

THE MINIMAL ERROR CONJUGATE GRADIENT METHOD IS A REGULARIZATION METHOD

MARTIN HANKE

(Communicated by Palle E. T. Jorgensen)

ABSTRACT. The regularizing properties of the conjugate gradient iteration, applied to the normal equation of a linear ill-posed problem, were established by Nemirovskii in 1986. A seemingly more attractive variant of this algorithm is the minimal error method suggested by King. The present paper analyzes the regularizing properties of the minimal error method. It is shown that the discrepancy principle is no regularizing stopping rule for the minimal error method. Instead, a different stopping rule is suggested which leads to order-optimal convergence rates.

1. INTRODUCTION

It is well known that many inverse problems in natural sciences are ill-posed in the sense that the solution lacks stability with respect to data perturbations. As a consequence, any resolving algorithm must incorporate some sort of regularization. The different techniques that have been suggested in the past include direct inversion methods like Tikhonov regularization [5] on the one hand, and iterative methods like Landweber iteration [6] on the other hand. Roughly speaking, any regularization method is based on some approximation process, and seeks to balance the approximation error versus the propagated data error. In iterative regularization methods the approximation process is the iteration itself, hence regularization is obtained by stopping the iterative procedure early: *the stopping index is the regularization parameter.*

While the original Landweber iteration is far too slow to be competitive, faster algorithms like the conjugate gradient iteration (CG) for linear ill-posed problems are much more powerful. In fact, for large-scale applications, e.g., discretized higher dimensional problems, direct inversion algorithms are prohibitively expensive, and conjugate gradient type methods may be the only alternative. The regularizing properties of these methods are difficult to analyze,

Received by the editors May 9, 1994.

1991 *Mathematics Subject Classification.* Primary 65J20, 65J10; Secondary 65R30, 45L05, 47A58.

Key words and phrases. Linear ill-posed problems, iterative regularization, conjugate gradients, minimal error method, discrepancy principle, order-optimal error bounds.

This research was supported in part by NATO grant CRG 930044.

and widely depend on the employed stopping rule. For CG (applied to the normal equation) Nemirovskii [10] suggested the discrepancy principle as stopping criterion, and established order-optimal error bounds for the corresponding approximations. Later on, King [8] considered the *minimal error method* (ME) for solving linear ill-posed problems. ME is a variant of CG, which—for exactly given right-hand side data—minimizes the iteration error rather than the data fit in the same subspace. Hence, ME seems to be somewhat superior to CG. However, as King writes at the end of his paper, “the stability of the ME method remains an open question”.

The purpose of this paper is to answer this question. After a brief review of the basic properties of the ME iteration, it will be shown in Sect. 3 that the iteration diverges if and only if the right-hand side does not belong to the range of the given operator. Nevertheless, if the data are small perturbations of some “true data”, then there is an appropriate stopping rule (stated explicitly in Sect. 5) with which ME provides regularized solutions with order-optimal accuracy. A surprising result, on the other hand, is the following: the discrepancy principle that has turned out appropriate for the CG algorithm *fails* for ME. This is shown in Sect. 4.

2. THE MINIMAL ERROR METHOD

Let $T : \mathcal{X} \rightarrow \mathcal{Y}$ be a bounded linear operator between Hilbert spaces \mathcal{X} and \mathcal{Y} with range $\mathcal{R}(T)$. Throughout, for both spaces \mathcal{X} and \mathcal{Y} the same symbol $\|\cdot\|$ will be used for the respective norms. Given the linear equation

$$(2.1) \quad Tx = y,$$

the aim is to recover x from noisy approximations y^δ of y . No preassumption will be stipulated on the nature of the noise, however, it will be assumed that an upper bound δ for the noise level is known, i.e.,

$$(2.2) \quad \|y - y^\delta\| \leq \delta.$$

The problem as posed above is commonly referred to as being ill-posed if the mapping $y \mapsto x$ is discontinuous, i.e., if the generalized inverse T^\dagger (cf. Groetsch [4]) of T is unbounded. Recall that the domain of T^\dagger is

$$\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \overline{\mathcal{R}(T)}^\perp.$$

