Kernel polynomials for the solution of indefinite and ill-posed problems

Dedicated to R.S. Varga on the occasion of his 70th birthday

Harald Frankenberger [a] and Martin Hanke [b,*]

[a] *Fachbereich Mathematik, Universität Kaiserslautern, D–67653 Kaiserslautern, Germany*
[b] *Fachbereich Mathematik, Johannes Gutenberg-Universität Mainz, D–55099 Mainz, Germany*
E-mail: hanke@math.uni-mainz.de

We introduce a new family of semiiterative schemes for the solution of ill-posed linear equations with selfadjoint and indefinite operators. These schemes avoid the normal equation system and thus benefit directly from the structure of the problem. As input our method requires an enclosing interval of the spectrum of the indefinite operator, based on some a priori knowledge. In particular, for positive operators the schemes are mathematically equivalent to the so-called ν-methods of Brakhage. In a way, they can therefore be seen as appropriate extensions of the ν-methods to the indefinite case. This extension is achieved by substituting the orthogonal polynomials employed by Brakhage in the definition of the ν-methods by appropriate kernel polynomials. We determine the rate of convergence of the new methods and establish their regularizing properties.

Keywords: Semiiterative methods, ill-posed problems, regularization, indefinite problems, ν-methods

AMS Subject classification: 65J20, 65F10, 41A10

1. Introduction

This paper is concerned with the approximate solution of linear operator equations

$$Tx = y\,, \tag{1.1}$$

where $T : \mathcal{X} \to \mathcal{X}$ is a selfadjoint and indefinite operator in a Hilbert space $\mathcal{X}$. We presume that T is injective, i.e., solutions x of (1.1) are unique; however, we shall explicitly focus on the case that the range $\mathcal{R}(T)$ of T is only a dense subspace of $\mathcal{X}$. In this case the operator T is not continuously invertible because T^{-1} is unbounded by the Open Mapping Theorem.

* Corresponding author.

When T^{-1} is unbounded, equation (1.1) is usually referred to as being *ill-posed* because the solution x of (1.1) does not depend continuously on the given right-hand side y. Even worse, in general (1.1) will not have a solution.

Ill-posed problems often appear in the context of so-called *inverse problems*. The prototype is an integral equation of the first kind with a smooth kernel function, e.g.,

$$\int_{\mathbb{R}^2} e^{-\gamma((\xi-\xi')^2+(\eta-\eta')^2)} x(\xi',\eta')\, d(\xi',\eta') = y(\xi,\eta), \qquad \xi,\eta \in \mathbb{R}\,, \tag{1.2}$$

which may serve as a simple mathematical model for imaging through atmospheric turbulence; the associated integral operator is selfadjoint and positive. Today astronomers use more elaborate kernel functions which often ressemble the Gaussian kernel in (1.2), cf., e.g., Bertero and Boccacci [3]. While the corresponding integral operator may still be selfadjoint, it typically is no longer positive but rather indefinite, although the dominating part of the spectrum belongs to the positive axis (see [17] for an example).

Ill-posed problems need some sort of regularization to encompass the loss of stability due to the unboundedness of T^{-1}. One can, for instance, use Tikhonov regularization to approximate the solution x. Another option that we employ in this paper is the use of iterative schemes for solving (1.1): here, regularization is incorporated via an early termination of the iteration which prevents unbounded data error propagation.

Most iterative methods for ill-posed problems are based on the normal equation $T^*Tx = T^*y$ for (1.1). When T is selfadjoint, however, we cannot benefit from the symmetry in this way. Calvetti, Reichel and Zhang [5] have recently suggested an iterative scheme which directly works with (1.1) when T is selfadjoint and indefinite. The purpose of this paper is to provide alternative algorithms which reduce to established schemes (the so-called ν-methods [4]) in the case when T is semidefinite. Like the method of Calvetti, Reichel and Zhang our new schemes belong to the class of semiiterative methods which are generated by appropriate sequences of polynomials; in our case these are kernel polynomials for certain generalized Jacobi polynomials. As a consequence, our iterates can be computed with short recurrences.

The outline of this paper is as follows. In Section 2 we review the basic facts about semiiterative methods for selfadjoint linear equations with particular emphasis on the specific subtleties for ill-posed problems. Then, in Section 3, we provide well-known results on kernel polynomials, which are subsequently used to define the new method in Section 4. We go on and prove rate of convergence estimates in Section 5 before we turn in Section 6 to the regularizing properties of our algorithm. Finally, in Section 7, we present some numerical results to illustrate our findings.

2. Semiiterative methods

Semiiterative methods were introduced in full generality by R.S. Varga in his book [24], although the terminology has been coined for a specific example somewhat earlier in [13]. In [24] *semiiterative* methods refer to hybrid schemes, which in each iteration carry out one step of residual correction (or Richardson iteration)

$$x_{n+1} = x_n + (y - Tx_n)\,, \tag{2.1}$$

followed by a linear combination of some or all previous iterates. Typically the iteration is initialized with $x_0 = 0$.

Semiiterative methods can alternatively be described by two sequences $\{r_n\}$ and $\{g_n\}$ of polynomials where g_n belongs to the set Π_{n-1} of polynomials of degree $n-1$ or less and $r_n \in \Pi_n$ with

$$r_n(\lambda) = 1 - \lambda g_n(\lambda)\,. \tag{2.2}$$

The polynomial g_n defines the n^{th} iterate via

$$x_n = g_n(T)y\,, \tag{2.3}$$

hence the error equals

$$x - x_n = x - g_n(T)Tx = r_n(T)x\,. \tag{2.4}$$

For the basic iteration (2.1), for example, we have

$$g_n(\lambda) = \sum_{j=0}^{n-1}(1-\lambda)^j \qquad \text{and} \qquad r_n(\lambda) = (1-\lambda)^n\,.$$

