BIT manuscript No.
(will be inserted by the editor)

Wavelet-based multilevel methods for linear ill-posed problems

E. Klann · R. Ramlau · L. Reichel

Received: date / Accepted: date

Abstract The representation of linear operator equations in terms of wavelet bases yields a multilevel framework, which can be exploited for iterative solution. This paper describes cascadic multilevel methods that employ conjugate gradient-type methods on each level. The iterations are on each level terminated by a stopping rule based on the discrepancy principle.

Keywords ill-posed problem · wavelet · multilevel method · minimal residual method

Mathematics Subject Classification (2000) 65R30 · 65R32 · 65T60 · 65N55

1 Introduction

Many problems in applied mathematics and engineering can be formulated as Fredholm integral equations of the first kind,

$$\int_a^b \kappa(t,s)x(s)ds = g(t), \qquad -\infty < a \le t \le b < \infty, \tag{1.1}$$

where the kernel κ and the right-hand side g are smooth, real-valued functions. For ease of notation, we consider in most of this paper integral equations in one

E. Klann
Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstr. 69, A-4040 Linz, Austria. E-mail: esther.klann@oeaw.ac.at

R. Ramlau
Johannes Kepler University of Linz, Institute for Industrial Mathematics, Altenbergerstr. 69, A-4040 Linz, Austria. E-mail: ronny.ramlau@jku.at

L. Reichel
Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu

space dimension; however, the methods described generalize to higher space dimensions by the use of tensor products. This is illustrated with a computed example in Section 4.

The determination of the solution x of (1.1) is an ill-posed problem in the sense of Hadamard, because i) the integral equation might not have a solution, ii) the solution might not be unique, and iii) the solution, if it exists and is unique, might not depend continuously on the data (the right-hand side). The computation of a meaningful approximate solution of (1.1) in finite precision arithmetic generally requires the use of numerical methods specifically designed for the solution of ill-posed problems; see, e.g., Engl et al. [9] and Groetsch [11] for discussions on ill-posed problems and methods for their solution.

We will assume that equation (1.1) is consistent and has a solution in the Hilbert space $\mathcal{L}_2([a,b])$ with norm $\|\cdot\|$. Often one is interested in determining the unique solution of minimal norm. We denote this solution by $\hat{x}$.

A common difficulty in applications is that the right-hand side function g in (1.1) is not available. Instead a corrupted version, which we denote by g^δ, is known. The error $e = g^\delta - g$ may stem from measurement or discretization errors and is referred to as "noise." We will assume that a bound δ for the norm of the error,

$$\|e\| \leq \delta, \tag{1.2}$$

is available. Then the discrepancy principle, described in Section 3, can be applied.

Our task is to determine an approximate solution x^δ of

$$\int_a^b \kappa(t,s)x(s)ds = g^\delta(t), \qquad a < t < b, \tag{1.3}$$

such that x^δ provides an accurate approximation of $\hat{x}$. Equation (1.3) is not required to be consistent; inconsistent equations are treated as least-squares problems.

It is convenient to write (1.1) and (1.3) in the form

$$Ax = g \tag{1.4}$$

and

$$Ax = g^\delta, \tag{1.5}$$

respectively, where $A : \mathcal{L}_2([\alpha,\beta]) \to \mathcal{L}_2([\alpha,\beta])$ is a compact operator. Thus, A has an unbounded inverse and may be singular. The right-hand side function g in (1.4) is assumed to be in the range of A, denoted by $\mathcal{R}(A)$.

We determine an approximation of the minimal-norm solution $\hat{x}$ of (1.4) by first replacing the operator A in (1.5) by an operator $A_{\rm reg}$ that approximates A and has a bounded inverse, and then solving the modified equation so obtained,

$$A_{\rm reg}x = g^\delta. \tag{1.6}$$

This replacement is referred to as regularization. The regularized operator A_{reg} should be chosen so that the solution x^δ of (1.6) is an accurate approximation of $\hat{x}$.

The possibly best understood regularization method is due to Tikhonov. In the simplest form of Tikhonov regularization, the inverse of A_{reg} is given by

$$(A_{\text{reg}})^{-1} = (A^*A + \lambda I)^{-1}A^*,$$

where A^* denotes the adjoint of A, I is the identity operator, and $\lambda > 0$ is a regularization parameter; see, e.g., Engl et al. [9] and Groetsch [11] for discussions on Tikhonov regularization. Several two- and multi-level methods for use in conjunction with Tikhonov regularization have been described in the literature; see, e.g., Chen et al. [4], Hanke and Vogel [13], Huckle and Staudacher [15], Jacobsen et al. [16], and King [18].

The multilevel methods of the present paper are applied to the unregularized problem (1.5). An approximate solution is determined on each level by a few iterations of a conjugate gradient-type method. Regularization is achieved by restricting the number of iterations on each level with the aid of the discrepancy principle. Thus, the operator A_{reg} is defined implicitly by the number of iterations carried out on each level, and by the restriction and extension operators applied during the course of the computations. The multilevel methods considered are cascadic, i.e., the computations proceed from coarser to finer levels. First an approximate solution is determined on the coarsest level by a conjugate gradient-type method. The iterations on this level are terminated by the discrepancy principle. The computed approximate solution so obtained is extended to the next finer level and there used as initial approximation for conjugate gradient-type iterations. The iterations are terminated by the discrepancy principle and the computed approximate solution so obtained is extended to the next finer level. The computations proceed in this fashion until an approximate solution of (1.5), that satisfies the discrepancy principle on the finest level, has been determined. The operator equation (1.5) is on each level represented in terms of a wavelet basis.

When the operator A and its restrictions to coarser levels are symmetric and positive semidefinite, we either can apply the Conjugate Residual (CR) or the MR-II iterative methods on each level. These are minimal residual methods of conjugate gradient-type. Their properties, when applied to ill-posed problems, have been analyzed by Hanke [12], Nemirovskii [26], and Plato [27]. The MR-II method also can be used when A and its restrictions to coarser levels are symmetric indefinite. Nonsymmetric operator equations are solved by CGNR, the conjugate gradient method applied to the associated normal equations. We show in Section 3 that our cascadic multiresolution methods based on these iterative methods are regularization methods in a well-defined sense.

Multilevel methods designed for the solution of the unregularized problem (1.5) also are considered in [7, 8, 22–24, 28, 30]. These methods differ from our approach in that they are not wavelet-based. Español [10] describes multilevel

methods based on wavelets for the solution of (1.5). The latter methods differ from our schemes in that full V-cycles are employed and the stopping criterion for the iterations on each level is not based on the discrepancy principle.

We remark that the design of cascadic multilevel methods of the present paper for the solution of ill-posed operator equations (1.5) differs significantly from the design of available cascadic multigrid methods for well-posed boundary value problems for elliptic partial differential equations. The latter kind of methods have been known for some time; see, e.g., Bornemann and Deuflhard [2]. The difference in design depends on that different smoothers have to be used to dampen highly oscillatory solution components and that the number of iterations on each level with multilevel methods for (1.5) has to be kept sufficiently small in order to avoid propagation of the error e in g^δ into the computed approximate solution. Smoothers are iterative methods constructed to dampen highly oscillatory eigenfunction components in the computed approximate solution. These eigenfunctions represent "noise." They are associated with eigenvalues close to the origin of compact operators, and generally with eigenvalues of largest magnitude of elliptic operators. This difference calls for the application of different smoothers for the damping of these eigenfunctions. In the present paper, we carry out smoothing by application of a few iterations with a conjugate gradient-type method on each level. The number of iterations performed is important. If too many iterations are carried out, then the error e in the right-hand side g^δ in (1.5) is propagated into the computed approximate solution and destroys the accuracy. Conversely, if too few iterations are performed, then unnecessarily low accuray in the computed approximate solution results. The main purpose of this paper is to determine a suitable number of iterations on each level. The combination of the discrepancy principle with estimation of the error in the right-hand side on each level makes it possible to prescribe how many iterations to carry out on each level, and to secure that the multilevel method is a regularization method. We note that this kind of analysis is not required for the design of multigrid methods for the solution of well-posed operator equations, such as boundary value problems for elliptic partial differential equations, and it has not been carried out for wavelet-based multilevel methods for the solution of ill-posed problems of the form (1.5).

This paper is organized as follows. Section 2 introduces the wavelet representations of (1.5) used on the different levels. The CGNR, CR, and MR-II iterative methods, as well as the discrepancy principle, are reviewed in Section 3, where also regularizing properties of the multilevel methods are discussed. Section 4 presents a few computed examples and Section 5 contains concluding remarks.

2 Wavelet representations

We describe wavelet representations of the operator A, the argument x, and the right-hand sides g and g^δ in equations (1.4) and (1.5), respectively. These

representations are the foundation for our multilevel methods. For a comprehensive treatment of wavelet analysis we refer the reader to [6,19,20,31].

