive constant C_η which does neither depend on l nor on $l_{\min}$. Then, the norm equivalence (3.5) *holds true for*

$$\begin{aligned}\Gamma_{\mathrm{u}} &= \frac{1}{1 + 2\,C_\eta \left(1 + (C_\eta + 1)\,\sigma_{l_{\min}}\right) \sigma_{l_{\min}}},\\ \Gamma_{\mathrm{o}} &= \left(1 + \sigma_{l_{\min}}^2\right)\left(1 + 2\,C_\eta\, \sigma_{l_{\min}}\right),\end{aligned}$$

where $\sigma_{l_{\min}} = \eta_{l_{\min}}/\sqrt{\alpha}$. The convergence rate ρ_1 of the undamped *($\omega = 1$) iteration* (3.4) *fulfills*

$$\rho_1 \le 2\,C_\eta\, \sigma_{l_{\min}} \left(1 + \sigma_{l_{\min}}\,(1 + \max\{C_\eta, \sigma_{l_{\min}}\})\right).$$

The above theorem, which will be proved in the appendix on page 17, has to be interpreted in the following ways:

1. For a fixed *splitting level* $L = l - l_{\min} \ge 1$ and a fixed noise level $\varepsilon > 0$ the convergence rate ρ_1 tends to zero. Moreover,

$$\rho_1(l) = \mathcal{O}\left(\eta_{l-L}/\sqrt{\alpha}\right) \quad \text{as } l \to \infty, \tag{3.7}$$

where the constant in the above $\mathcal{O}$–expression does neither depend on l and L nor on α.

2. For a given noise level $\varepsilon > 0$ the undamped iteration admits a convergence rate ρ_1 which is bounded smaller than 1 uniformly in l provided that $\eta_{l_{\min}} < \sqrt{\alpha}/\left(2C_\eta\,(C_\eta + 2)\right)$. In this case $l_{\min}$ depends on ε and it increases when ε decreases.

3. If $l_{\min}$ is fixed independently of ε and l then the undamped iteration may even diverge for ε too small. Still, we can use the undamped iteration as a preconditioner for the conjugate gradient method applied to (2.1). For more details on preconditioning the conjugate gradient method we refer e. g. to Deuflhard and Hohmann [11] and Hackbusch [15].

As an initial guess for the iteration (3.4) we suggest to take $u_l^0 = f_{l_{\min}}^{\varepsilon,\alpha}$, the solution of (2.7) with respect to $V_{l_{\min}}$. The effort to calculate $f_{l_{\min}}^{\varepsilon,\alpha}$ is less than the computational work for one iteration step.

Lemma 3.3 *Let $u_l^0 = f_{l_{\min}}^{\varepsilon,\alpha}$. Then,*

$$\|f_l^{\varepsilon,\alpha} - u_l^0\|_a \le \sqrt{\|K\|^2 + \alpha}\, \inf_{v \in V_{l_{\min}}} \|f_l^{\varepsilon,\alpha} - v\|_X.$$

Proof: The proof is analogous to the proof of Céa's lemma, see e. g. Ciarlet [5]. ■

3.2 Representation in coordinates

In this section we present a matrix version of the iteration (3.4) given suitable Riesz-bases on V_l and W_l. Therefore, we assume that X is a function space over the compact interval $[a,b] \subset I\!R$; $X = L^2(a,b)$ for example. The results obtained can easily be generalized to multiple tensor products of X with itself.

Let $\varphi \in X$ be a compactly supported function satisfying the refinement equation

$$\varphi(x) \;=\; \sqrt{2} \sum_{k=0}^{M_\varphi - 1} h_k \, \varphi(2x - k) \tag{3.8}$$

with coefficients $h_k \in I\!R$ (φ is called scaling function in the wavelet terminology, see e. g. Daubechies [9]). For the ease of presentation we neglect – at the present time – necessary boundary modifications and suppose that

$$V_l = \operatorname{span}\{\varphi_{l,k} \mid k = 0, \ldots, n_l - 1\} \subset X \tag{3.9}$$

for all $l \geq l^\star > 0$ where $\varphi_{l,k}(x) := 2^{l/2}\varphi(2x - k)$. Further let there exist another function ψ, the wavelet, such that $W_l = \operatorname{span}\{\psi_{l,k} \mid k = 0, \ldots, m_l - 1\}$ and

$$\psi(x) \;=\; \sqrt{2} \sum_{k=0}^{M_\psi - 1} g_k \, \varphi(2x - k) \tag{3.10}$$

with coefficients $g_k \in I\!R$.

Since the sum of two functions $f_l = \sum_k c_k^l \varphi_{l,k} \in V_l$ and $g_l = \sum_k d_k^l \psi_{l,k} \in W_l$ is in V_{l+1}, it can be expressed by $f_l + g_l = \sum_k c_k^{l+1} \varphi_{l+1,k}$. Applying both refinement equations (3.8) and (3.10) we get the relation

$$c_k^{l+1} \;=\; \sum_i h_{k-2i} \, c_i^l \;+\; \sum_j g_{k-2j} \, d_j^l$$

which we write in matrix notation as

$$c^{l+1} \;=\; H_{l+1}^t c^l \;+\; G_{l+1}^t d^l. \tag{3.11}$$

Clearly, $H_{l+1} : I\!R^{n_{l+1}} \to I\!R^{n_l}$ and $G_{l+1} : I\!R^{n_{l+1}} \to I\!R^{m_l}$.

The solution $f_l^{\varepsilon,\alpha}$ of the variational problem (2.7) resp. of the normal equation (2.1) can be expanded in the basis of V_l as $f_l^{\varepsilon,\alpha} = \sum_k (\xi_l)_k \varphi_{l,k}$. The vector $\xi_l \in I\!R^{n_l}$ of the expansion coefficients is the unique solution of the linear system

$$A_l \, \xi_l \;=\; \beta_l \tag{3.12}$$

where the entries of the positive definite matrix A_l and of the right hand side β_l are given by

$$(A_l)_{i,j} \;=\; \langle K\varphi_{l,i}, K\varphi_{l,j}\rangle_Y \;+\; \alpha\,\langle \varphi_{l,i}, \varphi_{l,j}\rangle_X \quad \text{and} \quad (\beta_l)_j \;=\; \langle g^\varepsilon, K\varphi_{l,j}\rangle_Y\,. \tag{3.13}$$

Lemma 3.4 *Let $\mathcal{T}_j$ and $\mathcal{R}_{l_{\min}}$ be the operators introduced in (3.2) and (3.3), respectively. Further define the restrictions*

$$\begin{aligned}\mathcal{H}_{l,j} &:= H_{j+1}\, H_j \cdots H_{l-1}\, H_l \,: I\!R^{n_l} \to I\!R^{n_j}, \\ \mathcal{G}_{l,j} &:= G_{j+1}\, H_j \cdots H_{l-1}\, H_l \,: I\!R^{n_l} \to I\!R^{m_j}\end{aligned}$$

for $j \le l-2$ and set $\mathcal{H}_{l,l-1} := H_l$ and $\mathcal{G}_{l,l-1} := G_l$.