In the context of ill-posed problems, the iteration (2.1) is attributed to Fridman [11]; it converges whenever T is selfadjoint and positive with $\|T\| < 2$, in the sequel we fix throughout

$$\|T\| = 1\,.$$

In view of (2.4), $|r_n(\lambda)|$ should be as small as possible for elements λ from the spectrum $\sigma(T)$ of T in order to have a small error $x - x_n$. On the other hand, (2.2) implies the constraint

$$r_n(0) = 1 \tag{2.5}$$

which has to be satisfied. For ill-posed problems this constraint is crucial because $\lambda = 0$ is an accumulation point of the spectrum of T, and hence, the best one can hope for is that, as $n \to \infty$,

$$r_n(\lambda) \longrightarrow 0\,, \qquad \text{pointwise on } \sigma(T) \setminus \{0\}\,. \tag{2.6}$$

If, in addition, the polynomial sequence

$$\{r_n\} \text{ is uniformly bounded on } \sigma(T) \tag{2.7}$$

then we have convergence $x_n \to x$ as $n \to \infty$ by the Banach-Steinhaus Theorem, compare Theorem 4.1 in [6].

We note that for the Fridman iteration (2.1), for example, (2.6) and (2.7) are satisfied when T is selfadjoint and positive. Nevertheless, the shortcoming of the Fridman iteration is its slow convergence. For ill-posed problems the rate of convergence of $x_n \to x$ as $n \to \infty$ cannot be measured in general terms: Since $r_n(0) = 1$ and the spectrum of T clusters at the origin, it follows from (2.4) and (2.5) that the error $\|x - x_n\|$ can be arbitrarily close to $\|x\|$. To obtain error bounds it is necessary to have a priori information concerning x like, e.g.,

$$x \in \mathcal{R}(|T|^\nu) \qquad \text{for some } \nu > 0\,. \tag{2.8}$$

Here, $|T|$ denotes the square root of the positive operator T^2; for $\nu \in \mathbb{N}$, $\mathcal{R}(|T|^\nu)$ can be replaced by the more familiar $\mathcal{R}(T^\nu)$. Since T of (1.2) is a smoothing operator and since $\nu > 0$, (2.8) can be considered to be a generalized smoothness assumption. Inserting (2.8) into (2.4), i.e., setting $x = |T|^\nu w$ for some $w \in \mathcal{X}$ we obtain from the Spectral Mapping Theorem

$$\|x - x_n\| = \|r_n(T)|T|^\nu w\| \le \|r_n(T)|T|^\nu\|\,\|w\| \le \sup_{\lambda\in\sigma(T)} |\lambda|^\nu |r_n(\lambda)|\,\|w\|\,. \tag{2.9}$$

Estimating $|\lambda|^\nu |r_n(\lambda)|$ therefore provides a general error bound. The supremum of $|\lambda|^\nu |r_n(\lambda)|$ over $\sigma(T)$ or the enclosing interval [0,1] is hence referred to as *modulus of convergence*. For Fridman's iteration the upper bound in (2.9) becomes

$$\sup_{\lambda\in\sigma(T)} |\lambda|^\nu |r_n(\lambda)| \le \sup_{\lambda\in[0,1]} \lambda^\nu (1-\lambda)^n = \frac{\nu^\nu n^n}{(n+\nu)^{n+\nu}} \le \nu^\nu n^{-\nu}\,.$$

For example, when $x \in \mathcal{R}(T)$ then this implies a rate of convergence of $x_n \to x$ of the order $O(n^{-1})$ as $n \to \infty$.

Faster schemes have been developed in the 1980's by Schock, Nemirovskii and Polyak, and Brakhage (see [14] for a survey and additional references). In particular, the so-called ν-methods by Brakhage [4] — a family of semiiterative methods parameterized by the real number $\nu > 0$ associated with (2.8) — have modulus of convergence

$$\sup_{\lambda\in[0,1]} |\lambda|^\nu |r_n(\lambda)| = O(n^{-2\nu})\,, \qquad n \to \infty\,, \tag{2.10}$$

under the condition (2.8).

Still, all these methods share the disadvantage that they require T to be semidefinite in order to take advantage of the symmetry of T. Recently, Calvetti, Reichel and Zhang [5] suggested a semiiterative method for selfadjoint problems

with indefinite T. While mathematically appealing, their method may be somewhat complicated to implement and does not lead to a sequence converging to x as $n \to \infty$ but rather to some regularized approximation of x.

In contrast, the methods to be presented below, converge to x in exact arithmetic with exactly given right-hand side y; in the presence of data errors regularized approximations of x can be obtained by terminating the iteration before instabilities become notable. Furthermore, these methods can be viewed as an extension of the ν-methods to the case when T is selfadjoint and indefinite.

3. Kernel polynomials

The key idea in [4] (see also [14]) for reducing the number of iterations in the Fridman iteration (2.1) without raising the costs of each individual step (one multiplication with T and a few linear combinations of elements in $\mathcal{X}$) was to choose orthogonal polynomials $\{r_n\}$ with respect to some weight function supported on $[0, 1]$. The efficiency of such a scheme is a consequence of the three-term recurrence relation of the orthogonal polynomials. The weight function corresponding to the ν-method ($\nu > 0$) is

$$w(t) = t^{2\nu}/\sqrt{t(1-t)}\,, \qquad 0 < t < 1\,. \tag{3.1}$$

The polynomials r_n are therefore translated Jacobi polynomials with a multiplicative scaling to achieve the normalization (2.5); such a normalization is possible because all roots of r_n belong to the open interval $(0, 1)$, which garantees that $r_n(0)$ is nonzero.

Given an indefinite operator T with spectrum

$$\sigma(T) \subset [a, 1]\,, \qquad -1 < a < 0\,,$$

and a weight function w supported on $[a, 1]$, a normalization (2.5) for the corresponding orthogonal polynomials might no longer be possible because the origin is an interior point of $[a, 1]$, and hence, can be a root of some member of the orthogonal family, cf. Figure 4.1. Motivated by some work of Fischer and Prestin [9] concerning polynomial wavelets we were therefore led to the use of *kernel polynomials* rather than orthogonal polynomials for the definition of our new semi-iterative scheme. In this we also follow Stiefel [22] who, back in 1955, already recommended the use of kernel polynomials for some applications in numerical linear algebra; see also Fischer [7].