2.1 Linear operator equations in a wavelet basis setting

We consider wavelet bases of $\mathcal{L}_2(\Omega)$ which are built by two types of functions: a scaling function ϕ and a wavelet function Ψ. The basis then consists of scaled and shifted versions, $\phi_{j,k}$ and $\Psi_{j,k}$, of these functions. For standard wavelet bases on $\mathbb{R}$, the index j denotes the scale and k denotes the space parameter,

$$\Psi_{j,k} := 2^{j/2}\Psi(2^j \cdot -k) \quad \text{and} \quad \phi_{j,k} := 2^{j/2}\phi(2^j \cdot -k).$$

We further assume that the set $\{\phi_{0,k}, \Psi_{j,k}\}_{j\geq 0, k\in\mathbb{Z}}$ forms an orthonormal basis for $\mathcal{L}_2(\mathbb{R})$. The wavelet expansion of a function $x \in \mathcal{L}_2(\mathbb{R})$ is then given by

$$x = \sum_{k\in\mathbb{Z}} \langle\, x, \phi_{0,k} \,\rangle \phi_{0,k} + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} \langle\, x, \Psi_{j,k} \,\rangle \Psi_{j,k} \ . \tag{2.1}$$

We further assume that the functions ϕ and Ψ have compact support; this holds, e.g., for the family of Daubechies wavelets [6]. In many applications the domain Ω is bounded. With some effort wavelet bases can be adapted to fairly general domains $\Omega \subset \mathbb{R}^d$; see [5] for a survey of these adaptations. The functions ϕ and Ψ may then change their form near the boundary of the domain. For simplicity we will consider here only the case $\Omega = [0,1]$. A wavelet basis for $\mathcal{L}_2([0,1])$ can be constructed as follows: Taking an orthonormal wavelet basis $\{\Phi_{0,k}, \Psi_{j,k} : j \geq 0, k \in \mathbb{Z}\}$ for $\mathcal{L}_2(\mathbb{R})$, we define

$$\Phi^{per}_{j,k}(t) = \sum_{l\in\mathbb{Z}} \Phi_{j,k}(t+l), \tag{2.2}$$

$$\Psi^{per}_{j,k}(t) = \sum_{l\in\mathbb{Z}} \Psi_{j,k}(t+l). \tag{2.3}$$

Then

$$\{\Phi^{per}_{0,0}\} \cup \{\Psi^{per}_{j,k} : j \in \mathbb{N}\ , k = 0, \ldots, 2^j - 1\} \tag{2.4}$$

is an orthonormal basis for $\mathcal{L}_2([0,1])$; see, e.g., [6]. Therefore, for any $x \in \mathcal{L}_2([0,1])$, we have the expansion

$$x = \langle\, x, \Phi^{per}_{0,0} \,\rangle \Phi^{per}_{0,0} + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} \langle\, x, \Psi^{per}_{j,k} \,\rangle \Psi^{per}_{j,k}\,. \tag{2.5}$$

The above described method to obtain a wavelet basis on the interval is computationally efficient, but has the disadvantage that an artificial discontinuity is introduced for non-periodic functions. This can be avoided by using, e.g., the construction by Meyer [21]. However, the methods described in our paper can be used with any construction for wavelet bases on the interval as long

as the resulting wavelet spaces $\mathcal{W}_j$ are finite dimensional. Application of the operator $A : \mathcal{L}_2([0,1]) \to \mathcal{L}_2([0,1])$ to (2.5) yields

$$Ax = \langle\, x, \Phi_{0,0}^{per} \,\rangle A\Phi_{0,0}^{per} + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \langle\, x, \Psi_{j,k}^{per} \,\rangle A\Psi_{j,k}^{per}.$$

On the other hand, expansions of the form (2.5) are valid for any element of $\mathcal{L}_2([0,1])$ and, in particular, we obtain from (1.4),

$$g = Ax = \langle\, Ax, \Phi_{0,0}^{per} \,\rangle \Phi_{0,0}^{per} + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \langle\, Ax, \Psi_{j,k}^{per} \,\rangle \Psi_{j,k}^{per}.$$

We will now express (1.4) as a linear system of equations with an infinite-dimensional matrix. For notational simplicity, we switch to a function system with only one index. To this end, we set

$$\begin{aligned} \phi_0(t) &= \Phi_{0,0}^{per}(t) \\ \phi_1(t) &= \Psi_{0,0}^{per}(t) \\ \phi_{2^j+k} &= \Psi_{j,k}^{per}(t), \qquad 0 \le k \le 2^j - 1, \qquad j = 0, 1, \dots \ . \end{aligned} \tag{2.6}$$

For a general orthonormal function system $\{\phi_m\}_{m\in\mathbb{N}^+}$, we have

$$\begin{aligned} x &= \sum_l \langle\, x, \phi_l \,\rangle \phi_l, \\ Ax &= \sum_l \langle\, x, \phi_l \,\rangle A\phi_l. \end{aligned}$$

For the kth wavelet coefficient of $g = Ax$, we obtain

$$g_k = \langle\, Ax, \phi_k \,\rangle = \sum_l \langle\, x, \phi_l \,\rangle \langle\, A\phi_l, \phi_k \,\rangle.$$

Introduce

$$\mathbf{A} = [a_{kl}]_{k,l}, \quad \mathbf{x} = [x_l]_l, \quad \text{and} \quad \mathbf{g} = [g_k]_k,$$

where

$$a_{kl} = \langle\, A\phi_l, \phi_k \,\rangle, \quad x_l = \langle\, x, \phi_l \,\rangle, \quad \text{and} \quad g_k = \langle\, g, \phi_k \,\rangle,$$

for $k, l = 0, 1, \dots$. The operator equation (1.4) now can be written as an (infinite-dimensional) matrix-vector relation between $\mathbf{x}$ and $\mathbf{g}$,

$$\mathbf{A}\mathbf{x} = \mathbf{g}.$$

The same considerations yield a matrix representation for the adjoint operator A^*.

Remark 2.1 The operator A^* has a matrix representation given by $\mathbf{A}^T$.

2.2 Definition of the restrictions of A and $\mathbf{A}$

Equivalently to the wavelet expansion (2.1) of a function $x \in \mathcal{L}_2([0,1])$, we consider a decomposition of $\mathcal{L}_2([0,1])$ itself. Define the spaces $\mathcal{V}_0^{per} = \text{span}\{\phi_{0,0}^{per}\}$ and $\mathcal{W}_j^{per} = \text{span}\{\Psi_{j,k}^{per}; 0 \le k \le 2^j - 1\}$ for $j \ge 0$. Due to the fact that the set $\{\phi_{0,0}^{per}, \Psi_{j,k}^{per}\}_{j \ge 0, 0 \le k \le 2^j - 1}$ constitutes an orthonormal wavelet basis for $\mathcal{L}_2([0,1])$, we know that for fixed j, the function system $\{\Psi_{j,k}^{per}\}_{k \in \mathbb{Z}}$ forms a basis for the space $\mathcal{W}_j^{per}$. In particular, the decomposition

$$\mathcal{L}_2([0,1]) = \mathcal{V}_0^{per} \oplus \bigoplus_{j=0}^{\infty} \mathcal{W}_j^{per}$$

holds, and setting $\mathcal{W}_{-1}^{per} = \mathcal{V}_0^{per}$, we obtain

$$\mathcal{L}_2([0,1]) = \bigoplus_{j=-1}^{\infty} \mathcal{W}_j^{per}.$$

Now we can introduce a family of restrictions of the operator A by using the multilevel structure of the wavelet setting.

Definition 2.1 For $i = -1, 0, 1, \dots$, define the spaces

$$\mathcal{U}_i = \bigoplus_{l=-1}^{i} \mathcal{W}_l^{per} \quad \subset \ \mathcal{L}_2([0,1]). \tag{2.7}$$

Let $P_i : \mathcal{L}_2([0,1]) \to \mathcal{U}_i$ denote the orthogonal projection onto $\mathcal{U}_i$ and introduce the restricted operators $A_i : \mathcal{U}_i \to \mathcal{U}_i$ by

$$A_i := P_i A P_i$$

for $i = -1, 0, 1, \dots$.

Lemma 2.1 *The projection P_i on the space $\mathcal{U}_i$ has the representation*

$$\begin{aligned} P_i x &= \langle x, \Phi_{0,0}^{per} \rangle \Phi_{0,0}^{per} + \sum_{l=0}^{i} \sum_{k \le 2^l - 1} \langle x, \Psi_{l,k}^{per} \rangle \Psi_{l,k}^{per} \\ &= \sum_{n=0}^{2^{i+1}-1} \langle x, \phi_n \rangle \phi_n \ . \end{aligned}$$

Proof The assertion follows directly with the help of the index transformation (2.6) and from the fact that the functions $\Psi_{j,k}^{per}, k \le 2^j - 1$, form a basis for $\mathcal{W}^{per}$.

Proposition 2.1 *The matrix representation $\mathbf{A}_j = [a_{kl}^j]$ of the restricted operator A_j has the entries*

$$a_{kl}^j = \begin{cases} a_{kl} & \text{for} \quad k, l = 0, \cdots, 2^j - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

Proof We have

$$(P_j x)_l = \langle\, P_j x, \phi_l \,\rangle = \begin{cases} 0 & \text{for } l > 2^j - 1, \\ \langle\, x, \phi_l \,\rangle = x_l & \text{for } l \le 2^j - 1. \end{cases}$$

Hence,

$$(AP_j x)_k = \sum_{l=0}^{\infty} a_{kl} (P_j x)_l = \sum_{l=0}^{2^j-1} a_{kl} x_l = \sum_{l=0}^{\infty} a_{kl}^j x_l. \tag{2.9}$$

By assumption,

$$(P_j A P_j x)_k = (\mathbf{A}_i \mathbf{x})_k = \sum_{l=0}^{\infty} a_{kl}^j x_l.$$

Since $(P_j y)_k = 0$ for all $k > 2^j - 1$, it follows that $(P_j A P_j x)_k = 0$ for $k > 2^j - 1$, i.e.,

$$\sum_{l=0}^{\infty} a_{kl}^j x_l = 0 \quad \text{for } k > 2^j - 1.$$

Hence, $a_{kl}^j = 0$ for $k > m$.

Lemma 2.2 *The matrix representation of the restricted adjoint operators A_i^* is given by $\mathbf{A}_i^T$.*

Proof Since P_i is an orthogonal projection, we have $P_i^* = P_i$. Therefore,

$$A_i^* = (P_i A P_i)^* = P_i A^* P_i.$$

The same arguments as in the proof of Proposition 2.1 now shows the lemma.