In the sequel, some of the properties of the two iteration methods CG (applied to the normal equation $T^*Tx = T^*y$) and ME will be reviewed. While the CG method is treated in a number of text books (cf., e.g., Chapters II.5 and III.4 in [4]), details on the ME method (cf. Algorithm 1) can only be found in a few articles, for example in [8]. To fix notation, let x_k^δ and X_k^δ , $k = 0, 1, 2, \dots$, always denote the iterates of ME and CG, respectively. The superscript δ emphasizes that the perturbed right-hand side y^δ is used for the iteration. It is assumed that the iteration starts with $x_0^\delta = X_0^\delta = 0$. Nevertheless, it is easy to incorporate some initial guess x_* by changing the right-hand side of (2.1) to $y^\delta - Tx_*$, and then solving for $x - x_*$.

```

r0 = y - Tx0
d0 = T* r0
k = 0
while (not stop) do
    αk = ||rk||2 / ||dk||2
    xk+1 = xk + αk dk
    rk+1 = rk - αk T dk
    βk+1 = ||rk+1||2 / ||rk||2
    dk+1 = T* rk+1 + βk+1 dk
    k = k + 1
end while.
    
```

ALGORITHM 1. ME iteration

For both algorithms, CG and ME, the k th iterate belongs to the k th Krylov subspace \mathcal{K}_k with respect to T^*y^δ and T^*T , i.e.,

$$\mathcal{K}_k = \text{span}\{T^*y^\delta, (T^*T)T^*y^\delta, (T^*T)^2T^*y^\delta, \dots, (T^*T)^{k-1}T^*y^\delta\}.$$

Consequently, there exist polynomials q_{k-1} and Q_{k-1} of degree $k-1$ with

$$(2.3) \quad x_k^\delta = q_{k-1}(T^*T)T^*y^\delta, \quad X_k^\delta = Q_{k-1}(T^*T)T^*y^\delta.$$

Introducing the so-called residual polynomials

$$(2.4) \quad p_k = 1 - \lambda q_{k-1}, \quad P_k = 1 - \lambda Q_{k-1},$$

one immediately concludes that

$$(2.5) \quad y^\delta - Tx_k^\delta = p_k(TT^*)y^\delta, \quad y^\delta - TX_k^\delta = P_k(TT^*)y^\delta.$$

Note from the definition (2.4) that

$$p_k, P_k \in \Pi_k^0 := \{p : p \text{ is a polynomial of degree } k \text{ with } p(0) = 1\}.$$

An important feature of the two algorithms is their optimality property: while the k th CG iterate X_k^δ minimizes the residual norm $\|y^\delta - Tx\|$ among all elements $x \in \mathcal{K}_k$ (cf. [4, Lemma 2.5.2]), the iterate x_k^δ of the ME method minimizes the error norm

$$\|x - T^\dagger y^\delta\| \longrightarrow \min \quad \text{over } x \in \mathcal{K}_k,$$

provided $y^\delta \in \mathcal{R}(T)$ (cf. [8]). From this follows readily the convergence of the ME iterates x_k^δ to $T^\dagger y^\delta$ in the case $y^\delta \in \mathcal{R}(T)$. However, it must be stressed that the condition $y^\delta \in \mathcal{R}(T)$ will rarely be fulfilled in ill-posed problems.

The key observation for the analysis of the two iterative schemes ME and CG is the fact that the residual polynomials $\{p_k\}$ and $\{P_k\}$ form two systems of real orthogonal polynomials. Denote by $\{E_\lambda\}$ the resolution of the identity of the selfadjoint semidefinite operator TT^* (cf. [4]), and define the inner product

$$(2.6) \quad [\varphi, \psi] := \int_0^\infty \varphi(\lambda)\psi(\lambda) d\|E_\lambda y^\delta\|^2.$$

Then one has (cf. [8] for ME, and [4, Lemma 2.5.4] for CG)

$$(2.7) \quad [p_k, p_l] = 0, \quad [P_k, \lambda P_l] = 0, \quad k, l \in \mathbf{N}_0, \quad k \neq l.$$

Note that, in view of (2.5) and (2.6),

$$(2.8) \quad \begin{aligned} [p_k, p_k] &= \|p_k(TT^*)y^\delta\|^2 = \|y^\delta - TX_k^\delta\|^2, \\ [P_k, P_k] &= \|P_k(TT^*)y^\delta\|^2 = \|y^\delta - TX_k^\delta\|^2. \end{aligned}$$

It must be emphasized that only finitely many orthogonal polynomials exist when the integral in (2.6) degenerates to a finite sum. In this case the CG iteration terminates after finitely many iterations with the exact solution $T^\dagger y^\delta$ of the perturbed problem. If $y^\delta \in \mathcal{R}(T)$, then the same holds true for the ME iteration; otherwise, however, y^δ has a nontrivial component in the orthogonal complement of $\mathcal{R}(T)$ and the iteration breaks down with division by zero. In the sequel it will always be assumed that k is smaller than one of these ultimate termination indices.