Let $\{p_n\}$ be the sequence of orthonormal polynomials for a given weight function w over $[a, 1]$. Then we require the associated kernel polynomials

$$K_n(t, \tau) = \sum_{k=0}^{n} p_k(t) p_k(\tau)\,, \qquad t, \tau \in \mathbb{R}\,,$$

and the *Christoffel functions*

$$\lambda_{n+1}(\tau) = 1/K_n(\tau,\tau)\,, \qquad \tau \in \mathbb{R}\,,$$

the latter being positive over $\mathbb{R}$ because $K_n(\tau,\tau) \geq |p_0(\tau)|^2 > 0$. Using the kernel polynomials we can define

$$r_n(t) = K_n(t,0)/K_n(0,0) = \lambda_{n+1}(0)K_n(t,0)\,. \tag{3.2}$$

Obviously, r_n is a polynomial of degree n which satisfies the normalization constraint (2.5).

Remark 3.1. We emphasize that Brakhage's ν-methods for the semidefinite case can be introduced in much the same way. The reason is that the n^{th} orthogonal polynomial for the weight function w of (3.1) is — up to a multiplicative constant — the kernel polynomial $K_n(\,\cdot\,,0)$ for the weight function

$$\tilde{w}(t) = w(t)/t = t^{2(\nu-1/2)}/\sqrt{t(1-t)}\,, \qquad 0 < t < 1\,; \tag{3.1$'$}$$

of course, the latter defines a weight function only for $\nu > 1/4$.

To demonstrate the efficiency of each individual step of a semiiterative method generated by (3.2) we make use of the Christoffel-Darboux identity which states that

$$K_n(t,\tau) = \frac{1}{a_n}\frac{p_{n+1}(t)p_n(\tau) - p_n(t)p_{n+1}(\tau)}{t-\tau}\,, \tag{3.3}$$

where a_n is given by the three-term recurrence relation for the sequence $\{p_n\}$:

$$p_{n+1}(t) = (a_n t + b_n)p_n(t) - c_n p_{n-1}(t)\,, \qquad n \in \mathbb{N}_0\,, \tag{3.4}$$

with $p_{-1} \equiv 0$, $p_0 \equiv (\int_a^1 w(t)\,dt)^{-1/2}$, and $a_n > 0$ for all $n \in \mathbb{N}$. Inserting (3.3) and (3.4) into (3.2) we eventually obtain Algorithm 3.1 for the recursive computation of the iterates x_n of (2.3) and their residuals

$$d_n = y - Tx_n = r_n(T)y\,,$$

where r_n is given by (3.2). The details of this derivation are left to the reader; we merely note that the intermediate quantities $\beta_n \in \mathbb{R}$ and $z_n \in \mathcal{X}$ in Algorithm 3.1 satisfy

$$\frac{\beta_n}{(\int_a^1 w(t)\,dt)^{1/2}} = p_n(0) \qquad \text{and} \qquad \frac{z_n}{(\int_a^1 w(t)\,dt)^{1/2}} = q_{n+1}(T)y$$

with

$$q_{n+1}(\lambda) = \frac{p_{n+1}(\lambda) - p_{n+1}(0)}{\lambda} \in \Pi_n\,.$$

`On Input:` $\{a_n, b_n, c_n : n = 0, 1, 2, \ldots\}$, i.e., the recursion coefficients in (3.4) of the orthonormal polynomials p_n

`On Output:` $x_n = g_n(T)y\,,\ \ d_n = y - Tx_n\,,\ \ n = 0, 1, 2, \ldots$

$z_{-1} = z_0 = 0$

$\beta_{-1} = 0\,,\ \beta_0 = 1$

$\alpha_0 = \beta_0^2$

$x_0 = 0\,,\ d_0 = y$

`for` $n = 0, 1, 2, \ldots$ `do`

$z_{n+1} = a_n(Tz_n + \beta_n y) + b_n z_n - c_n z_{n-1}$

$\beta_{n+1} = b_n\beta_n - c_n\beta_{n-1}$

$\alpha_{n+1} = \alpha_n + \beta_{n+1}^2$

$x_{n+1} = \frac{\alpha_n}{\alpha_{n+1}}\, x_n - \frac{\beta_{n+1}}{\alpha_{n+1}}\, z_{n+1}$

$d_{n+1} = \frac{\alpha_n}{\alpha_{n+1}}\, d_n + \frac{\beta_{n+1}}{\alpha_{n+1}}\, (Tz_{n+1} + \beta_{n+1} y)$

`end for`

Algorithm 3.1: Computation of the semiiterative method based on (3.2)

4. Generalized Jacobi polynomials

In (3.1) and (3.1′) the order of the zeros of w and $\tilde{w}$ at $t = 0$ is crucial to establish the improved rate of convergence (2.10). We are therefore led to study the one parameter family (with parameter $\mu \geq 1$)

$$w_\mu(t) \;=\; |t|^{2(\mu-1)}/\sqrt{(1-t)(t-a)}\,, \qquad a < t < 1\,, \tag{4.1}$$

which is obtained from the (translated and rescaled) Chebyshev weight by adding a multiple root at the origin; note that, in general, $t = 0$ is an interior point of the interval $[a, 1]$. For $\mu = 1$ we obtain the familiar Chebyshev weight function on the interval $[a, 1]$.

