ELSEVIER

Linear Algebra and its Applications 298 (1999) 99–113

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

A unified approach to Krylov subspace methods for the Drazin-inverse solution of singular nonsymmetric linear systems

Avram Sidi

Computer Science Department, Technion, Israel Institute of Technology, Haifa 32000, Israel

Received 3 March 1999; accepted 13 July 1999

Submitted by M. Neumann

Abstract

Consider the linear system $Ax = b$, where $A \in \mathbb{C}^{N\times N}$ is a singular matrix. In the present work we propose a general framework within which Krylov subspace methods for Drazin-inverse solution of this system can be derived in a convenient way. The Krylov subspace methods known to us to date treat only the cases in which A is hermitian and its index $\mathrm{ind}(A)$ is unity necessarily. In the present work A is not required to be hermitian. It can have any type of spectrum and $\mathrm{ind}(A)$ is arbitrary. We show that, as is the case with nonsingular systems, the Krylov subspace methods developed here terminate in a finite number of steps that is at most $N - \mathrm{ind}(A)$. For one of the methods derived here we also provide an analysis by which we are able to bound the errors, the relevant bounds decreasing with increasing dimension of the Krylov subspaces involved. The results of this paper are applicable to consistent systems as well as to inconsistent ones. An interesting feature of the approach to singular systems presented in this work is that it is formulated as a generalization of the standard Krylov subspace approach to nonsingular systems. Indeed, our approach here reduces to that relevant for nonsingular systems upon setting $\mathrm{ind}(A) = 0$ everywhere. © 1999 Elsevier Science Inc. All rights reserved.

AMS classification: 15A06; 15A09; 65F10; 65F50

Keywords: Singular linear systems; Drazin-inverse solution; Projection methods; Krylov subspace methods

E-mail address: asidi@cs.technion.ac.il (A. Sidi)

0024-3795/99/$ - see front matter © 1999 Elsevier Science Inc. All rights reserved.
PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 1 5 3 - 6

1. Introduction

Consider the linear system

$$Ax = b, \tag{1.1}$$

where $A \in \mathbb{C}^{N\times N}$ is a singular matrix and $\mathrm{ind}(A)$ is arbitrary. Here $\mathrm{ind}(A)$, the index of A, is the size of the largest Jordan block corresponding to the zero eigenvalue of A. The purpose of this paper is to present a general framework within which Krylov subspace methods for the Drazin-inverse solution of (1.1) can be developed and their properties discussed. We recall that the Drazin-inverse solution of (1.1) is the vector $A^{\mathrm{D}}b$, where A^{D} denotes the Drazin inverse of the singular matrix A. For the Drazin inverse and its properties, see e.g., [2] or [4].

We do not put any restriction on the matrix A. Thus, A is not necessarily hermitian or hermitian positive semidefinite. In addition to its index being arbitrary, its spectrum can be of any type. Neither do we put any restrictions on the linear system (1.1). This system may be consistent or inconsistent. We only assume that $\mathrm{ind}(A)$ is known.

Before we embark on the subject matter of this paper it may be worth reminding ourselves that Krylov subspace methods are meant to be applied especially to linear systems that involve sparse matrices, which is where they are most useful.

The subject of Krylov subspace methods for computing Drazin-inverse solutions has been treated in several papers. First, the method of Conjugate Gradients (CG) can be applied when A is hermitian positive semidefinite and (1.1) is consistent, see [13]. It is shown in [18] that the method of Arnoldi [1] and the method of Generalized Conjugate Residuals (GCR) of Eisenstat et al. [8] and the method of Lanczos [14] as well can be applied to nonhermitian but consistent singular systems when $\mathrm{ind}(A)$ is unity, and error bounds are also given. In addition, Sidi [18] provides a complete convergence theory for these methods and others in the presence of initial iterations via the Richardson iterative scheme.