For $v_l = \sum_k c_k^l \varphi_{l,k} \in V_l$ we have that

$$\mathcal{T}_j v_l = \sum_{k=0}^{n_l-1} \left(\alpha^{-1}\, \mathcal{G}_{l,j}^t\, B_j^{-1}\, \mathcal{G}_{l,j}\, A_l\, c^l\right)_k \varphi_{l,k}, \quad l_{\min} \le j \le l-1,$$

where B_j is the Gramian matrix $(B_j)_{r,s} = \langle \psi_{j,r}, \psi_{j,s}\rangle_X$, and

$$\mathcal{R}_{l_{\min}} v_l = \sum_{k=0}^{n_l-1} \left(\mathcal{H}_{l,l_{\min}}^t\, A_{l_{\min}}^{-1}\, \mathcal{H}_{l,l_{\min}}\, A_l\, c^l\right)_k \varphi_{l,k}.$$

Proof: Since $\mathcal{T}_j \varphi_{l,k} \in W_j$ there exists a matrix $T \in I\!R^{n_l \times m_j}$ with entries $T_{k,i}$ such that $\mathcal{T}_j \varphi_{l,k} = \sum_i T_{k,i}\, \psi_{j,i}$. A function $w_j = \sum_k d_k^j \psi_{j,k} \in W_j$ admits the expansion $w_j = \sum_k \left(\mathcal{G}_{l,j}^t d^j\right)_k \varphi_{l,k}$ which follows inductively from (3.11). Hence,

$$b_j(\mathcal{T}_j v_l, w_j) = \alpha\, \langle B_j\, T\, c^l, d^j\rangle_{I\!R^{m_j}} \quad \text{and} \quad a(v_l, w_j) = \langle \mathcal{G}_{l,j}\, A_l\, c^l, d^j\rangle_{I\!R^{m_j}}.$$

Finally, the equation (3.2) gives $T = \alpha^{-1}\, B_j^{-1}\, \mathcal{G}_{l,j}\, A_l$ and we get $\mathcal{T}_j v_l = \sum_i (T c^l)_i \psi_{j,i} = \sum_k (\mathcal{G}_{l,j}^t T c^l)_k \varphi_{l,k}$.

The representation of $\mathcal{R}_{l_{\min}}$ can be proved in the same way. ∎

Now, the abstract iteration (3.4) translated into an iteration acting on (3.12) reads

$$z_l^{\mu+1} = z_l^{\mu} - \omega\, C_{l,l_{\min}} \left(A_l z_l^{\mu} - \beta_l\right), \quad \mu = 0, 1, 2, \ldots, \tag{3.14}$$

where

$$C_{l,l_{\min}} = \mathcal{H}_{l,l_{\min}}^t\, A_{l_{\min}}^{-1}\, \mathcal{H}_{l,l_{\min}} + \alpha^{-1} \sum_{j=l_{\min}}^{l-1} \mathcal{G}_{l,j}^t\, B_j^{-1}\, \mathcal{G}_{l,j} \tag{3.15}$$

and with an arbitrary starting guess $z_l^0 \in I\!R^{n_l}$. The starting guess due to Lemma 3.3 is

$$z_l^0 = \mathcal{H}_{l,l_{\min}}^t \xi_{l_{\min}} = \mathcal{H}_{l,l_{\min}}^t A_{l_{\min}}^{-1} \mathcal{H}_{l,l_{\min}} \beta_l. \tag{3.16}$$

In the next section we will see that the application of B_j^{-1} and hence of $C_{l,l_{\min}}$ to a vector can be realized in a very efficient way.

4 Application to integral equations

We demonstrate the performance of our proposed iterative scheme for the solution of integral equations. For simplicity we limit ourselves to problems on the unit interval, but the algorithm can be carried over to problems on Cartesian products of intervals.

Let $K : L^2(0,1) \to L^2(0,1)$ be an integral operator with non-degenerate kernel k which is square integrable over the unit square $[0,1]^2$. Then, the integral equation of the first kind,

$$Kf(\cdot) \;=\; \int_0^1 k(\cdot,y)\,f(y)\,dy \;=\; g(\cdot), \tag{4.1}$$

is ill-posed.

To estimate the decay rate of γ_l as l goes to infinity, we consider scaling function spaces V_l (3.9) of *order* N, that is, the polynomials up to degree $N-1$ restricted to $[0,1]$ are in V_l.

We associate the discretization step-size $\delta_l = 2^{-l}$ to V_l and we will use D^r to denote the (generalized) differential operator of order r.

Lemma 4.1 *Let V_l be the spaces* (3.9) *of order N. Further, let K be the integral operator* (4.1) *and suppose that the composition $D^r K^*$ is a bounded operator on $L^2(0,1)$ for one $r \in \{1,\ldots,N\}$. Then,*

$$\gamma_l \;=\; \|(I-P_l)K^*\| \;\le\; C_N\;\|D^r\,K^*\|\;\delta_l^r\,. \tag{4.2}$$

Proof: If f is sufficiently smooth we have the following approximation result: $\|\,(I-P_l)f\,\|_{L^2} \;\le\; C_N\;\|D^r f\|_{L^2}\;\delta_l^r$ for $r=1,\ldots,N$, see e. g. Strang [23], Strang and Fix [24]. The estimate for γ_l is a straightforward consequence. ∎

In the framework of scaling function spaces of order $N \ge 1$, Theorem 3.2 applies because the requirements (3.6) are met for $\eta_l = C_N\,\|D^r\,K^*\|\,\delta_l^r$ with $r \ge 1$ and the decay rate (3.7) reads now

$$\rho_1(l) \;=\; \mathcal{O}\big(\delta_{l-L}^r/\sqrt{\alpha}\big) \quad \text{as } l \to \infty. \tag{4.3}$$

The constant in (4.3) is independent of l, L and α.

4.1 Wavelet and spline spaces

Let V_l, $l \ge 0$, be the space that contains all functions being constant on the intervals $[2^{-l}k, 2^{-l}(k+1)[$ $(k=0,\ldots,2^l-1)$. This is a simple but admissible choice: V_l is of the form (3.9) when we set $\varphi = \chi_{[0,1[}$, the characteristic function of the interval $[0,1[$. The complement spaces W_l have also the dimension 2^l and they are spanned by the Haar-wavelet $\psi = \chi_{[0,1/2[} - \chi_{[1/2,1[}$, see Daubechies [8]. The refinement equations (3.8) and (3.10) hold true with $h_0 = h_1 = 1/\sqrt{2}$ and $g_0 = -g_1 = 1/\sqrt{2}$, respectively ($M_\varphi = M_\psi = 2$). Clearly, the order of V_l is $N=1$.