Our cascadic multilevel method proceeds from the coarsest to the finest level. A computed approximation on a coarse level has to be mapped onto the next finer level by *extension* (prolongation). The mapped approximation is then used as initial approximation for the computation on this level. In the proposed wavelet setting (2.7), this means that we need to extend an element from the space $\mathcal{U}_{i-1}$, so that it belongs to the space $\mathcal{U}_i$. Note that

$$x \in \mathcal{U}_{i-1} \quad \Leftrightarrow \quad x = \sum_{j=-1}^{i-1} \sum_{k \le 2^j - 1} x_{j,k} \Psi_{j,k}^{per}.$$

Definition 2.2 The extension operator $E_i : \mathcal{U}_{i-1} \to \mathcal{U}_i$ is defined by

$$E_i x := \sum_{j=-1}^{i} \sum_{k \le 2^j - 1} \tilde{x}_{j,k} \Psi_{j,k}^{per}$$

with

$$\tilde{x}_{j,k} = \begin{cases} x_{j,k} & \text{for } j \le i-1, \quad k \le 2^j - 1, \\ 0 & \text{for } j = i, \quad k \le 2^i - 1. \end{cases}$$

We have $E_i = I$ in the sense of infinite functional extensions. However, in a finite-dimensional setting, it holds

$$E_i : \mathbb{R}^{2^{i-1}-1} \to \mathbb{R}^{2^i-1}.$$

It is convenient to require the extension operator to satisfy the condition

$$\mathcal{R}(E_i) \subset \overline{\mathcal{R}(A_i^*)} = \mathcal{N}(A_i)^\perp, \tag{2.10}$$

which secures that we do not get a solution component in the null space of the operator when we go to the next larger space. This is important if we want to determine the minimal norm solution.

Proposition 2.2 *The extension operator E_i restricted to $\overline{\mathcal{R}(A_{i-1}^*)} \subset \mathcal{U}_i$ satisfies condition (2.10), i.e.,*

$$\mathcal{R}(E_{i|\overline{\mathcal{R}(A_{i-1}^*)}}) \subset \overline{\mathcal{R}(A_i^*)} = \mathcal{N}(A_i)^\perp.$$

Proof Consider the extension operator on $\mathcal{R}(A_{i-1}^*)$,

$$E_i : \overline{\mathcal{R}(A_{i-1}^*)} \to \mathcal{U}_i.$$

Since $A_{i-1}^* = P_{i-1}A^*P_{i-1}$ and $\mathcal{R}(P_{i-1}) = \mathcal{U}_{i-1} \subset \mathcal{U}_i = \mathcal{R}(P_i)$, it follows that

$$\mathcal{R}(A^*P_{i-1}) \subset \mathcal{R}(A^*P_i). \tag{2.11}$$

We have $\mathcal{R}(A_i^*) = \mathcal{R}(A^*P_i) \cap \mathcal{U}_i$ and $\mathcal{R}(E_i) = \mathcal{R}(A_{i-1}^*) \cap \mathcal{U}_i$. Therefore, we obtain from (2.11) that

$$\mathcal{R}(E_i) \subset \mathcal{R}(A_i^*) \cap \mathcal{U}_i = \mathcal{R}(A_i^*).$$

Remark 2.2 We will see below that in our multilevel methods the computed approximate solutions $x_{i,k_i^{\delta_i}}^{\delta_i}$ on level i, indeed, are elements of $\mathcal{R}(A_i^*)$, provided the iterative method used is either CGNR or MR-II, and the initial approximate solution $x_0^{\delta_0}$ is chosen appropriately; see Algorithm 3.2 for the notation.

2.3 Data approximation

As already mentioned in Section 1, the right-hand side g of (1.4) is assumed not to be available. Instead a noise-contaminated approximation g^δ of g is explicitly known. We assume that a bound, δ, for the norm of the noise, is available; cf. (1.2). For the multilevel method, we need bounds for $\|g_i - g_i^\delta\|$, where $g_i = P_i g$ and $g_i^\delta = P_i g^\delta$ are the projections onto $\mathcal{U}_i$ of the exact and

noise-contaminated right-hand sides of (1.4) and (1.5), respectively. Here $\|\cdot\|$ denotes the Euclidean norm on level i. Such a bound is furnished by

$$\|g_i - g_i^\delta\|^2 = \sum_{j=-1}^{i} \sum_{k \le 2^j - 1} |\langle\, g - g^\delta, \Psi_{j,k}\,\rangle|^2$$
$$= \sum_{j=-1}^{\infty} \sum_{k \le 2^j - 1} |\langle\, g - g^\delta, \Psi_{j,k}\,\rangle|^2 \le \delta^2. \tag{2.12}$$

However, in general this bound is quite crude. An approach to estimating the norm of the noise in g_i^δ based on coefficient-wise soft shrinkage is presented in Section 4. In order to express a possible link between the noise and the projection level, we use the notation $g_i^{\delta_i} := P_i g^\delta$, and we seek to determine bounds of the form

$$\|g_i - g_i^{\delta_i}\| \le \delta_i, \qquad i = 1, 2, \dots \ . \tag{2.13}$$

3 Iterative methods

We first review regularization properties of the CGNR, CR, and MR-II iterative methods shown by Hanke [12], Nemirovskii [26], and Plato [27], and then discuss their use in multilevel schemes.

3.1 Conjugate gradient-type methods

This subsection reviews results for one-level CGNR, CR, and MR-II iterative methods applied to the approximate solution of (1.5). Let x_0^δ denote the initial approximate solution and define the associated residual error $r_0^\delta = g^\delta - Ax_0^\delta$.

Assume that the operator A is non-selfadjoint, and let A^* denote its adjoint. CGNR is the conjugate gradient method applied to the normal equations

$$A^*Ax = A^*g^\delta \tag{3.1}$$

associated with (1.5). Introduce the Krylov subspaces

$$\mathcal{K}_k(A^*A, A^*r_0^\delta) = \mathrm{span}\{A^*r_0^\delta, (A^*A)A^*r_0^\delta, \dots, (A^*A)^{k-1}A^*r_0^\delta\}, \quad k = 1, 2, \dots \ .$$

The kth iterate, x_k^δ, determined by CGNR applied to (1.5) with initial iterate x_0^δ satisfies

$$\|Ax_k^\delta - g^\delta\| = \min_{x \in x_0^\delta + \mathcal{K}_k(A^*A, A^*r_0^\delta)} \|Ax - g^\delta\|, \ \ x_k^\delta \in \ x_0^\delta + \mathcal{K}_k(A^*A, A^*r_0^\delta), \tag{3.2}$$

which shows that CGNR is a minimal residual method; see, e.g., Björck [1] for implementations. Note that if x_0^δ is orthogonal to $\mathcal{N}(A)$, then so are all the iterates x_k^δ, $k = 1, 2, \dots$. This is the case, e.g., when $x_0^\delta = 0$. If in addition g^δ is noise-free, i.e., if g^δ is replaced by g in (3.1) and (3.2), then the iterates

x_k^δ determined by CGNR converge to $\hat{x}$, the minimal-norm solution of (1.4), when k increases. The evaluation of x_k^δ requires k applications of the operator A and k applications of the adjoint A^*.

The residual error $r = g^\delta - Ax$ is sometimes referred to as the discrepancy associated with x. The *discrepancy principle* furnishes a criterion for choosing the number of CGNR iterations. Let g^δ satisfy (1.2) for some $\delta \geq 0$, and let $\tau > 1$ be a constant independent of δ. Then x is said to satisfy the discrepancy principle if

$$\|g^\delta - Ax\| \leq \tau\delta. \tag{3.3}$$

Stopping Rule 3.1 *Let g^δ, δ, and τ be the same as in (3.3). Terminate the iterations when, for the first time,*

$$\|g^\delta - Ax_k^\delta\| \leq \tau\delta. \tag{3.4}$$

We denote the stopping index by k^δ.

We are interested in the behavior of the iterates $x_{k^\delta}^\delta$ when reducing $\delta > 0$, while keeping $\tau > 1$ fixed. Note that k^δ typically increases when δ is decreased to zero. We remark that Stopping Rule 3.1 is only meaningful when the associated noise-free problem (1.4) is consistent.

An iterative method equipped with Stopping Rule 3.1 is said to be a *regularization method* if the computed iterates $x_{k^\delta}^\delta$ satisfy

$$\lim_{\delta \searrow 0} \sup_{\|g - g^\delta\| \leq \delta} \|\hat{x} - x_{k^\delta}^\delta\| = 0, \tag{3.5}$$

where $\hat{x}$ is the minimal-norm solution of (1.4). The constant $\tau > 1$ in Stopping Rule 3.1 is kept fixed as δ is decreased to zero. The following proposition shows that CGNR is a regularization method when applied to the solution of (1.5). Proofs are provided by Nemirovskii [26] and Hanke [12, Theorem 3.12].

Proposition 3.1 *Let $x_0^\delta \in \mathcal{N}(A)^\perp$, let equation (1.4) be consistent, and assume that g^δ satisfies (1.2) for some $\delta > 0$. Terminate the CGNR-iterations according to Stopping Rule 3.1 with $\tau > 1$ a fixed constant. Let k^δ denote the stopping index and $x_{k^\delta}^\delta$ the associated iterate. Then $x_{k^\delta}^\delta \to \hat{x}$ as $\delta \searrow 0$, where $\hat{x}$ denotes the minimal-norm solution of (1.4).*

We turn to the situation when the operator A is self-adjoint and possibly indefinite. The iterates x_k^δ, $k = 1, 2, \ldots$, determined by the MR-II method satisfy

$$\|Ax_k^\delta - g^\delta\| = \min_{x \in x_0^\delta + \mathcal{K}_k(A, Ar_0^\delta)} \|Ax - g^\delta\|, \qquad x_k^\delta \in x_0^\delta + \mathcal{K}_k(A, Ar_0^\delta). \tag{3.6}$$

Thus, MR-II is a minimal residual method. Moreover, $x_0^\delta \in \mathcal{N}(A)^\perp$ implies that $x_k^\delta \in \mathcal{N}(A)^\perp$ for $k \geq 1$. The iterates x_k^δ can be computed with recursion formulas with few terms, similarly as for the conjugate gradient methods. Implementations are provided in [3,12]. The computation of the iterate x_k^δ requires $k+1$ applications of the operator A. The following analog of Proposition 3.1 is shown by Hanke [12, Theorem 6.15].