The weight functions (4.1) belong to the class of so-called *generalized Jacobi weights*, cf., e.g., Nevai [20, p. 169]. Those weight functions have raised considerable interest, and a number of remarkable results have been proven. We require two of these, which take the following form for our specific weight functions:

Theorem A (Nevai [20, p. 120]). The Christoffel functions $\lambda_n(w_\mu; t)$ corresponding to (4.1) satisfy

$$\lambda_n(w_\mu; t) \;\sim\; \frac{1}{n}\left(\,|t| \,+\, \frac{1}{n}\right)^{2(\mu-1)}\,, \qquad n \to \infty\,,$$

uniformly for $a \leq t \leq 1$.

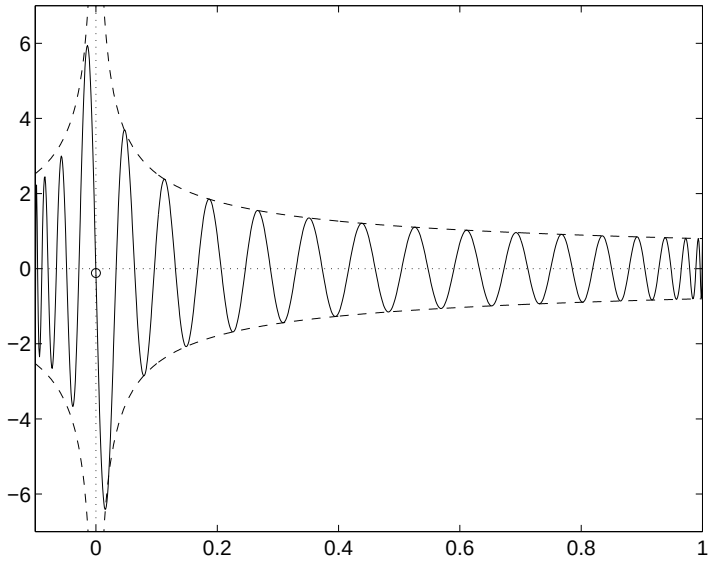

Figure 4.1. Orthonormal polynomial p_{39} $(a = -0.1, \mu = 1.5)$.

Theorem B (Badkov [1,2], see also [20, p. 170]). The orthonormal polynomials $p_n(w_\mu; t)$, $n \in \mathbb{N}_0$, corresponding to (4.1) satisfy

$$\Big(|t| + \frac{1}{n} \Big)^{2(\mu-1)} |p_n(w_\mu; t)|^2 = O(1), \qquad n \to \infty,$$

uniformly for $a \le t \le 1$.

Here, the notation $\varphi_n \sim \psi_n$ is used to refer to the fact that the quotients φ_n/ψ_n and ψ_n/φ_n are bounded as $n \to \infty$. We shall make use of these two results to estimate the modulus of convergence and divergence of our new methods in Sections 5 and 6.

We refer to Figure 4.1 for a typical orthonormal polynomial p_n corresponding to (4.1). The small circle in this plot indicates that $p_n(0)$ may be close or even equal to zero, which is the reason why those polynomials should not be used for r_n in (2.4). Figure 4.2 contains the corresponding rescaled kernel polynomial r_n of (3.2). The dashed lines in these two figures show appropriate multiples of the asymptotic envelope functions $\lambda \mapsto \pm|\lambda|^{-(\mu-1)}$ and $\lambda \mapsto \pm|\lambda|^{-\mu}$, respectively.

As we have indicated before the recursion coefficients for the orthogonal polynomials corresponding to (4.1) enter into the recursive computation of the kernel polynomials in Algorithm 3.1. Except for a few exceptional values of $a \le 0$ those coefficients are not explicitly known, but they can always be computed recursively.

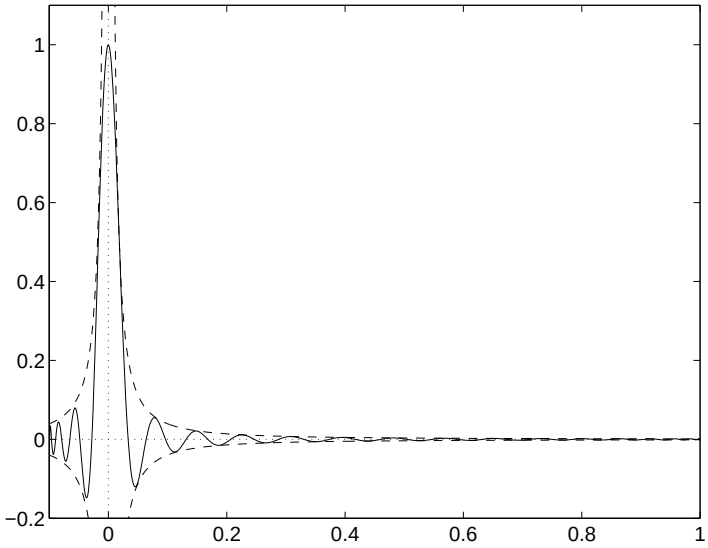

Figure 4.2. Rescaled kernel polynomial r_{39} $(a = -0.1, \mu = 1.5)$.

One option for their computation is the well-known *modified Chebyshev algorithm*, cf., e.g., Gautschi [12] or Fischer and Golub [8]; for a detailed investigation of this method with particular emphasis on the above weight function, we refer to [10]. One disadvantage with this scheme is its increasing complexity when $\mu \notin \mathbb{N}$.

Even for $\mu \in \mathbb{N}$ it is somewhat cheaper to use an alternative recursion suggested recently by Magnus [19] for this and a number of related weight functions. We state this scheme for our particular case in Algorithm 4.1 and refer to [19] and [10] for further details. As is shown there the quantities κ_n in this algorithm are positive for all $n \in \mathbb{N}$; furthermore, we have

$$a_n \longrightarrow \frac{4}{1+|a|}\,, \quad b_n \longrightarrow -2\,\frac{1+a}{1+|a|}\,, \quad c_n \longrightarrow 1\,, \qquad \text{as } n \to \infty\,, \tag{4.2}$$

cf. Szegö [23, Sect. 12.7].

Algorithm 4.1 takes just a few multiplications and one square root per iteration. In our implementation we have used the ten point Gauß-Chebyshev quadrature rule for the approximation of the two integrals in the definition of κ_1.