The asymptotic behaviour (4.3) is illustrated in Figure 1 where approximations to $\rho_1(l)$, $L=4$, are plotted for several l and $\alpha = 0.001$. The underlying kernel is $k_1(x,y) = x-y$, if $x \ge y$, and $k_1(x,y)=0$, otherwise. Besides the convergence speed we plotted the rate

$$q_l \;=\; \rho_1(l)\,/\,\rho_1(l-1). \tag{4.4}$$

Since (4.3) is valid with $r=1$ and $L=4$ we expect q_l to be bounded: $q_l \le 0.5$ for l large.

To improve the poor decay rate of ρ_1 obtained in the Haar-wavelet case we introduce scaling function spaces containing polynomials of higher degree. We present two different families of functions which both can be considered as further developments of the introductory example above.

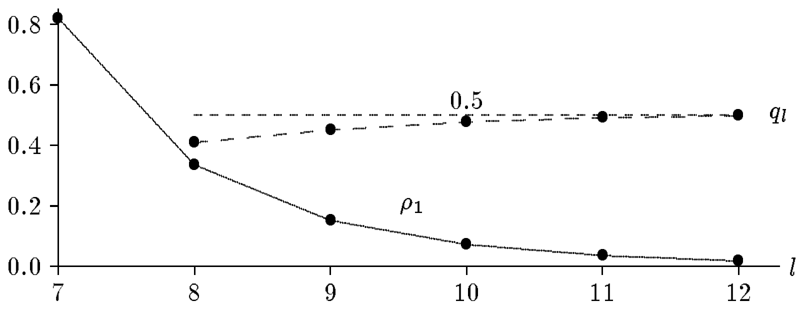

Figure 1: Solid curve: convergence rates ρ_1 of the undamped iteration (3.14), $l_{\min} = l - 4$, with respect to the Haar-wavelet and with $\alpha = 0.001$. The underlying kernel is $k_1(x, y) = x - y$, if $x \geq y$, and $k_1(x, y) = 0$, otherwise. Dashed curve: the rate q_l (4.4). The theoretical bound 0.5 for q_l is drawn as a dashed straight line. On level $l = 6$ the iteration diverges.

4.1.1 Daubechies wavelets on the interval

We will briefly recall the wavelet systems on the interval $[0, 1]$ constructed by Cohen et al. in [6]. This construction is a modification of the Daubechies wavelet family on the real line, see Daubechies [8].

Let $\varphi = \varphi_N$ and $\psi = \psi_N$ be the Daubechies scaling and wavelet function of order $N \geq 2$ which both have compact support in $[0, 2N - 1]$. Define $l^\star$ to be the smallest integer such that $2^{l^\star} \geq 2N$. For $l \geq l^\star$ there exist $2N$ *edge scaling functions* φ_k^0, φ_k^1, and $2N$ *edge wavelets* ψ_k^0, ψ_k^1, such that each of the sets

$$\begin{aligned} X_l &= \{\varphi_{l,k}^0, \mid 0 \leq k \leq N - 1\} \\ &\quad \cup \left\{\varphi_{l,k} \mid N \leq k \leq 2^l - N - 1\right\} \cup \left\{\varphi_{l,k}^1 \mid 2^l - N \leq k \leq 2^l - 1\right\}, \\ Y_l &= \{\psi_{l,k}^0 \mid 0 \leq k \leq N - 1\} \\ &\quad \cup \left\{\psi_{l,k} \mid N \leq k \leq 2^l - N - 1\right\} \cup \left\{\psi_{l,k}^1 \mid 2^l - N \leq k \leq 2^l - 1\right\}, \end{aligned}$$

is a family of orthonormal functions (for convenience we use the notation $f_{l,k}^0(x) = 2^{l/2} f_k^0(2^l x)$ and $f_{l,k}^1(x) = 2^{l/2} f_k^1(2^l(1 - x))$ only for edge functions). The spaces V_l^d and W_l^d of dimensions $n_l = m_l = 2^l$,

$$V_l^d := \operatorname{span} X_l \quad \text{and} \quad W_l^d := \operatorname{span} Y_l, \tag{4.5}$$

have all properties required in the previous sections, for instance, the order of V_l^d is N. The edge functions satisfy – as the interior functions – a kind of refinement equation. The corresponding coefficients are tabulated in [6].

Remark: Since the wavelet-basis in W_l^d is an orthonormal basis the corresponding Gramian matrix B_l coincides with the identity matrix which simplifies the structure of $C_{l,l_{\min}}$ (3.15).

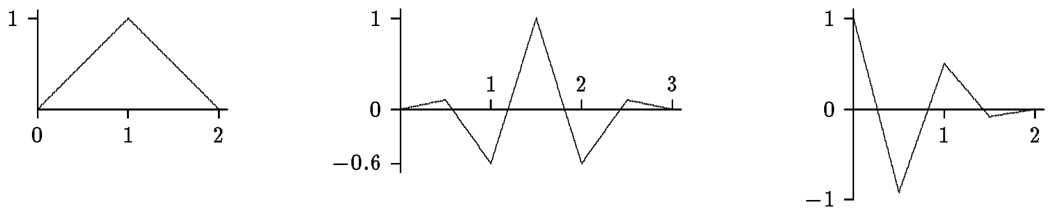

Figure 2: The linear B–Spline S_2 (4.6), its corresponding (pre-)wavelet ψ^s (4.8) and the edge spline-wavelet ψ^0 (4.9) for the left boundary (from left to right).

4.1.2 Spline spaces on the interval

The N–th order B–spline S_N is defined as the N–fold convolution of $\chi_{[0,1]}$. The B–splines have compact support in $[0, N]$ and they satisfy the refinement relation

$$S_N(x) = \sum_{k=0}^{N} 2^{1-N} \binom{N}{k} S_N(2x - k).$$

The corresponding B–spline-wavelet of order N is well known, see e. g. Chui [2] for a comprehensive introduction to spline-wavelets.

As in the case of the Daubechies wavelets we will need *edge splines* and *edge spline-wavelets* to yield approximation spaces of order N on $[0, 1]$. For $N = 2$ we will give explicit expressions. The general case is considered by Chui and Quak in [3].