Proposition 3.2 *Let the operator A be self-adjoint, let equation (1.4) be consistent, and assume that g^δ satisfies (1.2) for some $\delta > 0$. Let $x_0^\delta \in \mathcal{N}(A)^\perp$ and let the iterates x_k^δ be generated by MR-II. Terminate the iterations according to Stopping Rule 3.1 with $\tau > 1$ a fixed constant. Let k^δ denote the stopping index and $x_{k^\delta}^\delta$ the associated iterate. Then $x_{k^\delta}^\delta$ converges to $\hat{x}$, the minimal norm solution of (1.4), as $\delta \searrow 0$.*

When A is self-adjoint and positive semidefinite, the CR method can be applied. The kth iterate determined by this method satisfies

$$\|Ax_k^\delta - g^\delta\| = \min_{x \in x_0^\delta + \mathcal{K}_k(A, r_0^\delta)} \|Ax - g^\delta\|, \qquad x_k^\delta \in x_0^\delta + \mathcal{K}_k(A, r_0^\delta), \tag{3.7}$$

which shows that CR is a minimal residual method, but the Krylov subspace is different than for MR-II. An implementation is described in [12,29]. The computation of x_k^δ requires k applications of A. Proofs that CR is a regularization method when the iterates are terminated by Stopping Rule 3.1 are provided in [12,26,27]. The computed solution may have a component in $\mathcal{N}(A)$.

3.2 Cascadic multilevel methods

This section describes cascadic multilevel methods that are based on wavelet-decompositions of $\mathcal{L}_2([\alpha, \beta])$ and the CGNR, CR, or MR-II iterative methods. Recall from Section 2 that the operators P_i are orthogonal projections, for $1 \leq i \leq \ell$,

$$P_i : \mathcal{L}_2([a, b]) \to \mathcal{U}_i, \quad A_i := P_i A P_i, \quad g_i := P_i g, \quad g_i^{\delta_i} := P_i g^\delta, \tag{3.8}$$

where $\mathcal{U}_1 \subset \mathcal{U}_2 \subset \ldots \subset \mathcal{U}_\ell$ form a sequence of nested linear subspaces of $\mathcal{L}_2([\alpha, \beta])$. We also need the extension operators E_i, $2 \leq i \leq \ell$, of Definition 2.2 with the spaces $\mathcal{U}_i$ as above. We will use the inequalities (2.13). It follows from (2.12) that they hold for all $\delta_i = \delta \geq 0$.

Let IM denote one of the iterative methods CGNR, CR, or MR-II. The cascadic multilevel methods described by Algorithm 3.2 below first determine an approximate solution of $A_1 x = g_1^{\delta_1}$ in $\mathcal{U}_1$ by IM. The iterations with IM are terminated as soon as an iterate that satisfies a stopping rule related to the discrepancy principle has been computed. This iterate is mapped from $\mathcal{U}_1$ into $\mathcal{U}_2$ by the extension operator E_2. We then apply IM to compute a correction in $\mathcal{U}_2$ of this mapped iterate. Again, the IM iterations are terminated by a stopping rule related to the discrepancy principle. The approximate solution in $\mathcal{U}_2$ determined in this fashion is mapped into $\mathcal{U}_3$ by E_3. The computations are continued in this manner until an approximation of $\hat{x}$ has been computed in $\mathcal{U}_\ell$. The operator E_1 in the algorithm only is applied to the zero-vector and is assumed to satisfy $E_1(0) = 0 \in \mathcal{U}_1$.

Algorithm 3.2 *Multilevel Methods*

Input: A, g^δ, $\ell \geq 1$ (number of levels), $\delta_1, \delta_2, \ldots, \delta_\ell$, $c_1, c_2, \ldots, c_\ell$ (coefficients for the stopping rule);
Output: approximate solution $x_\ell^\delta \in \mathcal{U}_\ell$ of (1.4);

$x_0^{\delta_0} := 0$;
for $i := 1, 2, \ldots, \ell$ *do*
$x_{i,0}^{\delta_{i-1}} := E_i x_{i-1}^{\delta_{i-1}}$;
$\Delta x_{i,k_i^{\delta_i}}^{\delta_i} := IM(A_i, g_i^{\delta_i} - A_i x_{i,0}^{\delta_{i-1}})$;
$x_{i,k_i^{\delta_i}}^{\delta_i} := x_{i,0}^{\delta_{i-1}} + \Delta x_{i,k_i^{\delta_i}}^{\delta_i}$;
endfor
$x_\ell^\delta := x_{\ell,k_\ell^{\delta_\ell}}^{\delta_\ell}$; □

In the above algorithm,

$$\Delta x_{i,k_i^{\delta_i}}^{\delta_i} := \mathrm{IM}(A_i, g_i^{\delta_i} - A_i x_{i,0}^{\delta_{i-1}})$$

denotes the computation of the approximate solution $\Delta x_{i,k_i^{\delta_i}}^{\delta_i}$ of the equation

$$A_i z = g_i^{\delta_i} - A_i x_{i,0}^{\delta_{i-1}} \tag{3.9}$$

by application of $k_i^{\delta_i}$ steps of one of the iterative methods CGNR, CR, or MR-II, with initial iterate $\Delta x_{i,0} = 0$.

Lemma 3.1 *The approximate solutions $x_{i,k_i^{\delta_i}}^{\delta_i}$ computed by Algorithm 3.2 with the CGNR or MR-II method satisfy*

$$x_{i,k_i^{\delta_i}}^{\delta_i} \in \mathcal{N}(A_i)^\perp, \qquad i = 1, 2, \ldots, \ell.$$

Proof For the standard (one-level) CGNR and MR-II iterative methods applied to the approximate solution of (1.5), it is known that $x_0^\delta \in \mathcal{N}(A)^\perp$ implies $x_k^\delta \in \mathcal{N}(A)^\perp$ for all $k = 1, 2, \ldots$. In Algorithm 3.2, we have $x_0^{\delta_0} := 0$. Let $i = 1$ in the algorithm. Then $x_{1,0}^{\delta_0} := E_1 x_0^{\delta_0} = 0 \in \mathcal{U}_1 \cap \mathcal{N}(A_1)^\perp$. The correction $\Delta x_{1,k_1^{\delta_1}}^{\delta_1}$ is computed by either CGNR or MR-II from the initial iterate $\Delta x_{1,0} = 0 \in \mathcal{N}(A_1)^\perp$. It follows that $\Delta x_{1,k_1^{\delta_1}}^{\delta_1} \in \mathcal{N}(A_1)^\perp$ and, therefore,

$$x_{1,k_1^{\delta_1}}^{\delta_1} \in \mathcal{N}(A_1)^\perp.$$

Since $\mathcal{N}(A_1)^\perp = \overline{\mathcal{R}(A_1^*)}$, we obtain from Proposition 2.2 that $x_{2,0}^{\delta_1} := E_2 x_1^{\delta_1} \in \mathcal{N}(A_2)^\perp$. The correction $\Delta x_{2,k_2^{\delta_2}}^{\delta_2}$ of $x_{2,0}^{\delta_1}$ is computed by either CGNR or MR-II from the initial iterate $\Delta x_{2,0} = 0 \in \mathcal{N}(A_2)^\perp$, and it follows that $x_{2,k_2^{\delta_2}}^{\delta_2} \in \mathcal{N}(A_2)^\perp$. We obtain that $x_{i,k_i^{\delta_i}}^{\delta_i} \in \mathcal{N}(A_i)^\perp$ for $i = 1, 2, \ldots, \ell$.

The number of iterations on level i, $k_i^{\delta_i}$, is determined by the discrepancy principle, whose application can be justified theoretically when the noise-free equation associated with (3.9) is consistent. We briefly discuss when we can expect the projected error-free equation

$$A_i z = g_i \tag{3.10}$$

to be consistent. The following example illustrates that the projected equation is not consistent for all orthogonal projections P_i.

Example 3.1. Let $A, P_1 \in \mathbb{R}^{2\times 2}$ and $g \in \mathbb{R}^2$ be given by

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad P_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Then

$$A_1 = P_1 A P_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad g_1 = P_1 g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Hence, the equation $A_1 x = g_1$ is inconsistent. □

Whether the projected equation is consistent depends not only on A and P_i, but also on the right-hand side g.

Example 3.2. Let A and P_1 be the same as in Example 3.1, and assume that $g = [0, 1]^T \in \mathbb{R}^2$. Then the projected equation (3.10) is consistent. □

While symmetry of A is not sufficient to guarantee consistency of the projected equation (3.10), symmetry and positive semidefiniteness is.

Theorem 3.3 *Let the operator A in the consistent equation (1.4) be symmetric and positive semidefinite. Then the projected equation (3.10) is consistent.*

Proof Introduce the spectral decomposition

$$A = W \Lambda W^*,$$

where Λ is a diagonal operator and W is orthogonal. Let $\hat{x} = W^* x$ and $\hat{g} = W^* g$. Then (1.4) can be expressed as $\Lambda \hat{x} = \hat{g}$.

The projected equation (3.10) can be expressed as

$$P_i W \Lambda W^* P_i x_i = P_i W \hat{g} = P_i W \Lambda \hat{x}. \tag{3.11}$$

Since P_i is an orthogonal projection, so is $\hat{P}_i = W^* P_i W$. We obtain from (3.11) that

$$\hat{P}_i \Lambda \hat{P}_i \hat{x}_i = \hat{P}_i \Lambda \hat{x}.$$

These are the normal equations associated with the least-squares problem

$$\min_z \|\Lambda^{1/2} \hat{P}_i z - \Lambda^{1/2} \hat{x}\|.$$

The normal equations associated with any equation or least-squares problem are consistent. This shows the theorem.