The treatment of the singular *inconsistent* systems by Krylov subspace methods has proved to be much harder, however. This has been so even for the simplest cases in which $\mathrm{ind}(A) = 1$. To date we are aware of the CG type methods of Calvetti et al. [3] that apply to hermitian systems only. A recent work by Fischer et al. [9] provides a class of methods that form a slight generalization of those of [3] and apply to the same problems. It must be mentioned that it is not only the inconsistent singular systems that have caused problems. Singular consistent systems with $\mathrm{ind}(A) > 1$ have proved to be just as hard since the Drazin inverse solution $A^{\mathrm{D}}b$ of such a system $Ax = b$ is not necessarily an ordinary solution, that is to say, $A(A^{\mathrm{D}}b) = b$ does not necessarily hold.

Finally, we mention the vector extrapolation methods developed in [19] for treating the most general case of singular nonhermitian inconsistent systems with arbitrary $\mathrm{ind}(A)$. This paper too contains a detailed convergence analysis for the methods developed in it. This analysis is carried out in the presence of initial iterations with the Richardson iterative scheme.

In view of the above, the present work seems to be the first to present a unified framework for Krylov subspace methods for singular, hermitian or nonhermitian, and consistent or inconsistent linear systems with arbitrary index.

In Section 2 we give some technical preliminaries that we use in the rest of the paper. In Section 3 we present a general framework of projection methods within which many different methods for the Drazin-inverse solution can be defined in a very simple way. This framework is based on a careful study of the subspaces that play a role in the Drazin-inverse solution and of the nature of the residual vectors r_m. A pleasant feature of the approach we present is that when we set $\mathrm{ind}(A) = 0$ we fall back to the projection method formalism for nonsingular systems. In this sense our approach is a bona fide generalization of the approach taken to the solution of nonsingular systems by projection methods.

In Section 4 we show that the projection methods generated in Section 3 terminate in a finite number of steps, this number being at most $N - \mathrm{ind}(A)$. The main results of this section are Theorems 4.2 and 4.3.

In Section 5 we give some actual Krylov subspace methods of Arnoldi, GCR, and Lanczos types, and show their connection with the existing methods for symmetric systems.

In Section 6 we derive error bounds for the GCR type method of Section 5. The main result of this section is Theorem 6.1.

In the remainder of this paper we denote $\mathrm{ind}(A)$ by a for short.

2. General preliminaries

We shall consider methods that start with an arbitrary initial vector x_0 and compute a sequence of vectors $x_1, x_2, \ldots,$ that are of the general form

$$x_m = x_0 + q_{m-1}(A)r_0; \quad r_0 = b - Ax_0, \tag{2.1}$$

where $q_{m-1}(\lambda)$ is a polynomial in λ of degree at most $m - 1$. Let us define

$$p_m(\lambda) = 1 - \lambda q_{m-1}(\lambda). \tag{2.2}$$

We call $p_m(\lambda)$ the mth residual polynomial since

$$r_m = b - Ax_m = p_m(A)r_0. \tag{2.3}$$

Note that

$$p_m(0) = 1. \tag{2.4}$$

As shown in [6], necessary and sufficient conditions for the convergence of all sequences $\{x_m\}_{m=0}^{\infty}$ generated as above are that

$$\lim_{m\to\infty} p_m^{(i)}(0) = 0, \quad i = 1, 2, \ldots, a; \quad a = \mathrm{ind}(A) \tag{2.5}$$

and

$$\lim_{m\to\infty} p_m^{(i)}(\lambda_j) = 0, \quad i = 0, 1, \ldots, k_j - 1, \tag{2.6}$$

where λ_j are the nonzero eigenvalues of A and $k_j = \mathrm{ind}(A - \lambda_j I)$.

The conditions in (2.5) will, of course, be satisfied if

$$p_m^{(i)}(0) = 0, \quad i = 1, 2, \ldots, a, \quad \text{for all } m = 0, 1, \ldots \tag{2.7}$$

Our purpose in this work is to discuss the design of methods that will generate vectors x_m as described above, such that their corresponding polynomials $p_m(\lambda)$ satisfy (2.7) instead of (2.5), in addition to (2.4).

Before going on it will be convenient to introduce some notation that has been used before in [11,5]. We shall denote by Π_m the set of all polynomials of degree at most m. We shall also define

$$\Pi_m^0 = \{p \in \Pi_m\colon p(0) = 1 \text{ and } p^{(i)}(0) = 0,\ i = 1, \ldots, a\}. \tag{2.8}$$

Clearly, Π_m^0 is the collection of all polynomials of degree at most m that satisfy (2.4) and (2.7). Thus, the polynomials $p_m(\lambda)$ that we will be considering in the present work are all in Π_m^0.