Conferring to the orthogonal polynomial terminology (cf. Chihara [1, Theorem I.7.1]), it becomes obvious from (2.7) that the polynomials P_k are multiples of the kernel polynomials corresponding to the inner product (2.6) and the normalization point $\lambda_0 = 0$. From this the optimality property of the CG iteration as stated above follows readily (cf. [1, Theorem I.7.3]), i.e.,

$$(2.9) \quad [P_k, P_k] \leq [p, p] \quad \text{for any } p \in \Pi_k^0;$$

moreover, the value of the minimum is given by the so-called Christoffel number:

$$[P_k, P_k] = (\check{p}_0^2(0) + \check{p}_1^2(0) + \dots + \check{p}_k^2(0))^{-1},$$

where \check{p}_j denotes the orthonormal multiple of p_j . Since $p_k = \check{p}_k/\check{p}_k(0)$, the numbers $\check{p}_k(0)$ can be expressed in terms of the residual polynomials p_k , namely

$$\check{p}_k^{-2}(0) = \check{p}_k^{-2}(0)[\check{p}_k, \check{p}_k] = [p_k, p_k],$$

and hence,

$$[P_k, P_k] = \left(\sum_{j=0}^k [p_j, p_j]^{-1} \right)^{-1}.$$

In view of (2.8) this shows that the residual norm of the k th CG iterate can be determined recursively in the course of the ME iteration via the formula

$$(2.10) \quad \|y^\delta - TX_k^\delta\|^2 = \left(\sum_{j=0}^k \|y^\delta - TX_j^\delta\|^{-2} \right)^{-1},$$

which will be useful in designing an appropriate stopping rule in Sect. 5.

3. A DIVERGENCE RESULT

The aim of this section is to clarify the behavior of the ME method in case the right-hand side y^δ does not belong to $\mathcal{R}(T)$.

Theorem 3.1. *If $y^\delta \notin \mathcal{R}(T)$, then the ME iteration either breaks down after a finite number of steps, or $\|x_k^\delta\| \rightarrow \infty$ as $k \rightarrow \infty$.*

Proof. If the integral in (2.6) is just a finite sum, then the iteration breaks down according to the assumption on y^δ (cf. Sect. 2). Otherwise, the iteration does not terminate, and it has to be shown that boundedness of $\{\|x_k^\delta\|\}$ implies

that $y^\delta \in \mathcal{R}(T)$. To see this, expand q_{k-1} in terms of the orthogonal basis $\{P_j : j = 0, 1, \dots, k-1\}$:

$$q_{k-1} = \sum_{j=0}^{k-1} \frac{[q_{k-1}, \lambda P_j]}{[P_j, \lambda P_j]} P_j.$$

By virtue of (2.3), the norm of the k th iterate can now be rewritten as

$$\|x_k^\delta\|^2 = [q_{k-1}, \lambda q_{k-1}] = \sum_{j=0}^{k-1} \frac{[q_{k-1}, \lambda P_j]^2}{[P_j, \lambda P_j]}.$$

Since, for $0 \leq j \leq k-1$,

$$[q_{k-1}, \lambda P_j] = [\lambda q_{k-1}, P_j] = [1 - p_k, P_j] = [1, P_j],$$

this yields

$$(3.1) \quad \|x_k^\delta\|^2 = \sum_{j=0}^{k-1} \frac{[1, P_j]^2}{[P_j, \lambda P_j]} \geq \sum_{j=0}^{k-1} \frac{[1/\lambda, \lambda P_j]^2}{[P_j, \lambda P_j]}.$$

In other words, if $\|x_k^\delta\|$ remains bounded as $k \rightarrow \infty$, then this implies that $q_{k-1} \rightarrow 1/\lambda$ with respect to the inner product $[\cdot, \lambda \cdot]$, and hence,

$$\|x_k^\delta\|^2 \rightarrow \int_{0+}^{\infty} \frac{1}{\lambda} d\|E_\lambda y^\delta\|^2 < \infty.$$