Example 3.3. Let $A = P_i$. Then A is self-adjoint and positive semidefinite. Consistency of (1.4) implies consistency of (3.10). □

Example 3.4. Let $A = I - P_i$. Consistency of (1.4) implies that $g \in \mathcal{U}_i^{\perp}$. Therefore, the projected equation (3.10) simplifies to the consistent equation $0z = 0$. □

For nonsymmetric operators A, we have the following result.

Theorem 3.4 *Let the operator A_i and right-hand side g_i be defined by (3.8), and let $\mathcal{U}_i = \mathcal{R}(P_i)$. Assume that*

$$A^*\mathcal{U}_i = \mathcal{U}_i. \tag{3.12}$$

Then equation (3.10) is consistent.

Proof Let $\{\phi_j^{(i)}\}_{j=1}^{\infty}$ be an orthonormal basis for $\mathcal{U}_i$ and define the operator $U_i : \ell_2 \to \mathcal{U}_i$ and its adjoint $U_i^* : \mathcal{U}_i \to \ell_2$ by

$$U_i c = \sum_{j=1}^{\infty} (e_j^T c)\phi_j^{(i)}, \qquad c \in \ell_2,$$

$$U_i^* f = \sum_{j=1}^{\infty} \langle\, \phi_j^{(i)}, f \,\rangle e_j, \qquad f \in \mathcal{U}_i,$$

where $e_j = [0, \ldots, 0, 1, 0, 0, \ldots\ ]^T \in \ell_2$ denotes the jth axis vector. Note that $U_i^* U_i = I$ and $P_i = U_i U_i^*$.

Define the operator $\hat{U}_i : \ell_2 \to (A^*\mathcal{U}_i)^{\perp}$ by

$$\hat{U}_i c = \sum_{k=1}^{\infty} (e_k^T c)\hat{\phi}_k^{(i)}, \qquad c \in \ell_2,$$

where $\{\hat{\phi}_k^{(i)}\}_{k=1}^{\infty}$ is an orthonormal basis for $(A^*\mathcal{U}_i)^{\perp}$. In particular,

$$\langle\, A^*\phi_j^{(i)}, \hat{\phi}_k^{(i)} \,\rangle = \langle\, \phi_j^{(i)}, A\hat{\phi}_k^{(i)} \,\rangle = 0, \qquad j, k = 1, 2, \ldots\ . \tag{3.13}$$

It follows from (3.12) that $\{\phi_1^{(i)}, \hat{\phi}_1^{(i)}, \phi_2^{(i)}, \hat{\phi}_2^{(i)}, \phi_3^{(i)}, \hat{\phi}_3^{(i)}, \ldots\ \}$ is a basis for $\mathcal{L}_2([\alpha, \beta])$. Therefore the operator $\breve{U}_i : \ell_2 \to \mathcal{L}_2([\alpha, \beta])$ given by

$$\breve{U}_i c = \sum_{j=1}^{\infty} \left((e_{2j-1}^T c)\phi_j^{(i)} + (e_{2j}^T c)\hat{\phi}_j^{(i)} \right), \qquad c \in \ell_2,$$

is invertible.

The adjoints of $\hat{U}_i$ and $\breve{U}_i$ are

$$\hat{U}_i^* f = \sum_{k=1}^{\infty} \langle\, \hat{\phi}_k^{(i)}, f \,\rangle e_k,$$

$$\breve{U}_i^* f = \sum_{k=1}^{\infty} \left(\langle\, \phi_k^{(i)}, f \,\rangle e_{2k-1} + \langle\, \hat{\phi}_k^{(i)}, f \,\rangle e_{2k} \right).$$

Since $\breve{U}_i$ is invertible, the equation

$$\breve{U}_i^* A \breve{U}_i c = \breve{U}_i^* g, \qquad x = \breve{U}_i c, \tag{3.14}$$

is equivalent to (1.4) and, therefore, consistent. Further, in view of (3.13), we have

$$U_i^* A \hat{U}_i = 0.$$

Combining this equation with (3.14) shows that

$$U_i^* A U_i c = U_i^* g$$

is consistent. This equation is equivalent to (3.10).

Remark 3.1 In finite dimensions, an orthonormal basis for the space $A^*\mathcal{U}_i$ easily can be computed by QR-factorization of the matrix A^*U_i. Let $A \in \mathbb{R}^{m\times m}$ and $U_i \in \mathbb{R}^{m\times k}$ with $m > k$. Assume, analogously to the proof of Theorem 3.4, that U_i has orthonormal columns. Consider the QR-factorization

$$A^*U_i = Q \begin{bmatrix} R_i \\ 0 \end{bmatrix},$$

where A^* is the transpose of A, $Q \in \mathbb{R}^{m\times m}$ is orthogonal, and $R_i \in \mathbb{R}^{k\times k}$ is upper triangular. Partition $Q = [Q_1\ Q_2]$ so that Q_1 consists of the first k columns. These columns form an orthonormal basis for $\mathcal{R}(A^*U_i)$ and the columns of Q_2 form an orthonormal basis for $\mathcal{R}^\perp(A^*U_i)$.

Example 3.5. Let P_1 and g be the same as in Example 3.1, and assume that

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then $A\mathbb{R}^2 \neq \mathbb{R}^2$. Moreover, A_1 and g_1 are the same as in Example 3.1 and, therefore, the projected equation (3.10) is not consistent. □

Returning to Algorithm 3.2, we conclude from the above discussion that the equation

$$A_i z = g_i - A_i x_{i,0}^{\delta_{i-1}}, \tag{3.15}$$

associated with (3.9) might not be consistent.

Introduce the orthogonal projection onto the null space of A_i,

$$P_{\mathcal{N}(A_i)} : \mathcal{U}_i \to \mathcal{N}(A_i). \tag{3.16}$$

In our experience, the inconsistency error, $P_{\mathcal{N}(A_i)} g_i$, is for many problems smaller than the norm of the error in the available right-hand side $g_i^{\delta_i}$ and, therefore, generally can be ignored in the design of multilevel methods. For instance, this is typically the case in the common situation when A_i is severely ill-conditioned but nonsingular, as well as when A is symmetric and semidefinite; cf. Theorem 3.3.

In order to show convergence of the approximate solutions determined by Algorithm 3.2 on each level, we have to require the equations (3.15) to be

consistent for all i. We will discuss how this requirement can be relaxed at the end of this section. Thus, assume for now that the equations (3.15) are consistent for $1 \leq i \leq \ell$, and let equation i have minimal-norm solution

$$\Delta x_i = x_i - x_{i,0}^{\delta_{i-1}}. \tag{3.17}$$

Algorithm 3.2 implicitly defines a regularized operator A_{reg} by carrying out $k_i^{\delta_i}$ IM iterations on level i for $i = 1, 2, \ldots, \ell$. It is important not to compute too many iterations on each level because, in general, the iterates $\Delta x_{i,k_i}^{\delta_i}$ do not converge to the minimal-norm solution (3.17) of (3.15) as the number of iterations, k_i, increases without bound. We use the following stopping rule on each level.

Stopping Rule 3.5 *Let the δ_i be the same as in (2.13) and denote the iterates determined on level i by IM applied to the solution of (3.9) by $\Delta x_{i,k_i}^{\delta_i}$, $k_i = 1, 2, \ldots$, with initial iterate $\Delta x_{i,0} = 0$. Terminate the iterations as soon as an iterate $\Delta x_{i,k_i}^{\delta_i}$ such that*

$$\|g_i^{\delta_i} - A_i x_{i,0}^{\delta_{i-1}} - A_i \Delta x_{i,k_i}^{\delta_i}\| \leq \tau \delta_i, \tag{3.18}$$

has been computed, where $\tau > 1$ is a constant independent of the δ_i. We denote the termination index by $k_i^{\delta_i}$ and the corresponding iterate by $\Delta x_{i,k_i^{\delta_i}}^{\delta_i}$.

The following theorem discusses convergence of the approximate solution $x_{i,k_i^{\delta_i}}^{\delta_i}$ determined by Algorithm 3.2 on level i towards the minimal-norm solution x_i of the noise-free projected problems (3.10) as $\delta_i \searrow 0$ for $1 \leq i \leq \ell$.

Theorem 3.6 *Let $A_\ell = A$ and $g_\ell = g$. Assume that the equations (3.10) are consistent for $1 \leq i \leq \ell$ and that (2.10) holds. Let the projected contaminated right-hand sides $g_i^{\delta_i}$ satisfy (2.13). Let IM in Algorithm 3.2 stand for CGNR or MR-II. Terminate the iterations with IM in Algorithm 3.2 on levels $1, 2, \ldots, \ell$ according to Stopping Rule 3.5. This yields the iterates $x_{i,k_i^{\delta_i}}^{\delta_i}$ for levels $1 \leq i \leq \ell$. Then the multilevel method described by Algorithm 3.2 is a regularization method on each level, i.e.,*

$$\lim_{\delta_i \searrow 0} \sup_{\|g_i - g_i^{\delta_i}\| \leq \delta_i} \|x_i - x_{i,k_i^{\delta_i}}^{\delta_i}\| = 0, \qquad 1 \leq i \leq \ell, \tag{3.19}$$

where x_i is the minimal norm solution of (3.10) with $x_\ell = \hat{x}$.

Proof We first show (3.19) when IM in Algorithm 3.2 stands for CGNR. Convergence for MR-II is discussed below. The projection operators P_i are such that the δ_i can be decreased as δ is reduced; this follows from (2.12). The proof proceeds by selecting suitably small $\delta_i > 0$. This can be done by choosing a sufficiently small $\delta > 0$.