The graph of the linear B–Spline S_2 is shown in Figure 2. We have that

$$S_2(x) := S(x) = \frac{1}{2}\left(S(2x) + 2\,S(2x-1) + S(2x-2) \right). \tag{4.6}$$

To span the linear functions on $[0, 1]$, we need two edge splines S^0 and S^1 given by $S^0(x) = 1 - x$, if $0 \le x \le 1$, $S_0(x) = 0$, otherwise, and $S^1(x) = S^0(2 - x)$. The modified scaling equation is

$$S^0(x) = S^0(2x) + \frac{1}{2}\,S(2x).$$

The space V_l^s of dimension $n_l = 2^l + 1$,

$$V_l^s := \operatorname{span}\{\, S_l^0,\, S_l^1,\, S_{l,k} \mid k = 0, \ldots, 2^l - 2 \,\}, \tag{4.7}$$

where $S_l^0(x) := S^0(2^l x)$ and $S_l^1(x) := S_l^0(1-x)$ $(l \ge 1)$, coincides with the space of continuous functions which are affine linear on $[2^{-l}k, 2^{-l}(k+1)[$.

The interior wavelet ψ^s given by, see Chui and Wang [4],

$$\psi^s(x) = \frac{1}{10}\,S(2x) - \frac{3}{5}\,S(2x-1) + S(2x-2) - \frac{3}{5}\,S(2x-3) + \frac{1}{10}\,S(2x-4) \tag{4.8}$$

has compact support in $[0, 3]$, cf. Figure 2. The edge wavelets ψ^0 and ψ^1 can be calculated directly and they fulfill

$$\begin{aligned} \psi^0(x) &= S^0(2x) - \frac{11}{12}\,S(2x) + \frac{1}{2}\,S(2x-1) - \frac{1}{12}\,S(2x-2), \\ \psi^1(x) &= \psi^0(2-x), \end{aligned} \tag{4.9}$$

cf. Figure 2. Again, we set $\psi_l^0(x) := \psi^0(2^l x)$, $\psi_l^1(x) := \psi_l^0(1-x)$ $(l \geq 1)$, and define

$$W_l^s \;:=\; \operatorname{span}\{\, \psi_l^0,\, \psi_l^1,\, \psi_{l,k}^s \,|\, k = 0, \ldots, 2^l - 3 \,\}$$

for $l \geq l^\star = 2$. The dimension of W_l^s is $m_l = 2^l$ and the reader may convince himself that $V_{l+1}^s = V_l^s \oplus W_l^s$, $l \geq l^\star$.

Since the constructed bases in V_l^s and W_l^s are only Riesz-bases the functions ψ^0, ψ^1, ψ^s, are usually called semi-orthogonal wavelets or pre-wavelets in the wavelet literature.

Remark: The Gramian matrix B_l with respect to the basis in W_l^s is a band matrix with band width 3 and B_l differs from a Toeplitz matrix only at both corners of its diagonal. Utilizing these special features together with the Cholesky decomposition, see e. g. Deuflhard and Hohmann [11], it is possible to compute the action of B_l^{-1} on a vector with $\mathcal{O}(m_l)$ operations. The additional storage is independent of the dimension.

4.2 Numerical realization and computational complexity

4.2.1 Matrix generation

The sensitive point in using the proposed method lies in the computation of the entries of the matrix A_l (3.13).

Since the integrals $\langle f, \varphi_{l,j}\rangle_{L^2}$ can be calculated only approximately we choose a quadrature rule of the form

$$Q_{l,j}(f) \;=\; \delta_l^{1/2} \sum_{t=0}^{m} w_t\, f(x_{l,t,j}) \tag{4.10}$$

with weights $w_t \in I\!R$ and abscissae $x_{l,t,j} \in [0,1]$, see Sweldens and Piessens [25]. If $Q_{l,j}$ is exact for polynomials up to degree m and if $f \in \mathcal{C}^{m+1}$, then

$$|\langle f, \varphi_{l,j}\rangle_{L^2} \;-\; Q_{l,j}(f)| \;\leq\; C_Q \max_{x\in[0,1]} |f^{(m+1)}(x)|\; \delta_l^{m+3/2} \tag{4.11}$$

where the positive constant C_Q does not depend on j. The error estimate (4.11) follows from Peano's theorem, see e. g. Davis and Rabinowitz [10].

Approximating the action of K on $\varphi_{l,j}$ by the integration rule (4.10) we have to deal with a perturbation $\widetilde{K}_l = \widetilde{K} P_l$ of K_l where

$$\widetilde{K}\varphi_{l,j}(x) \;:=\; Q_{l,j}\big(k(x,\cdot)\big) \;=\; 2^{-l/2} \sum_{t=0}^{m} w_t\, k(x, x_{l,t,j}).$$

Provided the degree of accuracy m (4.11) of the quadrature rule is sufficiently high, we will show that the convergence analysis of the previous section carries over to the perturbed linear system

$$\widetilde{A}_l\, \widetilde{\xi}_l \;=\; \widetilde{\beta}_l \tag{4.12}$$

with
$$\begin{aligned}(\widetilde{A}_l)_{i,j} \;&=\; \langle \widetilde{K}\varphi_{l,i}, \widetilde{K}\varphi_{l,j}\rangle_{L^2} \;+\; \alpha\,\langle \varphi_{l,i}, \varphi_{l,j}\rangle_{L^2} \\ &=\; \delta_l \sum_{s=0}^{m}\sum_{t=0}^{m} w_s\, w_t \int_0^1 k(x, x_{l,s,i})\, k(x, x_{l,t,j})\, dx \;+\; \alpha\,\langle \varphi_{l,i}, \varphi_{l,j}\rangle_{L^2}\end{aligned} \tag{4.13}$$

and $(\widetilde{\beta}_l)_j = \delta_l^{1/2} \sum_t w_t \int g^\varepsilon(x)\, k(x, x_{l,t,j})\, dx$, cf. (3.13).

Lemma 4.2 *Let V_l be either V_l^d (4.5) or V_d^s (4.7) and let $\widetilde{K}_l$ be defined as above with the quadrature rule (4.10) of accuracy m. Further, suppose that the kernel k of the operator K (4.1) is $m+1$–times continuously differentiable with respect to y. Then,*

$$\widetilde{\gamma}_l := \|K - \widetilde{K}_l\| = \mathcal{O}(\delta_l^{m+1}).$$

Proof: Obviously, $\|K - \widetilde{K}_l\| \le \gamma_l + \|K_l - \widetilde{K}_l\|$. Using Cauchy's inequality for sums and the fact that we have a Riesz-basis on V_l, it is straightforward to verify that

$$\begin{aligned} \|(K_l - \widetilde{K}_l)f\|^2 \quad &\le \quad C\,\|f\|_{L^2}^2\; n_l \max_{0\le j\le n_l-1} \int_0^1 |(K-\widetilde{K})\varphi_{l,j}(x)|^2\,dx \\ &\overset{(4.11)}{\le} \quad C\,\|f\|_{L^2}^2\; n_l\; C_Q^2 \max_{x,y\in[0,1]} \left|\frac{\partial^{m+1}}{\partial y^{m+1}}\,k(x,y)\right|^2 \delta_l^{2m+3}. \end{aligned}$$

The assumption on the kernel gives $\|K_l - \widetilde{K}_l\| = \mathcal{O}(\delta_l^{m+1})$ as well as $\gamma_l = \mathcal{O}(\delta_l^{m+1})$ by Lemma 4.1. ■

Lemma 2.1 holds also true for $\widetilde{K}$ qualitatively: $\|\widetilde{K}Q_l\| = \|\widetilde{K}P_{l+1}(I-P_l)Q_l\| \le \|\widetilde{K}_{l+1} - K\| + \|K - \widetilde{K}_l\| = \widetilde{\gamma}_{l+1} + \widetilde{\gamma}_l$.