Of course, this is equivalent to $y^\delta \in \mathcal{D}(T^\dagger)$, and hence, in order to prove that $y^\delta \in \mathcal{R}(T)$ it remains to show that $E_0 y^\delta = 0$. Using (2.4) and the orthogonality of $\{P_j\}$ (cf. (2.7)), the following lower bound is obtained for $[1, P_j]$:

$$[1, P_j] = [P_j, P_j] + [Q_{j-1}, \lambda P_j] = [P_j, P_j] \geq \|E_0 y^\delta\|^2.$$

On the other hand, from the fact that the CG iterates minimize the residual norm in \mathcal{R}_k , one concludes that

$$[P_j, \lambda P_j] = \|T^*(y^\delta - TX_k^\delta)\|^2 \leq \|T\|^2 \|y^\delta - TX_k^\delta\|^2 \leq \|T\|^2 \|y^\delta\|^2.$$

Combining these two estimates it follows from (3.1) that

$$\|x_k^\delta\|^2 \geq \sum_{j=0}^{k-1} \frac{\|E_0 y^\delta\|^4}{\|T\|^2 \|y^\delta\|^2} = k \frac{\|E_0 y^\delta\|^4}{\|T\|^2 \|y^\delta\|^2},$$

and therefore $\{\|x_k^\delta\|\}$ can only be bounded if $E_0 y^\delta = 0$. Thus, it has been shown that $y^\delta \in \mathcal{R}(T)$, and the proof is complete. \square

It must be emphasized that Theorem 3.1 does not imply that the ME iteration process is inappropriate for the regularization of ill-posed problems. Rather, it has been shown that the iteration must not be carried too far to obtain regularized approximations for $T^\dagger y$.

4. THE DISCREPANCY PRINCIPLE

Given the upper bound δ for the perturbation (2.2) of the right-hand side and a fixed parameter $\tau > 1$, Nemirovskii [10] (see also [7]) suggested determining the stopping index $k = k(\delta)$ for the CG iteration as the smallest index for which

$$\|y^\delta - TX_k^\delta\| \leq \tau\delta.$$

This is a familiar generalization of the discrepancy principle for choosing the regularization parameter in Tikhonov regularization (cf. [5, Sect. 3.3]).

Assume that the exact solution $T^\dagger y$ has a certain smoothness, i.e., assume that there exist $\mu > 0$ and $w \in \mathcal{R}$ with

$$(4.1) \quad T^\dagger y = (T^*T)^\mu w, \quad \|w\| = \omega.$$

Nemirovskii has shown that the CG iterate $X_{k(\delta)}^\delta$ corresponding to the above parameter choice satisfies the error bound

$$(4.2) \quad \|T^\dagger y - X_{k(\delta)}^\delta\| \leq c\omega^{1/2\mu+1}\delta^{2\mu/2\mu+1},$$

where c is some positive constant (in the sequel, c is a generic constant with different meanings at different locations). Recall from regularization theory (cf. Louis [9] or Vainikko and Veretennikov [11]) that the error bound (4.2) is best possible under the given assumption (4.1): *any reasonable regularized solution should satisfy (4.2)*.

In the remainder of this section a counterexample will be constructed to show that the discrepancy principle will not even terminate the ME iteration for certain triplets T , y , and y^δ with $y \in \mathcal{R}(T)$ and $y^\delta \notin \mathcal{R}(T)$. A different stopping rule for the ME method which leads to order-optimal error bounds will be considered in the following section.

The construction of the counterexample is based on the particular inner products

$$(4.3) \quad [\varphi, \psi]_0 = \int_0^1 \varphi(\lambda)\psi(\lambda)\lambda^{2\nu}(1-\lambda)^{-1/2} d\lambda$$

and

$$(4.4) \quad [\varphi, \psi] = \delta^2\varphi(0)\psi(0) + [\varphi, \psi]_0,$$

where $\nu > 0$ is some fixed, arbitrary parameter. Think of $[\cdot, \cdot]_0$ and $[\cdot, \cdot]$ as the inner products (2.6) corresponding to some operator T and right-hand sides $y \in \mathcal{R}(T)$ and $y^\delta \notin \mathcal{R}(T)$, respectively. Since $[\lambda^{-\mu-1/2}, \lambda^{-\mu-1/2}]_0 < \infty$, $0 < \mu < \nu$, it follows that $T^\dagger y$ belongs to $\mathcal{R}((T^*T)^\mu)$ for all $0 < \mu < \nu$. By virtue of (4.4), $y^\delta - y$ is a perturbation of norm δ along the orthogonal complement of T 's range.