We will show that for an arbitrary $\epsilon > 0$, there are positive $\delta_1, \delta_2, \ldots, \delta_\ell$, depending on ϵ such that

$$\|x_i - x_{i,k_i^{\delta_i}}^{\delta_i}\| \leq \epsilon, \qquad 1 \leq i \leq \ell. \tag{3.20}$$

This then shows (3.19). First consider level $i = 1$ and apply CGNR to the equation

$$A_1 z = g_1^{\delta_1},$$

using Stopping Rule 3.5 with $x_{1,0}^{\delta_0} = 0$. By Proposition 3.1, there is a $\delta_1 > 0$ such that equation (3.20), for $i = 1$, holds for the computed approximate solution $x_{1,k_1^{\delta_1}}^{\delta_1} := \Delta x_{1,k_1^{\delta_1}}^{\delta_1}$.

We turn to level $i = 2$. Let $x_{2,0}^{\delta_1} = E_2 x_{1,k_1^{\delta_1}}^{\delta_1}$. In view of Lemma 3.1, $x_{2,0}^{\delta_1}$ has no component in $\mathcal{N}(A_2)$. It follows from (2.13), for $i = 2$, that

$$\|g_2 - A_2 x_{2,0}^{\delta_1} - (g_2^{\delta_2} - A_2 x_{2,0}^{\delta_1})\| \leq \delta_2.$$

Application of CGNR to (3.9) for $i = 2$, with initial iterate $\Delta x_{2,0} = 0$ and Stopping Rule 3.5, yields the approximate solution $\Delta x_{2,k_2^{\delta_2}}^{\delta_2}$. It follows from Proposition 3.1 that we may choose δ_2, so that

$$\|\Delta x_2 - \Delta x_{2,k_2^{\delta_2}}^{\delta_2}\| \leq \epsilon,$$

where Δx_2 is defined by (3.17). This is equivalent to (3.20) for $i = 2$. We now can proceed in this fashion for increasing values of i. This shows (3.20) for all i, and thereby the theorem when IM is CGNR.

The proof of (3.19) when IM stands for MR-II follows analogously by application of Proposition 3.2.

When A is symmetric and positive semidefinite, the CR method also can be used in Algorithm 3.2. However, the computed approximate solution on level i may have a component in $\mathcal{N}(A_i)$.

The limit (3.19) cannot be established for indices i for which equation (3.15) is inconsistent. The following result shows that the limit (3.19) holds for indices of consistent equations.

Corollary 3.1 *Let $\mathcal{I}$ be the set of the indices of the consistent equations (3.15), $1 \leq i \leq \ell$. Then $\ell \in \mathcal{I}$ by the consistency of (1.4). Let the conditions of Theorem 3.6 hold, except that only equations (3.15) with indices $i \in \mathcal{I}$ are required to be consistent. Let δ_i for $i \notin \mathcal{I}$ be so large that no iterations are carried out on these levels. Then*

$$\lim_{\delta_i \searrow 0} \sup_{\|g_i - g_i^{\delta_i}\| \leq \delta_i} \|x_i - x_{i,k_i^{\delta_i}}^{\delta_i}\| = 0, \qquad i \in \mathcal{I},$$

where x_i, for $i \in \mathcal{I}$, is the minimal norm solution of (3.10) with $x_\ell = \hat{x}$.

Proof The result follows from the observation that the convergence (3.19) of $x_{i,k_i^{\delta_i}}^{\delta_i}$ in the proof of Theorem 3.6 can be established independently of $x_{i-1}^{\delta_{i-1}}$. Note that $x_{i,0}^{\delta_{i-1}}$ is orthogonal to $\mathcal{N}(A_i)$ because of (2.10).

Remark 3.2 In actual computations it may be beneficial to carry out a few steps with an iterative method also on levels $i \notin \mathcal{I}$, where $\mathcal{I}$ is defined in Corollary 3.1, in order to determine a more accurate approximate solution of the consistent equation

$$A_i z = (I - P_{\mathcal{N}(A_i)})g_i - A_i x_{i,0}^{\delta_{i-1}},$$

than $x_{i,0}^{\delta_{i-1}}$, where $P_{\mathcal{N}(A_i)}$ is defined by (3.16). This is likely to reduce the computational effort required to determine an approximate solution of (1.5). However, it is important not to carry out too many iterations to avoid a large propagated error due to the error in g^δ in the computed approximation of $\hat{x}$.

We finally note that, generally, we are only interested in determining an accurate approximation of $\hat{x}$. Therefore, the consistency requirement on levels $1, 2, \ldots, \ell - 1$ can be dispensed with. Consistency on these levels is convenient only because it helps us decide how many iterations to carry out on these levels.

4 Computed examples

We first solve an integral equation defined on an interval, then consider the estimation of the norm of the noise on each level of the multilevel methods, and finally discuss a problem from computerized tomography in two space-dimensions.

Example 4.1. Consider the Fredholm integral equation of the first kind,

$$(Kx)(t) := \int_0^{\pi/2} \kappa(s,t)x(s)ds = g(t), \qquad 0 \le t \le \pi/2, \tag{4.1}$$

where $\kappa(s,t) := \exp(s\cos(t))$ and

$$g(t) := \sin^2(3t) - 0.9t^3 + t^2. \tag{4.2}$$

Submatrix	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
n_i	28	50	88	158	292	554	1071	2098

Table 4.1 Example 4.1: Orders n_i of the matrices A_i.

The system matrix is given by $A = (\langle K\Psi_{j,k}, \Psi_{m,n}\rangle)_{m,n\in\Lambda}$. The function $x(t)$, $t \in [0, \pi/2]$ is discretized with 2^{11} equidistant points. Using the MATLAB routine wavedec and the Daubechies-4-wavelet, we determine a decomposition on 8 levels. Specifically, we obtain matrices A_i of sizes $n_i \times n_i$, $1 \le i \le 8$. Table 4.1 shows the orders n_i. The matrix A_j is a submatrix of A_i if $j \le i$. The condition number of the matrix A_8 is computed to be $2 \cdot 10^{26}$ by the MATLAB function cond; thus, the matrix is numerically singular.

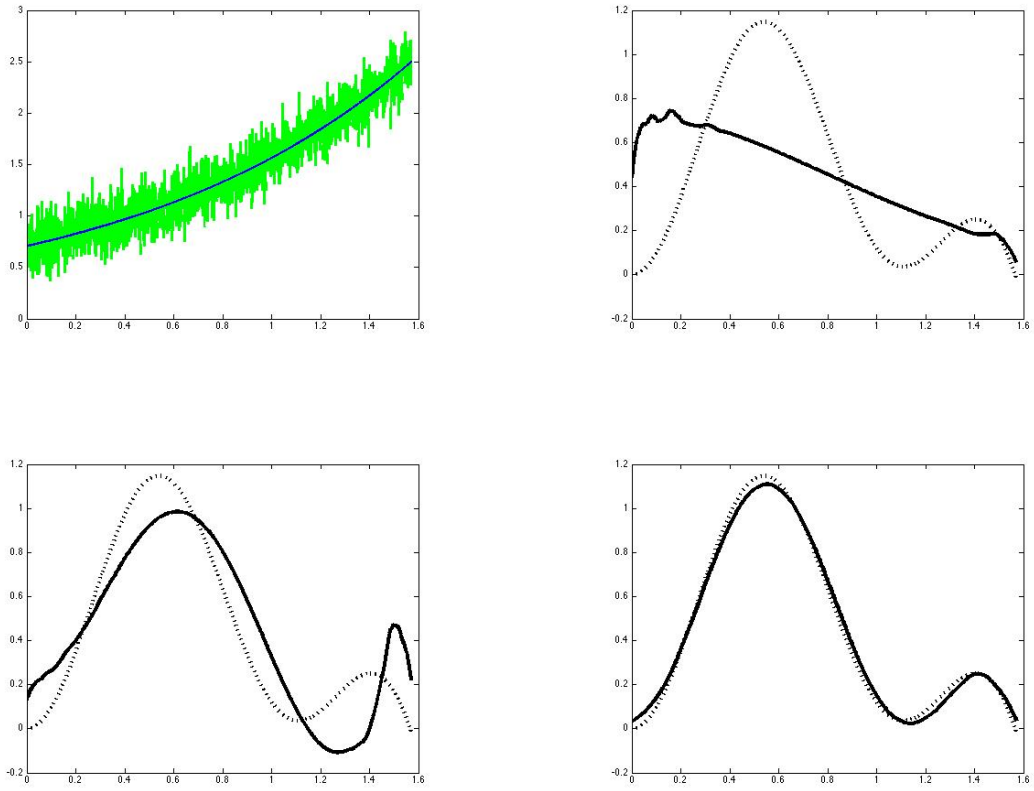

Fig. 4.1 Example 4.1: Exact right-hand side g and contaminated right-hand side g^δ with $\delta/\|g\| = 1 \cdot 10^{-1}$ (upper left), exact solution $\hat{x}$ and approximate solutions computed by ML-CGNR for $\delta/\|g\| = 1 \cdot 10^{-1}$ (upper right), for $\delta/\|g\| = 1 \cdot 10^{-2}$ (lower left), and for $\delta/\|g\| = 1 \cdot 10^{-5}$ (lower right). The exact solution is plotted as a dashed curve and the computed approximate solutions are displayed with solid curves.