Altogether we found: *The convergence result stated in Theorem* 3.2 *applies also to the iteration* (3.14) *when A_l is replaced by $\widetilde{A}_l$. Moreover, the optimal order r of the decay rate* (4.3) *is not affected by numerical integration as long as the quadrature rule has the degree of accuracy $r-1$.*

4.2.2 Computational complexity and implementation issues

We demonstrate the efficiency of the multilevel algorithm (3.14) for the solution of (4.12) by comparing its complexity with the complexity of the Cholesky decomposition which is a suitable direct solver to tackle (4.12).

We make the following general assumptions: We will not include the computation of $\widetilde{A}_l$ (4.13) into the operation count of the algorithm because this effort has to be raised independent of the specific solver we use for (4.12). The application of B_j^{-1} in (3.15) is performed according to the remark at the end of Section 4.1.2. Further, the auxiliary linear system with matrix $\widetilde{A}_{l_{\min}}$ is solved using Cholesky decomposition where the decomposition is computed once before the iteration starts. Finally, the underlying scaling function spaces have order N and the degree of accuracy of the used quadrature rule is $N-1$.

The different parts of the algorithm have the following operation count (we consider the leading terms only):

1. *Computation of the matrix* $\widetilde{A}_{l_{\min}}$ (4.13): $N^2\,n_{l_{\min}}^2/2$. Here, we did not take into account the evaluation of the kernel at the quadrature points and we supposed that the integral and the inner products are calculated exactly.
2. *Cholesky decomposition of* $\widetilde{A}_{l_{\min}}$ *and computation of the starting guess* (3.16): $n_{l_{\min}}^3/6$.
3. *One step of the iteration*: n_l^2. Since the application of $C_{l,l_{\min}}$ (3.15) to a vector can be realized by $\mathcal{O}(n_l + n_{l_{\min}}^2)$ operations, the evaluation of the defect dominates the complexity.

Thus, s steps of the iteration take essentially

$$O_s := s\,n_l^2 + n_{l_{\min}}^3/6 + N^2\,n_{l_{\min}}^2/2$$

operations.

If we fix the splitting level $L = l - l_{\min}$ then we have on one side that the convergence rate tends to zero for large l and only a few steps of the iteration yield the solution, cf. Table 1. On the other side O_s is dominated by $n^3_{l_{\min}} = n^3_{l-L}$ which gives the unfavourable complexity $O_s = \mathcal{O}(n_l^3)$. However, with a sophisticated implementation we can achieve both, a decreasing convergence rate and an optimal complexity.

The idea is as follows: we allow the splitting level L to grow but not too fast. We therefore define

$$L(l) := \lfloor l/3 \rfloor \quad \text{for } l \geq l^\star + 1, \tag{4.14}$$

where $\lfloor\cdot\rfloor$ denotes the 'greatest integer' and $l^\star$ is the smallest possible approximation level, cf. Sections 4.1.1 and 4.1.2. Now the asymptotic relation (4.3) reads

$$\rho_1(l) = \mathcal{O}\big(\eta_{l-L(l)}/\sqrt{\alpha}\big) = \mathcal{O}\big(\delta_l^{2r/3}/\sqrt{\alpha}\big) \quad \text{as } l \to \infty. \tag{4.15}$$

So, we sacrificed one third of the optimal decay rate of ρ_1 to have the following operation count $(n_l = 2^l)$

$$O_s \leq (s + 4/3)\, n_l^2 + 2N^2\, n_l^{4/3}.$$

In the latter implementation the undamped multilevel iteration can be viewed as a direct solver with complexity order $\mathcal{O}(n_l^2)$ which outperforms the Cholesky solver.

Remark: Since $C_{l,l_{\min}}$ (3.15) is a sum of matrices its action on a vector can be performed in parallel. This leads to a speed up when the iteration is implemented on a parallel machine.

Moreover, the iteration can even be accelerated with the help of matrix compression techniques presented by Harten and Yad-Shalom [16], see also Beylkin, Coifman and Rokhlin [1] and Dahmen, Prössdorf and Schneider [7]. The matrix-vector product $\widetilde{A}_l v_l$ can be calculated with only $\mathcal{O}(n_l)$ or $\mathcal{O}(n_l \log n_l)$ operations as far as the kernel k satisfies some additional requirements.

4.3 Numerical experiments

We shall provide numerical approximations to the convergence rates to demonstrate the theoretical results as well as the performance of the Schwarz relaxation (3.14).

For convenience, we limit ourselves to the Daubechies wavelet system of order $N = 2$ and to the linear spline system of the same order. In both cases we computed the weights of the quadrature rules such that affine linear functions are integrated exactly. The quadrature points are of the form $x_{l,t,j} = 2^{-l}(j + t)$. In all presented experiments the integral in (4.13) is evaluated exactly and all kernels are smooth enough to yield $\widetilde{\gamma}_l = \mathcal{O}(\delta_l^2)$.

The asymptotic behaviour (4.3) is illustrated in Figure 3 where approximations to ρ_1, $L = 5$, are plotted for several l and the regularization parameter $\alpha = 0.001$. The underlying kernel is $k_1(x, y) = x - y$, $x \geq y$, $k_1(x, y) = 0$, otherwise.

Next, we compute approximations to ρ_1 where the coarsest level $l_{\min} = 2$ is fixed. We know that $\rho_1 < 1$ uniformly in l if α is not too small. Here, $\alpha \geq 0.001$ is numerically sufficient, see Figure 4.