It is known (cf., e.g., Chihara [2]) that the orthogonal polynomials corresponding to $[\cdot, \cdot]$ of (4.4) can be expressed in terms of the orthogonal polynomials corresponding to $[\cdot, \cdot]_0$ and those corresponding to $[\cdot, \lambda^2 \cdot]_0$. Since the weight function in (4.3) is a translated Jacobi weight, these polynomials are all translated Jacobi polynomials $P_k^{(\alpha, \beta)}$. For the readers convenience, the precise statement of this result (in an appropriate form) and a simple proof of it are included below.

Proposition 4.1. For $k \geq 1$, the k th residual polynomial p_k of the ME iteration corresponding to the inner product (4.4) equals

$$(4.5) \quad p_k = p_k^{(2\nu)} - \vartheta_k \lambda p_{k-1}^{(2\nu+2)}, \quad \vartheta_k = [p_{k-1}^{(2\nu+2)}, \lambda p_{k-1}^{(2\nu+2)}]_0^{-1} \delta^2,$$

where, for $\alpha = 2\nu$ and $2\nu + 2$, respectively,

$$p_j^{(\alpha)}(\lambda) = P_j^{(\alpha, -1/2)}(1 - 2\lambda) / P_j^{(\alpha, -1/2)}(1), \quad j \geq 0.$$

Proof. Let $p \in \Pi_k^0$ be given by the right-hand side of (4.5). Then one obtains from the orthogonality of the Jacobi polynomials that, for $0 < j < k$,

$$[p, \lambda^j] = [p, \lambda^j]_0 = [p_k^{(2\nu)}, \lambda^j]_0 - \vartheta_k [p_{k-1}^{(2\nu+2)}, \lambda^2 \lambda^{j-1}]_0 = 0;$$

on the other hand, for $j = 0$ and $k \geq 1$,

$$[p, 1] = \delta^2 p(0) + [p_k^{(2\nu)}, 1]_0 - \vartheta_k [p_{k-1}^{(2\nu+2)}, \lambda]_0 = \delta^2 - \frac{[p_{k-1}^{(2\nu+2)}, \lambda]_0}{[p_{k-1}^{(2\nu+2)}, \lambda p_{k-1}^{(2\nu+2)}]_0} \delta^2.$$

From the orthogonality it is easy to see that the fraction on the right-hand side equals 1, and hence, p is orthogonal to all polynomials of lower degree with respect to $[\cdot, \cdot]$. \square

Consider now the residual norms for the ME iteration:

$$\begin{aligned} \|y^\delta - Tx_k^\delta\|^2 &= [p_k, p_k] \\ &= \delta^2 + [p_k^{(2\nu)} - \vartheta_k \lambda p_{k-1}^{(2\nu+2)}, p_k^{(2\nu)} - \vartheta_k \lambda p_{k-1}^{(2\nu+2)}]_0 \\ &= \delta^2 + [p_k^{(2\nu)}, p_k^{(2\nu)}]_0 - 2\vartheta_k [p_k^{(2\nu)}, \lambda p_{k-1}^{(2\nu+2)}]_0 + \vartheta_k^2 [p_{k-1}^{(2\nu+2)}, \lambda^2 p_{k-1}^{(2\nu+2)}]_0. \end{aligned}$$

It will be shown next that

$$[p_k^{(2\nu)}, \lambda p_{k-1}^{(2\nu+2)}]_0 < 0.$$

To this end, expand $\lambda p_{k-1}^{(2\nu+2)}$ in terms of $\{p_j^{(2\nu)}\}$, i.e.,

$$\lambda p_{k-1}^{(2\nu+2)} = \sum_{j=0}^k \gamma_{j,k} p_j^{(2\nu)}.$$

Because of their normalization at $\lambda = 0$, it follows that

$$p_{k-1}^{(2\nu+2)}(\lambda) / p_k^{(2\nu)}(\lambda) \longrightarrow -\infty, \quad \lambda \rightarrow +\infty,$$