5. Applications to ill-posed problems

As outlined in Sect. 3 we shall use the kernel polynomials corresponding to the weight function (4.1) (for some fixed $\mu \geq 1$) to define via (3.2) the polynomials r_n of our semiiterative method for the regularization of ill-posed and indefinite problems (1.1).

`On Input:` left end $a \in [-1, 0]$ of interval $[a, 1] \subset \sigma(T)$; index μ of w_μ
`On Output:` the recursion coefficients $\{a_n, b_n, c_n\}$ in (3.4) of $p_n(w_\mu; \cdot)$

$\xi^2 = a/(a-1)$

$\gamma_{-1} = 0$

$\kappa_0 = \kappa_{-1} = 0$

$$\kappa_1 = \int_{-1}^{1} t^2 \, \frac{|t^2 - \xi^2|^{2(\mu-1)}}{\sqrt{1-t^2}} \, dt \; \Big/ \int_{-1}^{1} \frac{|t^2 - \xi^2|^{2(\mu-1)}}{\sqrt{1-t^2}} \, dt$$

$\sigma = \Sigma = 0$

`for` $n = 0, 1, 2, \ldots$ `do`

$m = 2n + 1$
$\rho = m(\xi^2 + 1) + 2(\mu - 1) + 1/2 - 2\sigma - (m + 2\mu)\kappa_{m-1}$
$\sigma = \sigma + \kappa_{m-1}$
$\Sigma = \Sigma + \kappa_{m-1}^2 + 2\kappa_{m-2}\kappa_{m-1}$
$\tau = m\,\xi^2/2 - (\xi^2 + 1)\sigma + \Sigma$
$\kappa_{m+1} = (\rho - \tau/\kappa_m)/(m + 2(\mu - 1) + 1) - \kappa_m$
$\gamma_n = 2\sqrt{\kappa_m \kappa_{m+1}}$
$a_n = 2/((1 - a)\gamma_n)$
$b_n = -a_n(a + (1 - a)(\kappa_{m-1} + \kappa_m))$
$c_n = \gamma_{n-1}/\gamma_n$
`% +++++` insert here the main loop of Algorithm 3.1 `+++++`
`% +++++` to obtain the new semiiterative method `+++++`
$m = 2n + 2$
$\rho = m(\xi^2 + 1) + 2(\mu - 1) + 1/2 - 2\sigma - (m + 2\mu)\kappa_{m-1}$
$\sigma = \sigma + \kappa_{m-1}$
$\Sigma = \Sigma + \kappa_{m-1}^2 + 2\kappa_{m-2}\kappa_{m-1}$
$\tau = m\,\xi^2/2 - (\xi^2 + 1)\sigma + \Sigma$
$\kappa_{m+1} = (\rho - \tau/\kappa_m)/(m + 2(\mu - 1) + 1) - \kappa_m$
`end for`

Algorithm 4.1: Computation of the recursion coefficients

We emphasize that for $a = 0$ the weight function (4.1) becomes

$$w_\mu(t) \; = \; t^{2(\mu-1)}/\sqrt{t(1-t)}\,, \qquad 0 < t < 1\,.$$

In view of Remark 3.1 our new scheme therefore extends Brakhage's ν-methods (with $\nu = \mu - 1/2 \geq 1/2$) to the indefinite case.

On the other hand it has been shown in [10, Satz 6.15] that for $a = -1$ and every $n \in \mathbb{N}_0$ *both* the $(2n)^{th}$ *and* the $(2n+1)^{st}$ iterate of our new scheme

coincide with the n^{th} iterate of the ν-method with $\nu = \mu/2$ when applied to the semidefinite problem $T^2 x = Ty$, i.e., to the normal equation of (1.1).

As a consequence, for a between -1 and 0, there is an initial phase of the iteration where our indefinite scheme mimicks the performance of the ν-method with $\nu = \mu - 1/2$, although the long range asymptotic behaviour of the iteration corresponds to the ν-method with $\nu = \mu/2 \leq \mu - 1/2$ applied to the normal equation; see [10, Sect. 6.4] for some numerical illustration.

Now we turn to a convergence analysis of our scheme to establish the aforementioned asymptotic behaviour.

Theorem 5.1. Let $\mu \geq 1$ and $-1 \leq a < 0$. Then the sequence $\{r_n\}$ defined as above is uniformly bounded over $[a, 1]$ and there exists a constant c_μ such that for all $n \in \mathbb{N}$

$$|t|^\mu |r_n(t)| \;\leq\; c_\mu n^{-\mu}\,, \qquad a \leq t \leq 1\,.$$

Proof. From (3.2) we have

$$|r_n(t)| \;=\; \lambda_{n+1}(0)\,|K_n(t,0)| \;\leq\; \lambda_{n+1}(0) \sum_{k=0}^{n} |p_k(t)p_k(0)|\,.$$

Since $\lambda_n(0) \sim n^{-2\mu+1}$ by Theorem A and $|p_k(t)| = O(n^{\mu-1})$, for $a \leq t \leq 1$ and $0 \leq k \leq n$ by Theorem B, we obtain

$$|r_n(t)| \leq C n^{-2\mu+1} \sum_{k=0}^{n} n^{\mu-1} n^{\mu-1} = C$$

for some $C > 0$, and hence, the polynomials r_n, $n \in \mathbb{N}$, are uniformly bounded. By the Christoffel-Darboux formula, we have

$$\begin{aligned} |t|^\mu |r_n(t)| &= |t|^\mu\,|K_n(t,0)|\,/K_n(0,0) \\ &= \frac{1}{a_n}\,|t|^{\mu-1}\lambda_{n+1}(0)\,\bigl|p_{n+1}(t)p_n(0) \;-\; p_n(t)p_{n+1}(0)\bigr|\,. \end{aligned}$$