In both examples the convergence rates with respect to the linear splines are clearly smaller than those with respect to the Daubechies wavelet. The main reason for this observation is the higher regularity of the linear B–spline: The Daubechies wavelet of order 2 is Hölder

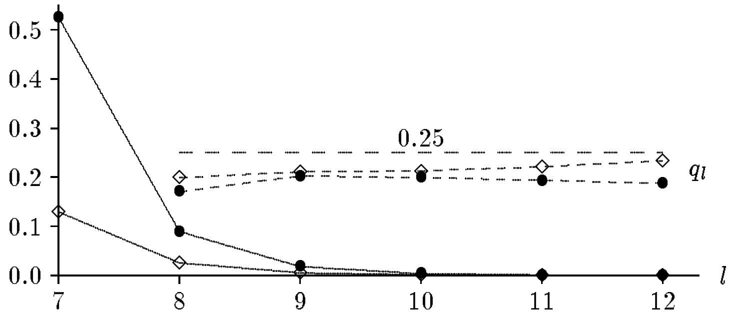

Figure 3: Solid curves: convergence rates ρ_l of the iteration (3.14), $l_{\min} = l - 5$, with respect to the kernel k_1 and with $\alpha = 0.001$ ($\bullet$: Daubechies wavelet, $N = 2$; $\diamond$: linear spline). Dashed curves: the rate q_l (4.4). The theoretical bound 0.25 for q_l is drawn as a dashed straight line.

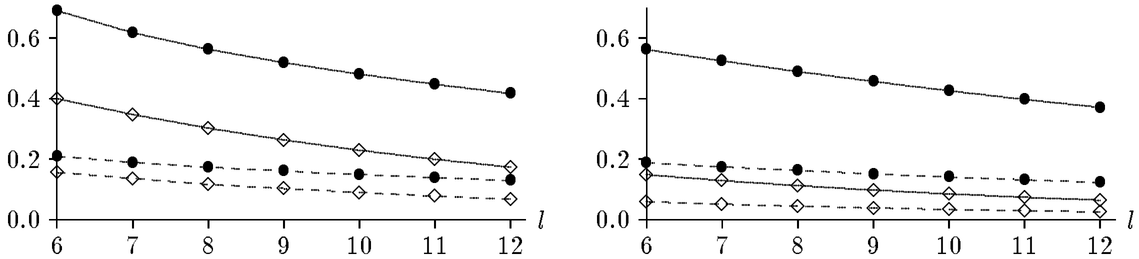

Figure 4: Convergence rates ρ_l with respect to the fixed coarsest level $l_{\min} = 2$ ($\bullet$: Daubechies wavelet, $N = 2$; $\diamond$: linear spline). Left: $k_2(x, y) = \cos(\pi\, x\, y)$, Right: $k_1(x, y) = x - y$, $x \geq y$, $k_1(x, y) = 0$, otherwise. Solid curves: $\alpha = 0.001$, dashed curves: $\alpha = 0.005$.

continuous with exponent 0.55, see Daubechies [9], whereas the linear B–spline is Lipschitz continuous.

Our last example illustrates the performance of the iteration (3.14) by an approximate solution of (4.1) with kernel k_1 and exact right hand side $g(x) = x^2(x^2 - 4x + 9)/12 + (1 - \cos(3\pi x))/\pi^2/18$. Thus, $f^\star(x) = (1 - x)^2 + \cos^2(3\pi x/2)$. In our computations we supposed that g^ε is known only at the discrete points $j\, \delta_l$ with $g^\varepsilon(j\, \delta_l) = g(j\, \delta_l) + \varepsilon_j$ where the random errors $\{\varepsilon_j\}$ are distributed uniformly in $[-\varepsilon, \varepsilon]$. The integrals in the $(\widetilde{\beta}_l)_j$'s are evaluated by the trapezoidal rule. Table 1 contains the number s of iteration steps to yield an Euclidean norm of the residue smaller than $10^{-4} \cdot \alpha \cdot \|z_l^s\|$. Here, $\{z_l^\mu\}_\mu$ denotes the sequence of iterates generated by (3.14) with starting guess either $z_l^0 = 0$ or $z_l^0 = \widetilde{u}_{l_{\min}} := \mathcal{H}^t_{l,l_{\min}} \widetilde{A}^{-1}_{l_{\min}} \mathcal{H}_{l,l_{\min}} \widetilde{\xi}_{l_{\min}}$ (3.16). Our stopping criterion guarantees a relative accuracy $\|z_l^s - \widetilde{\xi}_l\|/\|z_l^s\|$ bounded by 10^{-4}.

The approximate solutions with respect to the discretization levels $l = 6$ and $l = 11$ together with the minimum norm solution $f^\star$ are displayed in Figure 5. We considered an absolute error $\varepsilon = 0.04$ and got the optimal regularization parameter by trial and error. It took 14 ($l = 6$) and 1 ($l = 11$) iteration steps ($L = 4$) to bound the relative accuracy by 0.01.

		$l=7$	$l=8$	$l=9$	$l=10$	$l=11$	$l=12$
$l_{\min}=l-5$	$z_l^0=0$	49	7	4	3	3	2
	$z_l^0=\widetilde{u}_{l_{\min}}$	12	4	2	1	1	1
$l_{\min}=3$	$z_l^0=0$	7	7	7	7	7	7
	$z_l^0=\widetilde{u}_{l_{\min}}$	4	4	2	4	4	4

Table 1: Necessary iterations to guarantee a relative accuracy smaller than 10^{-4} ($\varepsilon = 0.04$, $\alpha = 0.001$, Daubechies wavelet, starting guess either $z_l^0 = 0$ or $z_l^0 = \widetilde{u}_{l_{\min}} := \mathcal{H}^t_{l,l_{\min}} \widetilde{A}^{-1}_{l_{\min}} \mathcal{H}_{l,l_{\min}} \widetilde{\xi}_{l_{\min}}$ (3.16)).

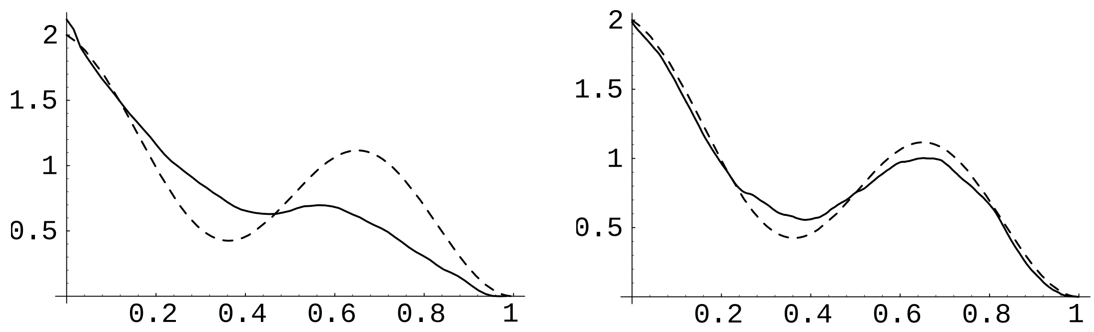

Figure 5: Approximate solutions (solid lines) and minimum norm solution (dashed lines) of (4.1) with kernel k_1 ($\varepsilon = 0.04$, Daubechies wavelet). Left: $l = 6$, $\alpha = 0.001$, right: $l = 11$, $\alpha = 1.5 \cdot 10^{-4}$. Each α is optimal in the sense that the L^2–error is minimized.