which implies that $\gamma_{k,k}$ is negative. Consequently,

$$[p_k^{(2\nu)}, \lambda p_{k-1}^{(2\nu+2)}]_0 = \sum_{j=0}^k \gamma_{j,k} [p_k^{(2\nu)}, p_j^{(2\nu)}]_0 = \gamma_{k,k} [p_k^{(2\nu)}, p_k^{(2\nu)}]_0 < 0,$$

as was to be shown. Since ϑ_k is positive (cf. (4.5)), this yields the following lower bound for the ME residuals:

$$(4.6) \quad \|y^\delta - Tx_k^\delta\|^2 \geq \delta^2 + [p_k^{(2\nu)}, p_k^{(2\nu)}]_0 + \delta^4 \frac{[p_{k-1}^{(2\nu+2)}, \lambda^2 p_{k-1}^{(2\nu+2)}]_0}{[p_{k-1}^{(2\nu+2)}, \lambda p_{k-1}^{(2\nu+2)}]_0^2}.$$

Using displays 7.391,1, 7.391,5, and 8.960 in [3], together with Stirling’s formula for the Gamma function, straightforward calculations show that

$$\begin{aligned} [p_k^{(2\nu)}, p_k^{(2\nu)}]_0 &\sim k^{-4\nu-1}, & k \rightarrow \infty, \\ [p_{k-1}^{(2\nu+2)}, \lambda^2 p_{k-1}^{(2\nu+2)}]_0 &\sim k^{-4\nu-5}, & k \rightarrow \infty, \\ [p_{k-1}^{(2\nu+2)}, \lambda p_{k-1}^{(2\nu+2)}]_0 &\sim k^{-4\nu-4}, & k \rightarrow \infty. \end{aligned}$$

Inserting now these asymptotics into (4.6), it follows that there exists $c > 0$, independent of k and δ , such that

$$\|y^\delta - Tx_k^\delta\|^2 \geq \delta^2 + ck^{-4\nu-1} (1 + \delta^4 k^{8\nu+4}), \quad k > 0.$$

The minimum (for $k \in \mathbb{N}$) of the right-hand side is bounded from below by

$$\delta^2 (1 + c\delta^{-1/2\nu+1})$$

for some $c > 0$, and hence, for δ sufficiently small, the norms of the ME residuals never reach the tolerance $\tau\delta$ in the discrepancy principle.

In other words, in this example the discrepancy principle does not terminate the iteration when δ becomes sufficiently small. However, as shown in Theorem 3.1, the iteration diverges for any $\delta > 0$ since $y^\delta \notin \mathcal{R}(T)$.

5. AN ORDER-OPTIMAL STOPPING RULE

The reason for the failure of the discrepancy principle is the suboptimal decay rate of the residual norm. If, in the above example, the squared residual norms for the unperturbed problem (i.e., $[p_k^{(2\nu)}, p_k^{(2\nu)}]_0$) would converge like $k^{-4\nu-2}$ instead of $k^{-4\nu-1}$, then the resulting stopping index would be order-optimal. The former rate, on the other hand, corresponds to the asymptotic behavior of the CG residuals. This is the motivation for using the CG residuals rather than the ME residuals for the stopping criterion. Note that it is possible—by virtue of (2.10)—to compute the norm of the CG residual in the course of the ME iteration.

Stopping Rule 5.1. Fix $\tau > 1$ and terminate the ME iteration when $Tx_k^\delta = y^\delta$, or when, for the first time,

$$\sum_{j=0}^k \|y^\delta - Tx_j^\delta\|^{-2} \geq (\tau\delta)^{-2}.$$

The resulting stopping index will be denoted by $k(\delta)$.

Remark. It follows from (2.10) that $k(\delta)$ is precisely the stopping index that would be chosen by the discrepancy principle for the CG iteration. Hence, as is shown in [10], $k(\delta) < \infty$ is well defined by this stopping rule. Another remark should concern potential breakdowns during the ME iteration. As pointed out in Sect. 2, a breakdown will occur in the k th iterative step, if the integral in (2.6) degenerates to a finite sum of k nonvanishing terms, one of which is $\varphi(0)\psi(0)\|E_0 y^\delta\|^2$. Note that no similar term occurs in the inner product $[\varphi, \lambda\psi]$. It follows that the CG iteration converges after $k - 1$ steps with

$X_{k-1}^\delta = T^\dagger y^\delta$, and hence, $k(\delta) \leq k - 1$. In other words, no breakdown can occur in the ME iteration up to the stopping index $k(\delta)$.