From Theorem A we conclude that $\lambda_n(0) \sim n^{-2\mu+1}$, whereas according to Theorem B $|p_k(0)| = O(n^{\mu-1})$ for $k = n$ and $n+1$, respectively. Similarly, since $\mu \geq 1$, Theorem B implies that

$$|t|^{\mu-1}|p_k(t)| \;=\; O(1)$$

for $k = n$ and $n+1$, uniformly for $a \leq t \leq 1$. Finally, because of (4.2), the factor $1/a_n$ remains bounded for $n \in \mathbb{N}$. From this follows the second assertion. □

As a corollary we obtain

Corollary 5.2. For every $\nu \in [0, \mu]$ there exists c_ν such that

$$|t|^\nu |r_n(t)| \;\leq\; c_\nu n^{-\nu}\,, \qquad a \leq t \leq 1\,.$$

Proof. Using the uniform boundedness of r_n in $[a,1]$ and the asymptotic estimate in Theorem 5.1 we obtain, since $0 \le \nu/\mu \le 1$,

$$|t|^\nu \, |r_n(t)| \;=\; \left(|t|^\mu \, |r_n(t)|\right)^{\nu/\mu} |r_n(t)|^{1-\nu/\mu} \;\le\; c_\mu^{\nu/\mu} \, n^{-\nu} \, C^{1-\nu/\mu} \, .$$

□

We have thus estimated the modulus of convergence of our new scheme. Note that it is not as fast as the ν-method applied to a semidefinite problem, cf. (2.10). Rather, the rate of convergence is the same as for the ν-method (with $\nu = \mu/2$) when applied to the normal equation $T^2 x = Ty$. To see this, recall that

$$x \in \mathcal{R}(|T|^\mu) \;=\; \mathcal{R}(|T^2|^{\mu/2}) \, ,$$

and hence, the ν-method (with $\nu = \mu/2$) will have a rate of convergence like $O(n^{-2\mu/2}) = O(n^{-\mu})$.

6. Regularizing properties

We now turn to an investigation of the self-regularizing properties of our new algorithm. To this end we introduce its so-called *modulus of divergence*, i.e., the numbers

$$G_n := \max_{a \le t \le 1} |g_n(t)| \, , \qquad n \in \mathbb{N} \, . \tag{6.1}$$

Note that G_n describes the worst-case propagation of data errors in the right-hand side within the first n iterations, because in view of (2.3) the iterates x_n and $\tilde{x}_n$ corresponding to right-hand sides y and $\tilde{y}$ satisfy

$$\|x_n - \tilde{x}_n\| \;=\; \|g_n(T)(y-\tilde{y})\| \;\le\; \|g_n(T)\| \, \|y-\tilde{y}\| \;\le\; G_n \|y-\tilde{y}\| \tag{6.2}$$

by the Spectral Mapping Theorem.

Since $g_n(\lambda) = (1 - r_n(\lambda))/\lambda$ by (2.2), where r_n converges to zero pointwise in $[a,1] \setminus \{0\}$, the polynomials g_n converge pointwise to the function $\lambda \mapsto 1/\lambda$ in $[a,1] \setminus \{0\}$, and hence, G_n of (6.1) will converge to infinity as $n \to \infty$. Since this corresponds to an unlimited growth of the propagated data error as the iteration proceeds, it is essential to derive sharp estimates for the modulus of divergence.

Theorem 6.1. Let $\mu \ge 1$ and $-1 \le a < 0$. Then there exists a constant C_μ such that

$$G_n \;\le\; C_\mu n \, , \qquad n \in \mathbb{N} \, .$$

Proof. From the representation (2.2) we obtain for $|\lambda| > |a|/2$ with $\lambda \in [a,1]$ that

$$|g_n(\lambda)| \;\leq\; \frac{1+|r_n(\lambda)|}{|\lambda|} \;\leq\; \frac{2}{|a|}\,(1+\max_{a\leq\tau\leq 1}|r_n(\tau)|)\,.$$

Since the polynomials r_n are uniformly bounded over $[a,1]$ by Theorem 5.1 we conclude that the polynomials g_n are also uniformly bounded on $[a,1]\backslash[a/2,|a|/2]$.

Next, we consider arguments $\lambda \in [a/2,|a|/2]$. From (2.2) and (2.5) we obtain

$$g_n(\lambda) \;=\; \frac{1-r_n(\lambda)}{\lambda} \;=\; \frac{r_n(0)-r_n(\lambda)}{\lambda} \;=\; -r_n'(\xi)$$

for some $\xi \in (a/2,|a|/2)$ by the Mean-Value Theorem. Since the sequence $\{r_n\}$ is uniformly bounded over $[a,|a|]$, Bernstein's inequality (cf., e.g., [23, Theorem 1.22.3]) yields the bound

$$|r_n'(\lambda)| \;\leq\; cn/\sqrt{a^2-\lambda^2} \;\leq\; \frac{2c}{\sqrt{3}\,|a|}\,n\,, \qquad |\lambda| \leq |a|/2\,,$$

for the derivative of r_n, where $c>0$ is some appropriate constant independent of $n \in \mathbb{N}$. It follows that

$$|g_n(\lambda)| \;\leq\; \frac{2c}{\sqrt{3}\,|a|}\,n\,, \qquad a/2 \leq \lambda \leq |a|/2\,,$$

and the proof is complete. □

Using the modulus of convergence and the modulus of divergence we obtain a pretty realistic description of the *semiconvergence phenomenon* observed for iterative regularization methods in practice. Let us assume that the exact solution $x = T^{-1}y$ satisfies the smoothness condition $x = |T|^\nu w$ with some $0 < \nu \leq \mu$ and $w \in \mathcal{X}$, and assume we are given perturbed data $\tilde{y}$: Then we can use (2.4) and (6.2) and rewrite the error of the n^{th} iterate $\tilde{x}_n$ corresponding to the given data $\tilde{y}$ as

$$x - \tilde{x}_n \;=\; x - x_n + x_n - \tilde{x}_n \;=\; r_n(T)|T|^\nu w + g_n(T)(y-\tilde{y})\,.$$