Note that $p_m(\lambda) = 1$ is the only member of Π_m^0 for $m = 0, 1, \ldots, a$, while for $m > a$ all polynomials in Π_m^0 are of the form $p(\lambda) = 1 - \sum_{i=1}^{m-a} c_i \lambda^{a+i}$.

Finally, we will work with the standard Euclidean inner product

$$(x, y) \equiv x^* y \tag{2.9}$$

for which $(y, x) = \overline{(x, y)}$ and $(\alpha x, \beta y) = \overline{\alpha}\beta(x, y)$ for any $\alpha, \beta \in \mathbb{C}$ and any $x, y \in \mathbb{C}^N$. Also, by x is orthogonal to y we shall mean $(x, y) = 0$.

In addition, we will let $\|x\|$ stand for the l_2-norm of $x \in \mathbb{C}^N$, i.e., $\|x\| = \sqrt{(x, x)}$. We will also denote by $\|B\|$, $B \in \mathbb{C}^{N\times N}$, the norm of the matrix B induced by the l_2-norm in $\mathbb{C}^N$.

3. Derivation of Krylov subspace methods

3.1. *Review of some facts about* $\{x_m\}$

Let us denote by $\hat{\mathscr{S}}$ the direct sum of the invariant subspaces of A corresponding to its nonzero eigenvalues λ_j, and by $\tilde{\mathscr{S}}$, its invariant subspace corresponding to its zero eigenvalue. Thus, $\hat{\mathscr{S}}$ is $\mathscr{R}(A^a)$, the range of A^a, and $\tilde{\mathscr{S}}$ is $\mathscr{N}(A^a)$, the nullspace of A^a. Every vector in $\mathbb{C}^N$ can be written as the sum of two unique vectors, one in $\hat{\mathscr{S}}$ and the other in $\tilde{\mathscr{S}}$.

Let us resolve b. Then $b = \hat{b} + \tilde{b}$, where $\hat{b} \in \hat{\mathscr{S}}$ and $\tilde{b} \in \tilde{\mathscr{S}}$, and $A^{\mathrm{D}}b$, the Drazin-inverse solution to $Ax = b$, is the unique vector in $\hat{\mathscr{S}}$ that satisfies the consistent system $Ax = \hat{b}$.

Let us also resolve x_0. Then $x_0 = \hat{x}_0 + \tilde{x}_0$, where $\hat{x}_0 \in \hat{\mathscr{S}}$ and $\tilde{x}_0 \in \tilde{\mathscr{S}}$, and, as shown in Theorem 4.1 of Climent et al. [5],

$$x_m - A^{\mathrm{D}}b = p_m(A)(\hat{x}_0 - A^{\mathrm{D}}b) + \tilde{x}_0. \tag{3.1}$$

That is to say, $x_m = \hat{x}_m + \tilde{x}_m$ with $\hat{x}_m = A^{\mathrm{D}}b + p_m(A)(\hat{x}_0 - A^{\mathrm{D}}b) \in \hat{\mathscr{S}}$ and $\tilde{x}_m = \tilde{x}_0 \in \tilde{\mathscr{S}}$ for all m. Obviously, we can eliminate $\tilde{x}_m = \tilde{x}_0$ by picking $x_0 = 0$ or $x_0 = A^a\xi$ for any $\xi \in \mathbb{C}^N$. In either case $\tilde{x}_0 = 0$.

Let us denote $\hat{e}_m = \hat{x}_m - A^{\mathrm{D}}b,\ m = 0, 1, \ldots,$ for short. Clearly, we want to drive $\hat{e}_m$ to zero. Now as $\hat{e}_0 \in \hat{\mathscr{S}}$ and $\hat{e}_m = p_m(A)\hat{e}_0$, the behavior of $\hat{e}_m$ is determined by the action of $p_m(A)$ on $\hat{\mathscr{S}}$. Of course, what we want is to make $p_m(A)$ on $\hat{\mathscr{S}}$ small with increasing m. The question now is how to choose $p_m(\lambda)$ to achieve precisely this goal.