The main result of this paper establishes the order-optimality of this new stopping rule for ME.

Theorem 5.1. *Assume that $y \in \mathcal{R}(T)$, $\|y - y^\delta\| \leq \delta$; and let the stopping index $k(\delta)$ of the ME iteration be chosen by Stopping Rule 5.1. If the exact solution $T^\dagger y$ satisfies (4.1) for some $\mu > 0$, then the following error bound holds for some constant c depending on μ only:*

$$\|T^\dagger y - x_{k(\delta)}^\delta\| \leq c \omega^{1/2\mu+1} \delta^{2\mu/2\mu+1}.$$

Proof. For brevity, let $k = k(\delta)$ throughout the proof. The key idea of the proof is to compare x_k^δ and X_k^δ , since the CG iterate X_k^δ satisfies the asserted bound (cf. (4.2)). It therefore remains to show that

$$(5.1) \quad \|x_k^\delta - X_k^\delta\| \leq c \omega^{1/2\mu+1} \delta^{2\mu/2\mu+1}$$

for some uniform constant $c > 0$. For $k = 0$ this is obviously true since the left-hand side of (5.1) is zero. For $k \geq 1$,

$$(5.2) \quad x_k^\delta - X_k^\delta = T^* q_{k-1} (TT^*) y^\delta - T^* Q_{k-1} (TT^*) y^\delta = \theta_{k-1} T^* d_{k-1} (TT^*) y^\delta,$$

where $\theta_{k-1} = q_{k-1}(0) - Q_{k-1}(0)$ is such that

$$(5.3) \quad d_{k-1} = \frac{q_{k-1} - Q_{k-1}}{\theta_{k-1}} = \frac{P_k - p_k}{\theta_{k-1} \lambda} \in \Pi_{k-1}^0.$$

For θ_{k-1} one can derive another representation by using the orthogonality relations (2.7) for the residual polynomials p_k and P_k (note that $(d_{k-1} - P_k)/\lambda$ is a polynomial of degree $k - 1$):

$$\begin{aligned} \theta_{k-1} [d_{k-1}, \lambda d_{k-1}] &= \left[\frac{P_k - p_k}{\lambda}, \lambda d_{k-1} \right] = [P_k, \lambda \frac{d_{k-1}}{\lambda}] - [p_k, d_{k-1}] \\ &= [P_k, \lambda \frac{d_{k-1} - P_k}{\lambda}] + [P_k, P_k] = [P_k, P_k]. \end{aligned}$$

This yields

$$\theta_{k-1} = \frac{[P_k, P_k]}{[d_{k-1}, \lambda d_{k-1}]},$$

and one obtains, by inserting this into (5.2),

$$\|x_k^\delta - X_k^\delta\|^2 = \frac{[P_k, P_k]^2}{[d_{k-1}, \lambda d_{k-1}]^2} [d_{k-1}, \lambda d_{k-1}] = \frac{[P_k, P_k]}{[d_{k-1}, \lambda d_{k-1}]} [P_k, P_k].$$

Together with (2.9) this leads to the inequality

$$(5.4) \quad \|x_k^\delta - X_k^\delta\|^2 \leq \frac{[d_{k-1}, d_{k-1}]}{[d_{k-1}, \lambda d_{k-1}]} [P_k, P_k].$$

Now, denote the zeros of P_k by $\{\lambda_{j,k} : j = 1, \dots, k\}$, with $\lambda_{1,k}$ and $\lambda_{k,k}$ the smallest and the largest zero, respectively. It is known (cf., e.g., Chihara [1, p. 36]) that the values $\{p_k(\lambda_{j,k})\}$ of the polynomial p_k alternate in sign as j runs from 1 through k , and by virtue of (5.3) the same must be true for

$\{d_{k-1}(\lambda_{j,k})\}$ since the $\lambda_{j,k}$ are positive. This implies that all $k-1$ roots of d_{k-1} belong to the interval $[\lambda_{1,k}, \lambda_{k,k}]$, and hence,

$$(5.5) \quad 0 \leq d_{k-1}(\lambda) \leq 1, \quad 0 \leq \lambda \leq \lambda_{1,k}.$$

To complete the proof, another result of Nemirovskii's analysis of the CG iteration is required (cf. [7, eq. (4.2)]), namely the fact that one can find c_ε with $0 < c_\varepsilon < \tau - 1$, such that (recall that $k = k(\delta)$)

$$\lambda_{1,k} \geq |P'_k(0)|^{-1} \geq (c_\varepsilon \frac{\delta}{\omega})^{2/2\mu+1} =: \varepsilon.$$