Using the triangle inequality and Theorems 5.1 and 6.1 we obtain

$$\begin{aligned} \|x-\tilde{x}_n\| &\leq \|r_n(T)|T|^\nu\|\,\|w\| \;+\; G_n\,\|y-\tilde{y}\| \\ &\leq c_\nu\|w\|\,n^{-\nu} \;+\; C_\mu\|y-\tilde{y}\|\,n\,. \end{aligned} \tag{6.3}$$

Assuming that the data error $\|y-\tilde{y}\|$ is small, the first term in (6.3) dominates in the beginning of the iteration, and the iteration error seemingly converges to zero in this initial stage. However, as n gets large, the influence of the propagated data error becomes notable and the iteration error will subsequently stagnate before it eventually diverges to infinity (compare Figure 7.2 for an illustration of this semiconvergence phenomenon).

Under these assumptions the best possible approximations of x are therefore obtained when the iteration is terminated at the transition from convergence to divergence. This is achieved by appropriate stopping rules like, for example, the *discrepancy principle* (see [6, Sect. 4.3]). With this stopping rule the iteration is terminated as soon as the residual norm $\|\tilde{y} - T\tilde{x}_n\|$ has been reduced to about the (presumably known) noise level $\|y - \tilde{y}\|$. As follows from the general theory developed in [6] the discrepancy principle yields what are called *order-optimal error bounds* as long as the smoothness assumption (2.8) is satisfied for some $0 < \nu \leq \mu - 1$. A somewhat more sophisticated stopping rule, which leads to order-optimal error bounds for the maximal range $0 < \nu \leq \mu$ has been developed in [10] for our new scheme; we refrain, however, from elaborating on this any further here.

7. Numerical results

We shall illustrate the performance of our method for the model problem `phillips` provided in Hansen's MATLAB regularization toolbox [18]. This problem,

$$\int_{-6}^{6} \phi(s-t)x(t)\,dt \;=\; y(s)\,, \qquad -6 \leq s \leq 6\,, \tag{7.1}$$

with

$$\phi(t) \;=\; \begin{cases} 1 + \cos(t\pi/3)\,, & 0 \leq |t| \leq 3\,, \\ 0\,, & 3 \leq |t|\,, \end{cases} \tag{7.2}$$

and $x = \phi \in \mathcal{X} = \mathcal{L}^2(-6,6)$ has first been suggested in the fundamental paper by Phillips [21] on Tikhonov-Phillips regularization. The corresponding convolution integral operator T is compact, selfadjoint and indefinite; after discretization and appropriate normalization (so that the largest eigenvalue becomes $\lambda = 1$) the smallest eigenvalue of T is $a \approx -0.0209$, so the problem is only mildly indefinite. Figure 7.1 illustrates the contribution of each individual eigenvector (plotted over the associated eigenvalue) to the total spectral mass of x. It can be seen that the negative eigenvalues play a minor but still significant role in the spectral distribution of x. Note that our discretized problem has dimension 128 and there is a whole cluster of positive and negative eigenvalues at the origin which cannot be distinguished in this plot. Still, the operator T is injective.

We now apply the new scheme with $\mu = 1.5$ and $a = -0.1$ to this problem. This may correspond to a situation with given a priori information that $\sigma(T) \subset [-0.1, 1]$. Figure 7.2 shows the errors of the first one hundred iterations given noisy data $\tilde{y}$ with a noise level of 1%, i.e., $\|\tilde{y} - y\|/\|y\| = 0.01$ ($\tilde{y} - y$ is a random Gaussian noise vector). For comparison, the results obtained with the ν-method ($\nu = \mu/2$, see the discussion in Section 5) when applied to the normal equation of this problem are also included in this figure (the dashed line).