Table 4.2 summarizes the computed results for standard (one-level) CGNR and (multilevel) ML-CGNR for several noise levels $\delta/\|g\|$. When an iterate from a coarser level has been refined, we first check whether the norm of the associated residual satisfies the stopping criterion for the present level, and if this is the case, then no iterations are carried out. The stopping indices $k_i^{\delta_i}$ for ML-CGNR in Table 4.2 show that only iterations on the coarsest level were carried out. This can be explained by the good approximation properties of the db4-wavelet: When the right-hand side g, given by (4.2), is projected onto the coarsest level, the (relative) approximation error of the projection is less than 1%. Since the solution error of CGNR itself with relative data error 10^{-5} is about 3%, it suffices to use the coarsest level when applying ML-CGNR. The acceleration of ML-CGNR, compared to CGNR, reported in Table 4.2 is the quotient of the number of multiplications required to evaluate

	CGNR		ML-CGNR		
$\frac{\delta}{\|g\|}$	k^δ	$\frac{\|x^\delta_{k^\delta} - \hat{x}\|}{\|\hat{x}\|}$	$k_i^{\delta_i}$	$\frac{\|x^{\delta_8}_{8,k_8^{\delta_8}} - \hat{x}\|}{\|\hat{x}\|}$	Accel.
$1 \cdot 10^{-1}$	2	$5.78 \cdot 10^{-1}$	2, 0, 0, 0, 0, 0, 0, 0	$5.66 \cdot 10^{-1}$	17.6
$1 \cdot 10^{-2}$	3	$5.35 \cdot 10^{-1}$	3, 0, 0, 0, 0, 0, 0, 0	$5.28 \cdot 10^{-1}$	23.5
$1 \cdot 10^{-3}$	5	$3.41 \cdot 10^{-1}$	5, 0, 0, 0, 0, 0, 0, 0	$3.24 \cdot 10^{-1}$	35.2
$1 \cdot 10^{-4}$	5	$3.32 \cdot 10^{-1}$	5, 0, 0, 0, 0, 0, 0, 0	$3.14 \cdot 10^{-1}$	35.2
$1 \cdot 10^{-5}$	8	$3.49 \cdot 10^{-2}$	8, 0, 0, 0, 0, 0, 0, 0	$3.46 \cdot 10^{-2}$	52.6

Table 4.2 Example 4.1: Termination indices k^δ for CGNR with Stopping Rule 3.1 determined by δ and $\tau = 1.1$, as well as relative errors in the computed approximate solutions $x^\delta_{k^\delta}$, and termination indices $k_1^{\delta_1}, k_2^{\delta_2}, \ldots, k_8^{\delta_8}$ (from left to right) for ML-CGNR with Stopping Rule 3.5 determined by the δ_i and $\tau = 1.1$, as well as relative errors in the computed approximate solutions $x^{\delta_8}_{8,k_8^{\delta_8}}$. In this example the $\delta_i = \|g_i - g_i^{\delta_i}\|$ are assumed to be known.

all the matrix-vector products for CGNR and the corresponding number of multiplications needed for ML-CGNR. Table 4.2 shows the acceleration to increase with decreasing error level. This is mainly due to the fact that more iterations are required for the solution for small noise levels.

The acceleration quotient is concerned with the numerical effort for the iterative process; the computational cost for the determining the matrix and right-hand side elements is not included. The computations required to evaluate the right-hand side entries is negligible, however, the computation of the matrix entries might not be. The computational effort for the latter depends on the application and the chosen discretization. We note that the same matrix entries can be used for many right-hand sides.

The results presented in Table 4.2 were obtained under the assumption that the projected noise levels $\delta_i = \|g_i - g_i^\delta\|$ are known. The knowledge of a tight bound for $\|g_i - g_i^\delta\|$ is helpful for determining a suitable number of iterations on each level, and yields high acceleration for ML-CGNR. This picture changes if we have no accurate bounds for the quantities $\|g_i - g_i^\delta\|$, and therefore set $\delta_i = \delta$ for all i. Then the acceleration is less pronounced, and for large noise levels there might be no acceleration. This observation indicates that good performance of ML-CGNR requires fairly accurate bounds for the errors $\|g_i - g_i^\delta\|$. The estimation of these errors is discussed in the following example. □

Example 4.2. The discussion at the end of the above example suggests that it is important for the good performance of ML-CGNR that accurate estimates for the quantities $\|g_i - g_i^{\delta_i}\|$ are available. Since the exact data g is unknown, so is g_i. This example describes a new approach, based on the analysis in [17], to estimate the norm of the noise in $g_i^{\delta_i}$. First, we seek to determine an approximation of the noise-free data g from the available noise-contaminated data g^δ. Denote this approximation by $S_\alpha g^\delta$. The latter is projected onto the space $\mathcal{U}_i$, and δ_i is estimated by

$$\tilde{\delta}_i := \|P_i S_\alpha g^\delta - g_i^\delta\|. \tag{4.3}$$

	$\frac{\delta}{\|g\|} = 1 \cdot 10^{-2}$		$\frac{\delta}{\|g\|} = 1 \cdot 10^{-4}$	
i	$\|g_i - g_i^\delta\|$	$\|P_i S_{\alpha_{opt}} g^\delta - g_i^\delta\|$	$\|g_i - g_i^\delta\|$	$\|P_i S_{\alpha_{opt}} g^\delta - g_i^\delta\|$
1	$1.4 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$2.3 \cdot 10^{-5}$	$1.4 \cdot 10^{-5}$
2	$2.0 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$2.9 \cdot 10^{-5}$	$2.6 \cdot 10^{-5}$
3	$3.1 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$3.7 \cdot 10^{-5}$	$3.4 \cdot 10^{-5}$
4	$4.3 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$4.8 \cdot 10^{-5}$	$4.8 \cdot 10^{-5}$
5	$6.3 \cdot 10^{-3}$	$7.6 \cdot 10^{-3}$	$6.5 \cdot 10^{-5}$	$6.7 \cdot 10^{-5}$
6	$9.3 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$9.2 \cdot 10^{-5}$	$9.2 \cdot 10^{-5}$
7	$1.3 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$
8	$1.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$

Table 4.3 Example 4.2: Comparison of the exact noise levels $\|g_i - g_i^\delta\|$ and their approximations $\|P_i S_{\alpha_{opt}} g^\delta - g_i^\delta\|$. The discrepancy principle (4.6) is used with $\tau = 1.4$.

In principle, any operator S_α that produces an accurate approximation of the error-free right-hand side g can be applied. We propose to use

$$S_\alpha g^\delta := \arg\min_{g \in \ell_2} \left\{ \|g^\delta - g\|_{\ell_2}^2 + \alpha \|g\|_{\ell_1} \right\}. \tag{4.4}$$

The minimizer of this functional can be determined by a coefficient-wise soft shrinkage,

$$\left(S_\alpha g^\delta\right)_k = \operatorname{sgn}(g^\delta)_k \max\left\{0, |(g^\delta)_k| - \alpha\right\}, \tag{4.5}$$

where sgn denotes the sign function. The regularization parameter $\alpha = \alpha_{opt}$ is chosen according to the discrepancy principle,

$$\alpha_{opt} = \min\{\alpha : \|S_\alpha g^\delta - g^\delta\| \geq \tau\delta\}. \tag{4.6}$$

The operator (4.4) together with the discrepancy principle (4.6) form a regularization method; see [17] for convergence rate results.

Table 4.3 displays the performance of the proposed scheme. The data vectors g and g^δ are associated with Example 4.1. The table shows that the norm of the noise can be estimated quite accurately on each one of the eight levels. The computational effort required is proportional to the number of elements in g^δ; see [17] for a discussion on the organization of the computations. □

Example 4.3. We present results for data reconstruction in computerized tomography. Specifically, we would like to determine an approximate solution of the equation

$$R^* R x = R^* g^\delta, \tag{4.7}$$

where g^δ is a noise-contaminated approximation of g with $\|R^*(g^\delta - g)\| \leq \delta$. The operator R denotes the 2d - Radon transform, $R : D \subset L_2(\mathbb{R}^2) \to L_2(\mathbb{R} \times S^1)$

$$(Rx)(s, \omega) = \int_{\mathbb{R}} x(s\omega + t\omega^\perp)\, dt$$

and R^* is its adjoint. The Radon transform considered between L_2 spaces is ill posed with degree 1/2, i.e., its singular values decrease like $n^{-1/2}$ as n

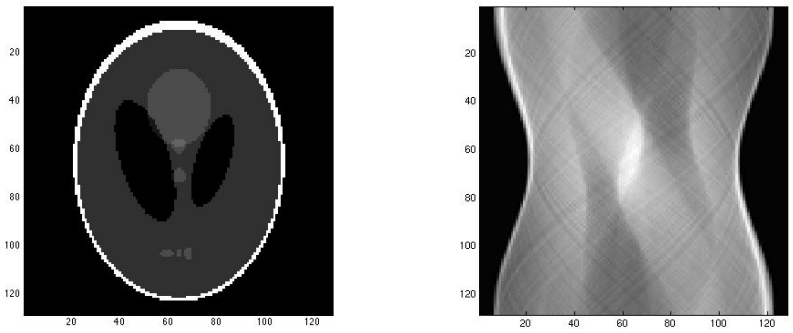

Fig. 4.2 Example 4.3: The phantom x_0 and its sinogram Rx_0.

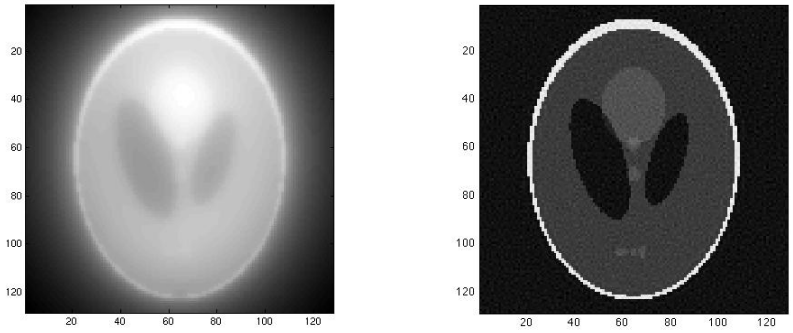

Fig. 4.3 Example 4.3: Noisy right-hand side $R^\star g^\delta$ with relative error $\|R^* g^\delta - R^* g\|/\|R^* g\| = 1 \cdot 10^{-3}$ (left), reconstruction of x_0 from this right-hand side (right).

increases. For a thorough analysis of the Radon transform, we refer to Natterer [25].