In view of this result, (5.5) holds true for $0 \leq \lambda \leq \varepsilon$, and therefore

$$\begin{aligned} [d_{k-1}, d_{k-1}] &= \|E_\varepsilon d_{k-1}(TT^*)y^\delta\|^2 + \|(I - E_\varepsilon)d_{k-1}(TT^*)y^\delta\|^2 \\ &\leq \|E_\varepsilon y^\delta\|^2 + \|(I - E_\varepsilon)d_{k-1}(TT^*)y^\delta\|^2 \\ &\leq (\|y^\delta - y\| + \|E_\varepsilon T(T^*T)^\mu \omega\|)^2 + \|(I - E_\varepsilon)d_{k-1}(TT^*)y^\delta\|^2 \\ &\leq (\delta + \varepsilon^{\mu+1/2}\omega)^2 + \varepsilon^{-1} \|(I - E_\varepsilon)T^* d_{k-1}(TT^*)y^\delta\|^2 \\ &\leq (1 + c_\varepsilon)^2 \delta^2 + \varepsilon^{-1} [d_{k-1}, \lambda d_{k-1}]. \end{aligned}$$

By definition of $k = k(\delta)$, and by the optimality property (2.9) of P_{k-1} ,

$$\tau^2 \delta^2 \leq [P_{k-1}, P_{k-1}] \leq [d_{k-1}, d_{k-1}],$$

and hence,

$$[d_{k-1}, d_{k-1}] \leq \left(\frac{1 + c_\varepsilon}{\tau}\right)^2 [d_{k-1}, d_{k-1}] + \varepsilon^{-1} [d_{k-1}, \lambda d_{k-1}].$$

In other words, since $c_\varepsilon < \tau - 1$, a uniform constant $c > 0$ has been found, such that

$$\frac{[d_{k-1}, d_{k-1}]}{[d_{k-1}, \lambda d_{k-1}]} \leq c \left(\frac{\omega}{\delta}\right)^{2/2\mu+1}.$$

Inserting this last inequality into (5.4) establishes

$$\|x_k^\delta - X_k^\delta\|^2 \leq c \left(\frac{\omega}{\delta}\right)^{2/2\mu+1} [P_k, P_k] \leq c \tau^2 \omega^{2/2\mu+1} \delta^{4\mu/2\mu+1},$$

since $[P_k, P_k] \leq \tau^2 \delta^2$ by the definition of the stopping index. Thus, (5.1) holds true, and the proof is complete. \square

REFERENCES

1. T. S. Chihara, *An introduction to orthogonal polynomials*, Gordon and Breach, New York, 1978.
2. ———, *Orthogonal polynomials and measures with end point masses*, Rocky Mountain J. Math. **15** (1985), 705–719.
3. I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic Press, New York, San Francisco, and London, 1965.
4. C. W. Groetsch, *Generalized inverses of linear operators*, Marcel Dekker, New York and Basel, 1977.
5. ———, *The theory of Tikhonov regularization for Fredholm equations of the first kind*, Pitman, Boston, London, and Melbourne, 1984.

6. M. Hanke, *Accelerated Landweber iterations for the solution of ill-posed equations*, Numer. Math. **60** (1991), 341–373.
7. ———, *A second look at Nemirovskii's analysis of the conjugate gradient method*, Beiträge zur Angewandten Analysis und Informatik (E. Schock, ed.), Verlag Shaker, Aachen, 1994, pp. 123–135.
8. J. T. King, *A minimal error conjugate gradient method for ill-posed problems*, J. Optim. Theory Appl. **60** (1989), 297–304.
9. A. K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.
10. A. S. Nemirovskii, *The regularization properties of the adjoint gradient method in ill-posed problems*, USSR Comput. Math. and Math. Phys. **26** (1986), no. 2, 7–16.
11. G. M. Vainikko and A. Y. Veretennikov, *Iteration procedures in ill-posed problems*, Nauka, Moscow, 1986. (Russian)

INSTITUT FÜR PRAKTISCHE MATHEMATIK, UNIVERSITÄT KARLSRUHE, D-76128 KARLSRUHE,
GERMANY

E-mail address: hanke@math.uni-karlsruhe.de