5 Discussion and conclusion

Our proposed multilevel iteration works most efficiently in its implementation discussed in Section 4.2.2 if

$$\eta_{l-L(l)} \ll \sqrt{\alpha}\,, \tag{5.1}$$

cf. (4.15). In view of our general assumption that the noise level $\varepsilon > 0$ is fixed, the requirement (5.1) holds for l such large that

$$\eta_{l-L(l)} \ll \sqrt{\alpha_0(\varepsilon)}\,, \tag{5.2}$$

where $\alpha_0(\varepsilon)$ is as in (2.4). The larger l the more efficient is the algorithm. The crucial condition $\varepsilon > 0$ is reasonable for real-life applications where noise can not be avoided due to the nature of the specific experiment and due to the limitations of the measuring instrument.

Readers familiar with ill-posed problems might see a serious drawback of the algorithm in condition (5.2) which is satisfied for large dimensions of the system (2.1). At a first glance, this seems to be contradiction to the rule: 'Don't discretize too fine to avoid noise amplification!'. The rule applies in the case the regularization is obtained only by discretization, see e. g. Natterer [19]. However, in our setting the stability of (2.1) comes from Tikhonov regularization and here it pays to refine the discretization: the solution $f_l^{\varepsilon,\alpha}$ of (2.1) is polluted by a discretization error as well as by the unavoidable noise error. Increasing the dimension reduces the discretization error without amplifying the noise. For large l, the remaining error $f_l^{\varepsilon,\alpha} - f^\star$ is caused only by the noise provided the regularization parameter α is selected according to one of the strategies mentioned in Section 2.1. This argument is supported by the error estimates given by Plato and Vainikko in [21] and by Figure 5 where the approximate solution on the left contains a relatively large discretization error as well as the noise error. In contrast, the discretization error of the approximate solution shown on the right side is negligible. Hence, this approximation is much closer to the exact minimum norm solution.

In the remainder of this section we focus on the comparison of King's multilevel method [17] with our algorithm. There are several differences: the most important one is – as already mentioned in the introduction – that King considers his iterative scheme as a regularization technique, that is, he uses the number of iteration steps to balance the discretization and the data error. In his setup relation (5.2) is superfluous, however, for the price of a slow convergence speed of his full multilevel algorithm. Indeed, even if (5.2) is satisfied, King's iteration has a splitting level dependent convergence rate, see [17, Corollary 3.4], which deteriorates for a increasing number of levels. In our case, provided (5.2) applies, the convergence becomes even faster with increasing discretization level and hence with increasing splitting level $L(l)$ (4.14). Of course, our algorithm can also be performed when (5.1) is strongly violated, for instance, if ε is zero. Its performance and regularization ability in the latter situation remains still to be investigated.

Besides the above comments our method outperforms King's method with respect to the following aspects which we only touch briefly.

- No damping parameter is needed to gain optimal convergence results.
- Each iteration step is parallelizable due to the additive structure.
- One iteration step is much cheaper (King's algorithm is a multiplicative iteration which requires two matrix-vector multiplications on each level of the multilevel process).

At the end we like to emphasize that it is unclear how the performance of King's method improves (if it improves at all!) with the smoothness of the operator and the order of the approximation spaces.

Appendix: Proof of Theorem 3.2

The proof of Theorem 3.2 will be given by providing bounds Γ_u and Γ_o for the inequality (3.5).

We note the estimate

$$b_j(w_j,w_j) \le a(w_j,w_j) \le \left(1 + \gamma_{l_{\min}}^2/\alpha\right) b_j(w_j,w_j) \quad \text{for all } w_j \in W_j,\ j \ge l_{\min}, \tag{A.1}$$

which is an immediate consequence from the definition of b_j (3.1) and from Lemma 2.1.

First, we consider the upper bound Γ_o. Let $v \in V_l$ then $v = P_{l_{\min}} v + \sum_j Q_j v$ due to (2.6). For the following estimate we will use the statements of Theorem 2.2 and Corollary 2.3 as well as the inequalities (A.1), (3.6), and $2|xy| \le x^2 + y^2$, $x, y \in I\!R$:

$$\begin{aligned} a(v,v) &= a(P_{l_{\min}}v, P_{l_{\min}}v) + \sum_{j=l_{\min}}^{l-1} a(Q_j v, Q_j v) \\ &\quad + 2\sum_{j=l_{\min}}^{l-1} a(P_{l_{\min}}v, Q_j v) + 2\sum_{j=l_{\min}}^{l-2}\sum_{i=j+1}^{l-1} a(Q_j v, Q_i v) \end{aligned}$$

$$
\begin{aligned}
&\le \left(1 + \eta_{l_{\min}}^2/\alpha\right) |||\,v\,|||^2 + 2 \sum_{j=l_{\min}}^{l-1} \eta_j\, a(P_{l_{\min}}v, P_{l_{\min}}v)^{1/2} a(Q_jv, Q_jv)^{1/2}/\sqrt{\alpha} \\
&\quad + 2 \sum_{j=l_{\min}}^{l-2} \sum_{i=j+1}^{l-1} \eta_i\, a(Q_jv, Q_jv)^{1/2} a(Q_iv, Q_iv)^{1/2}/\sqrt{\alpha} \\
&\le \left(1 + \eta_{l_{\min}}^2/\alpha\right) |||\,v\,|||^2 + \sum_{j=l_{\min}}^{l-1} \eta_j \left(a(P_{l_{\min}}v, P_{l_{\min}}v) + a(Q_jv, Q_jv)\right)/\sqrt{\alpha} \\
&\quad + \sum_{j=l_{\min}}^{l-2} \sum_{i=j+1}^{l-1} \eta_i \left(a(Q_jv, Q_jv) + a(Q_iv, Q_iv)\right)/\sqrt{\alpha} \\
&\le \left(1 + \eta_{l_{\min}}^2/\alpha\right) |||\,v\,|||^2 + C_\eta \eta_{l_{\min}} \Big(a(P_{l_{\min}}v, P_{l_{\min}}v) + \sum_{j=l_{\min}}^{l-1} a(Q_jv, Q_jv)\Big)/\sqrt{\alpha} \\
&\quad + C_\eta \eta_{l_{\min}} \sum_{j=l_{\min}}^{l-1} a(Q_jv, Q_jv)/\sqrt{\alpha} \\
&\le \left(1 + \eta_{l_{\min}}^2/\alpha\right) |||\,v\,|||^2 + 2\, C_\eta \left(\eta_{l_{\min}}/\sqrt{\alpha}\right) \left(1 + \eta_{l_{\min}}^2/\alpha\right) |||\,v\,|||^2.
\end{aligned}
$$

Thus, we have verified the right inequality in (3.5) with Γ_o from Theorem 3.2.