The reason for solving the normal equations (4.7) rather than the original equation $Rx = g^\delta$ is that the matrix representing R^*R is easier accessible than the matrix representation for R. Discretization is carried out by Haar wavelets. The image represented by the solution x has 128×128 pixels, and the matrix representation of R^*R is of size $2^{14} \times 2^{14}$ and has $2.7 \cdot 10^8$ entries. The size of the matrix necessitates the use of iterative solution methods and a computer with large memory.

We first seek to determine the function x_0 shown in Figure 4.2, which also shows the associated noise-free data Rx_0. The noise-contaminated right-hand

	CGNR		ML-CGNR							
$\frac{\|R^*(g-g^\delta)\|}{\|R^*g\|}$	k^δ	$\frac{\|x^\delta_{k^\delta}-x_0\|}{\|x_0\|}$	$k_1^{\delta_i}$	$k_2^{\delta_i}$	$k_3^{\delta_i}$	$k_4^{\delta_i}$	$k_5^{\delta_i}$	$k_6^{\delta_i}$	$\frac{\|x^{\delta_6}_{6,k_6^{\delta_6}}-x_0\|}{\|x_0\|}$	Accel.
$1\cdot10^{-1}$	6	$8.48\cdot10^{0}$	4,	4,	4,	3,	0,	4	$8.62\cdot10^{0}$	1.39
$1\cdot10^{-2}$	11	$9.36\cdot10^{-1}$	5,	7,	7,	7,	6,	5	$9.32\cdot10^{-1}$	1.85
$1\cdot10^{-3}$	17	$1.06\cdot10^{-1}$	7,	9,	10,	12,	11,	10	$1.08\cdot10^{-1}$	1.53
$1\cdot10^{-4}$	26	$1.23\cdot10^{-2}$	8,	12,	14,	18,	19,	19	$1.23\cdot10^{-2}$	1.27
$1\cdot10^{-5}$	36	$1.23\cdot10^{-3}$	10,	13,	19,	22,	25,	29	$1.23\cdot10^{-3}$	1.17

Table 4.4 Example 4.3: Solution of (4.7). For CGNR with Stopping Rule 3.1 determined by δ and $\tau = 1.1$: Termination indices k^δ and relative errors in the computed approximate solutions $x^\delta_{k^\delta}$; for ML-CGNR with Stopping Rule 3.5 determined by δ and $\tau = 1.1$: Termination indices $k_1^{\delta_1}, k_2^{\delta_2}, \ldots, k_6^{\delta_6}$ (from left to right) and relative errors in the computed approximate solutions $x^{\delta_6}_{6,k_6^{\delta_6}}$.

side R^*g^δ with 0.1% relative noise is displayed in Figure 4.3, which also shows the computed approximation of x_0 determined by 6-level ML-CGNR. We remark that due to the noise-damping property of $R^\star$, the noise in the data g^δ is usually much larger than in the right-hand side R^*g^δ. For the reconstruction we used the exact projected noise levels δ_i.

Details of the computations with standard CGNR and 6-level ML-CGNR are reported in Table 4.4. The table shows ML-CGNR to require fewer multiplications for matrix-vector product evaluation than standard CGNR; the reduction is 50% on the finest level for the noise level 1%.

Comparing the Tables 4.2 and 4.4 shows that the acceleration achieved in the present example is smaller than in Example 4.1. This is explained by the fact that the desired solution is not well approximated by the Haar basis. In particular, the finer levels still contain a lot of information; about 35% of its ℓ_2 norm is located at the finest level. This can be prevented by choosing a basis with better approximation properties, such as higher order wavelet bases from the Daubechies class, e.g., db2 or db4. Also the use of a different prolongation operator might help to speed up the convergence. Intuitively, our ML-CGNR implementation performs the best when the function to be determined contains only little information in the finer levels. This is in agreement with the results of Example 4.1. The following test supports this observation.

Starting with the function x_0, we create functions x_s by multiplying all wavelet coefficients of x_0 on level j by the factor $2^{-(j-1)\cdot s}$. In this way, the contribution of the finer levels to the norm of x_s is reduced for $s > 0$, and we expect a higher acceleration rate for ML-CGNR for larger values of s. This is confirmed by computations reported in Table 4.5. Multiplying the coefficients on wavelet level j by the factor $2^{-(j-1)\cdot s}$ with $s > 0$ increases the Besov smoothness. We conclude that the ML-CGNR algorithm performs particularly well whenever the desired solution has high Besov smoothness. □

s	Acceleration	$\frac{\|x_s - P_5 x_s\|}{\|x_s\|}$
0	$1.4 \cdot 10^0$	$4 \cdot 10^{-1}$
1	$1.9 \cdot 10^0$	$2 \cdot 10^{-2}$
3	$3.1 \cdot 10^1$	$2 \cdot 10^{-5}$
4	$7.0 \cdot 10^1$	$5 \cdot 10^{-7}$

Table 4.5 Example 4.3: Acceleration of ML-CGNR for the reconstruction of the functions x_s in dependence of s. The last column shows the relative approximation error of $P_5 x_s$ to x_s, where P_5 denotes the projection on the first 5 wavelet detail levels.

5 Conclusion

This paper describes cascadic wavelet-based multilevel methods using the minimal residual methods CGNR, CR, and MR-II as basic iteration schemes. Computed examples show ML-CGNR to be particularly effective for problems with a smooth solution that can be approximated fairly well already on the coarser levels.

Acknowledgements

L.R. would like to thank Heinz Engl and Ronny Ramlau for an enjoyable visit to RICAM, Linz, during which this work was begun, and Jocelyne Erhel for a pleasant visit to IRISA, Rennes, during which this work was continued.

References

1. Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
2. F. A. Bornemann and P. Deuflhard, The cascadic multigrid method for elliptic problems, Numer. Math., 75 (1996), pp. 135–152.
3. D. Calvetti, B. Lewis, and L. Reichel, On the choice of subspace for iterative methods for linear discrete ill-posed problems, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
4. Z. Chen, Y. Xu, and H. Yang, A multilevel augmentation method for solving ill-posed operator equations, Inverse Problems, 22 (2006), pp. 155–174.
5. A. Cohen, Numerical Analysis of Wavelet Methods, North-Holland, Amsterdam, 2003.
6. I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.
7. M. Donatelli and S. Serra-Capizzano, On the regularization power of multigrid-type algorithms, SIAM J. Sci. Comput., 27 (2006), pp. 2053–2076.
8. M. Donatelli and S. Serra-Capizzano, Filter factor analysis of an iterative multilevel regularizing method, Electron. Trans. Numer. Anal., 29 (2008), pp. 163–177.
9. H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
10. M. I. Español, Multilevel Methods for Discrete Ill-Posed Problems: Application to Deblurring, Ph. D. thesis, Department of Mathematics, Tufts University, Medford, MA, 2009.
11. C. W. Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, London, 1984.
12. M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman, Essex, 1995.

13. M. Hanke and C. R. Vogel, Two-level preconditioners for regularized inverse problems I: Theory, Numer. Math., 83 (1999), pp. 385–402.
14. P. C. Hansen, Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems, Numer. Algorithms, 6 (1994), pp. 1–35.
15. T. Huckle and J. Staudacher, Multigrid preconditioning and Toeplitz matrices, Electron. Trans. Numer. Anal., 13 (2002), pp. 81–105.
16. M. Jacobsen, P. C. Hansen, and M. A. Saunders, Subspace preconditioned LSQR for discrete ill-posed problems, BIT, 43 (2003), pp. 975–989.
17. L. Justen and R. Ramlau, A general framework for soft shrinkage with applications to blind deconvolution and wavelet denoising, Appl. Comput. Harmon. Anal., 26 (2009), pp. 43–63.
18. J. T. King, Multilevel algorithms for ill-posed problems, Numer. Math., 61 (1992), pp. 311–334.
19. A. K. Louis, P. Maass, and A. Rieder, Wavelets: Theory and Applications, Wiley, Chichester, 1997.
20. S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.
21. F. G. Meyer, Ondelettes sur l'Intervalle, Rev. Math. Iberoamer., 7 (1992), pp. 115–133.
22. S. Morigi, L. Reichel, and F. Sgallari, Noise-reducing cascadic multilevel methods for linear discrete ill-posed problems, Numer. Algorithms, 53 (2010), pp. 1–22.
23. S. Morigi, L. Reichel, and F. Sgallari, Cascadic multilevel methods for fast nonsymmetric blur- and noise-removal, Appl. Numer. Math., 60 (2010), pp. 378–396.
24. S. Morigi, L. Reichel, F. Sgallari, and A. Shyshkov, Cascadic multiresolution methods for image deblurring, SIAM J. Imag. Sci., 1 (2008), pp. 51–74.
25. F. Natterer, The Mathematics of Computerized Tomography, Teubner, Stuttgart, 1986.
26. A. S. Nemirovskii, The regularization properties of the adjoint gradient method in ill-posed problems, USSR Comput. Math. Math. Phys., 26 (2) (1986), pp. 7–16.
27. R. Plato, The method of conjugate residuals for solving the Galerkin equations associated with symmetric positive semidefinite ill-posed problems, SIAM J. Numer. Anal., 35 (1998), pp. 1621–1645.
28. L. Reichel and A. Shyshkov, Cascadic multilevel methods for ill-posed problems, J. Comput. Appl. Math., 233 (2010), pp. 1314–1325.
29. Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, Philadelphia, 2003.
30. O. Scherzer, An iterative multi level algorithm for solving nonlinear ill-posed problems, Numer. Math., 80 (1998), pp. 579–600.
31. G. Strang and T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1996.