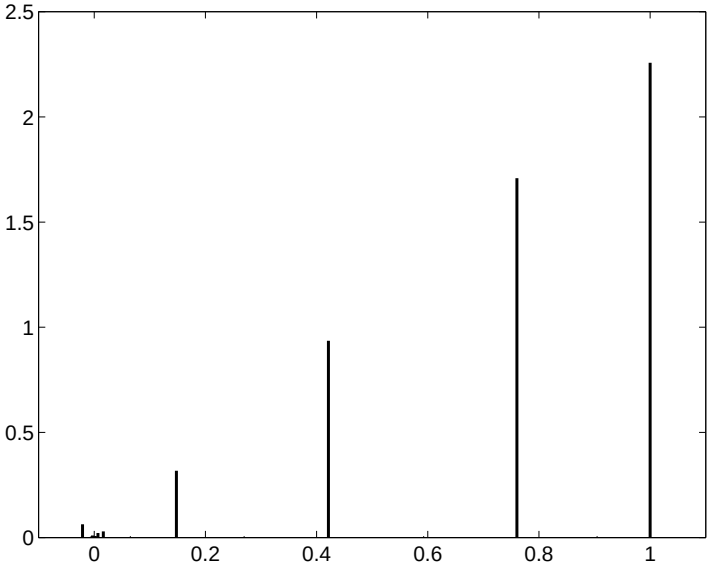

Figure 7.1. Spectral distribution of the exact solution x

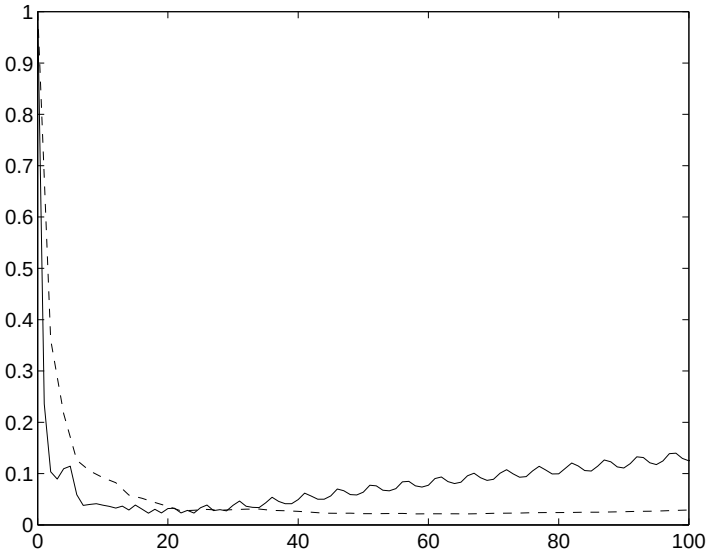

Figure 7.2. Relative error history ($a = -0.1, \mu = 1.5$) vs. iteration count

It can be seen that the new scheme is converging much faster in the initial phase of the iteration, but in the end also diverges more rapidly. Nevertheless, to obtain the best possible approximation of the true solution the new scheme only requires 18 iterations (see Table 7.1), whereas the 'classical' ν-method based on

	indefinite scheme	normal equation scheme
best iterate:	18	67
minimal error:	0.0229	0.0216

Table 7.1
Best regularized reconstruction of x

the normal equation requires more than twice as many iterations to achieve its optimal reconstruction; the quality of the two reconstructions is similar.

We mention that the wiggles in the error history in Figure 7.2 are a typical phenomenon of the new scheme. These wiggles are caused by erratic data noise propagation due to oscillations in the sequence $\{g_n(0)\}$. Our experiments indicate that this phenomenon is more pronounced when a is very close to the origin; the wiggles also appear to be the larger, the smaller is μ.

Another remark concerns the stability of Algorithm 4.1 for the computation of the recursion coefficients $\{a_n, b_n, c_n\}$ for the polynomials $p_n(w_\mu; \cdot)$. Our numerical experiments support the claim in [19] that the computation of these coefficients is pretty stable. In spite of this, however, there has not yet been a formal proof of this statement. Finally, the numerical approximation of the integrals required for the initialization of κ_1 in Algorithm 4.1 also did not lead to a notable deterioration of our iterative scheme.

We conclude with a comment on the competition between semiiterative methods and conjugate gradient type methods. While it is generally acknowledged that conjugate gradient type methods converge faster in terms of iteration count, semiiterative methods may have an advantage in computation time on parallel machines, cf., e.g., [16]. We do not want to elaborate further on this here; we only mention that the conjugate gradient type method MR-II developed in [15] obtains for this particular example the same accuracy with fewer iterations: its best iterate is obtained after seven iterations, the corresponding error equals 0.0200 (compare this with Table 7.1). For a more detailed numerical comparison we refer to [10].

References

[1] V. M. Badkov, *On boundedness in the mean of orthonormalized polynomials*, Math. Notes **13** (1973), 453–459.

[2] ———, *Convergence in the mean and almost everywhere of Fourier series in polynomials orthogonal on an interval*, Math. USSR Sbornik **24(2)** (1974), 223–256.

[3] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing, London, 1998.

[4] H. Brakhage, *On ill-posed problems and the methods of conjugate gradients*, Inverse and Ill-Posed Problems (Orlando) (H. W. Engl and C. W. Groetsch, eds.), Academic Press, 1987, pp. 165–175.

[5] D. Calvetti, L. Reichel, and Q. Zhang, *Iterative exponential filtering for large discrete ill-posed problems*, Numer. Math. **83** (1999), 535–556.

[6] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Doordrecht, 1996.

[7] B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Teubner-Wiley, New York, 1996.

[8] B. Fischer and G. H. Golub, *How to generate unknown orthogonal polynomials out of known orthogonal polynomials*, J. Comput. Appl. Math. **43** (1992), 99–115.

[9] B. Fischer and J. Prestin, *Wavelets based on orthogonal polynomials*, Math. Comp. **66** (1997), 1593–1618.

[10] H. Frankenberger, *Beiträge zur Behandlung linearer inkorrekt gestellter Probleme in Banach- und Hilberträumen*, Dissertation, Universität Kaiserslautern, 1998.

[11] V. Fridman, *Methods of successive approximations for Fredholm integral equations of the first kind*, Uspekhi Mat. Nauk **11(1)** (1956), 233–234, In Russian.

[12] W. Gautschi, *On generating orthogonal polynomials*, SIAM J. Sci. Stat. Comput **3** (1982), 289–317.

[13] G. H. Golub and R. S. Varga, *Chebyshev semiiterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods*, Numer. Math. **3** (1961), 147–168.

[14] M. Hanke, *Accelerated Landweber iterations for the solution of ill-posed equations*, Numer. Math. **60** (1991), 341–373.

[15] ———, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, Longman House, Harlow, 1995.

[16] M. Hanke, M. Hochbruck, and W. Niethammer, *Experiments with Krylov subspace methods on a massively parallel computer*, Appl. Math. **38** (1993), 440–451.

[17] M. Hanke and J. G. Nagy, *Restoration of atmospherically blurred images by symmetric indefinite conjungate gradient techniques*, Inverse Problems **12** (1996), 157–173.

[18] P. C. Hansen, *Regularization tools: A* MATLAB *package for analysis and solution of discrete ill-posed problems*, Numer. Alg. **6** (1994), 1–35.

[19] A. P. Magnus, *Asymptotics for the simplest generalized Jacobi polynomials recurrence coefficients from Freud's equations: numerical explorations*, Ann. Numer. Math. **2** (1995), 311–325.

[20] P. G. Nevai, *Orthogonal polynomials*, Mem. Amer. Math. Soc. **213** (1979).

[21] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. Assoc. Comput. Mach. **9** (1962), 84–97.

[22] E. L. Stiefel, *Kernel polynomials in linear algebra and their numerical applications*, US Nat. Bureau Stand., Appl. Math. Ser. **49** (1958), 1–22.

[23] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. 23, Amer. Math. Soc., Providence, Rhode Island, 1975.

[24] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.