Now, we concentrate on the lower bound Γ_u. The first inequality of the following estimate comes from (A.1) and the second one follows by the splitting $v = P_{l_{\min}}v + \sum_j Q_jv$ for $v \in V_l$:

$$
\begin{aligned}
|||\,v\,|||^2 &\le a(P_{l_{\min}}v, P_{l_{\min}}v) + \sum_{j=l_{\min}}^{l-1} a(Q_jv, Q_jv) \\
&\le a(v,v) + 2 \sum_{j=l_{\min}}^{l-1} |a(P_{l_{\min}}v, Q_jv)| + 2 \sum_{j=l_{\min}}^{l-2} \sum_{i=j+1}^{l-1} |a(Q_jv, Q_iv)| \,. \qquad \text{(A.2)}
\end{aligned}
$$

To proceed we supply Lemma A.1.

Lemma A.1 *Let P_l and Q_l be the projections defined in Section* 2 *and let a be the bilinear form* (2.8)*. If $i \neq j$ then*

$$
a(Q_iv, Q_jv) \le \left(\gamma_i \gamma_j/\alpha\right) a(v,v) \quad \textit{for all } v \in X.
$$

Further,

$$
a(P_lv, Q_jv) \le \left(1 + \gamma_l/\sqrt{\alpha}\right) \left(\gamma_j/\sqrt{\alpha}\right) a(v,v) \quad \textit{for } j \ge l \textit{ and for all } v \in X.
$$

Proof: Writing $\mathcal{A}$ for $K^*K + \alpha I$ we have that

$$
\begin{aligned}
a(Q_iv, Q_jv) &= \langle KQ_iv, KQ_jv\rangle_Y \\
&= \langle KQ_i\mathcal{A}^{-1/2}\mathcal{A}^{1/2}v\,,\, KQ_j\mathcal{A}^{-1/2}\mathcal{A}^{1/2}v\rangle_Y \\
&\le \Big(\|KQ_i\|\,\|KQ_j\|/\alpha\Big)\, a(v,v).
\end{aligned}
$$

Lemma 2.1 implies now the first statement. Taking into account that

$$\|K P_l \mathcal{A}^{-1/2}\| \leq \|K \mathcal{A}^{-1/2}\| + \|K(I - P_l)\mathcal{A}^{-1/2}\| \leq 1 + \gamma_l/\sqrt{\alpha}$$

one can prove the second statement analogously. ■

Applying Lemma A.1 to (A.2) we achieve

$$|||\, v \,|||^2 \leq a(v,v) + 2\left(1 + \eta_{l_{\min}}/\sqrt{\alpha}\right)\left(C_\eta\, \eta_{l_{\min}}/\sqrt{\alpha}\right) a(v,v) + 2\left(C_\eta^2\, \eta_{l_{\min}}^2/\alpha\right) a(v,v)\,.$$

Hence, Theorem 3.2 is proved.

References

[1] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Comm. Pure Appl. Math.*, 44:141–183, 1991.

[2] C. K. Chui. *An Introduction to Wavelets.* Wavelet Analysis and its Applications. Academic Press, Boston, 1992.

[3] C. K. Chui and E. Quak. Wavelets on bounded interval. In D. Braess and L. L. Schumaker, editors, *Numerical Methods in Approximation Theory, Vol. 9*, International Series of Numerical Mathematics, Vol. 105, pages 53–75. Birkhäuser Verlag, Basel, 1992.

[4] C. K. Chui and J. Z. Wang. On compactly supported spline wavelets. *Trans. Amer. Math. Soc.*, 330(2):903–915, 1992.

[5] P. G. Ciarlet. *The Finite Element Methods for Elliptic Problems.* Sudies in Mathematics and its Applications. North-Holland, Amsterdam, 1978.

[6] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *App. and Comp. Harm. Anal.*, 1(1):54–81, 1993.

[7] W. Dahmen, S. Prössdorf, and R. Schneider. Wavelet approximation methods for pseudodifferential equations II: Matrix compression and fast solution. *Adv. in Comp. Math.*, 1:259–335, 1993.

[8] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:906–966, 1988.

[9] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Series in Applied Mathematics. SIAM Publications, Philadelphia, 1992.

[10] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration.* Computer Science and Applied Mathematics. Academic Press, New York, 1975.

[11] P. Deuflhard and A. Hohmann. *Numerical Analysis: A First Course in Scientific Computation.* de Gruyter Textbook. de Gruyter, Berlin, New York, 1994.

[12] V. Dicken and P. Maaß. Wavelet-Galerkin methods for ill-posed problems. Technical report, Fachbereich Mathematik, Universität Potsdam, Germany, 1994.

[13] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *App. and Comp. Harm. Anal.*, 2(2):101–126, 1995.

[14] C. W. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind.* Pitman, Boston, 1984.

[15] W. Hackbusch. *Iterative solution of large sparse systems of equations.* Applied Mathematical Sciences. Springer-Verlag, New York, 1994.

[16] A. Harten and I. Yad-Shalom. Fast multiresolution algorithms for matrix-vector multiplication. *SIAM J. Numer. Anal.*, 31(4):1191–1218, 1994.

[17] J. T. King. Multilevel algorithms for ill-posed problems. *Numer. Math.*, 61:311–334, 1992.

[18] A. K. Louis. *Inverse und schlecht gestellte Probleme.* Studienbücher Mathematik. B. G. Teubner, Stuttgart, Germany, 1989. *English translation in preparation.*

[19] F. Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numer. Math.*, 28:329–341, 1977.

[20] P. Oswald. *Multilevel Finite Element Approximation: Theory and Applications.* Teubner Skripten zur Numerik. B. G. Teubner, Stuttgart, Germany, 1994.

[21] R. Plato and G. M. Vainikko. On the regularization of projection methods for solving ill-posed problems. *Numer. Math.*, 57:63–79, 1990.

[22] U. Rüde. *Mathematical and Computational Techniques for Multilevel Adaptive Methods.* Frontiers in Applied Mathematics. SIAM, Philadelphia, 1993.

[23] G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31:614–627, 1989.

[24] G. Strang and G. Fix. A Fourier analysis of the finite element variational method. In *Constructive Aspects of Functional Analysis*, Rome, 1973. Edizioni Cremonese.

[25] W. Sweldens and R. Piessens. Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions. *SIAM J. Numer. Anal.*, 31(4):1240–1264, 1994.

[26] X.-G. Xia and M. Z. Nashed. The Backus-Gilbert method for signals in reproducing Hilbert spaces and wavelet subspaces. *Inverse Problems*, 10(3):785–804, 1994.

[27] H. Yserentant. Old and new convergence proofs for multigrid methods. In *Acta Numerica*, pages 285–326. Cambridge University Press, New York, 1993.