When we know that the nonzero spectrum of A is contained in a certain set Ω of the complex λ-plane, then we can try to determine $p_m(\lambda)$ such that $|p_m(\lambda)|$ on Ω is sufficiently small. Better still, we may be able to determine $p_m(\lambda)$ by minimizing some norm of it that is defined on Ω. This kind of an approach has been taken in [11,5] for the case in which Ω is a real positive interval and the minimization is carried out in some l_2-norm. The treatment of [11] assumes $a = 1$, while that of [5] is valid for all values of a. Now these methods are semi-iterative methods and not projection methods. In this paper we are interested in projection methods that require no knowledge of the spectrum, and we should consider other ways of determining $p_m(\lambda)$.

3.2. *Review of projection methods for nonsingular systems*

At this point it is important to recall some essential facts about projection methods for the solution of $Bx = f$ when $B \in \mathbb{C}^{N\times N}$ is nonsingular. We choose an arbitrary vector x_0 and two m-dimensional subspaces V and W. Then we look for an approximation to the solution of $Bx = f$ that is of the form $x_m = x_0 + z_m,\ z_m \in V$, such that $r_m = f - Bx_m = r_0 - Bz_m$ is orthogonal to all vectors in W. If we denote by V and W also the $N \times m$ matrices whose columns span the subspaces V and W, respectively, then x_m is uniquely determined and given by $x_m = x_0 + V(W^*BV)^{-1}W^*r_0$, provided $\det(W^*BV) \neq 0$.

First, by the assumption that B is nonsingular, and by the fact that $r_m = -B(x_m - s)$, where s is the solution of $Bx = f$, we know that $\|r_m\|$ is a true norm of the error $(x_m - s)$, and hence driving r_m to zero is equivalent to driving x_m to s. Next, we expect r_m to decrease towards zero as we enlarge the subspaces V and W. The reason for this expectation is that r_m is becoming orthogonal to more and more vectors as we are increasing m. Indeed, if $m = N$, then $r_m = 0$ since any vector in $\mathbb{C}^N$ that is orthogonal to N linearly independent vectors must be the zero vector. We also know that r_m decreases on the average with increasing m in case of the Krylov subspace methods such as the method of Arnoldi, the method of Generalized Conjugate Residuals (GCR), and the method of Lanczos, all of which are projection methods whose right and left subspaces are Krylov subspaces. (For all three methods $V = \mathscr{K}_m(B; r_0)$, and $W = V$ for the method of Arnoldi, $W = \mathscr{K}_m(B; Br_0)$ for GCR, and $W = \mathscr{K}_m(B^*; \check{r}_0)$ with some arbitrary $\check{r}_0 \in \mathbb{C}^N$ for the method of

Lanczos. Here $\mathscr{K}_m(B;u) \equiv \text{span}\{u, Bu, \ldots, B^{m-1}u\}$ is an m-dimensional Krylov subspace.)

3.3. *General projection methods for $A^{\mathrm{D}}b$*

We now want to be able to use the methodology described above that involves computational work with the residual r_m to construct projection methods to approximate $A^{\mathrm{D}}b$. We first realize that when the system $Ax = b$ is inconsistent, we necessarily have $r_m = b - Ax_m \neq 0$ for all m, from which it becomes obvious that r_m cannot be driven to zero and hence direct application of the projection approach, as described in the previous section, will lead nowhere. Furthermore, $r_m \neq 0$ for all m may be the case even when $Ax = b$ is a consistent system, if $\text{ind}(A) > 1$. To see this we begin with $r_0 = b - Ax_0 = \hat{r}_0 + \tilde{r}_0$, where $\hat{r}_0 \in \hat{\mathscr{S}}$ and $\tilde{r}_0 \in \tilde{\mathscr{S}}$, which, upon substituting in (2.3), and using the fact that $p_m(A)v = v$ for $v \in \tilde{\mathscr{S}}$, gives

$$r_m = p_m(A)\hat{r}_0 + \tilde{r}_0. \tag{3.2}$$

That is to say, $r_m = \hat{r}_m + \tilde{r}_m$ with $\hat{r}_m = p_m(A)\hat{r}_0 \in \hat{\mathscr{S}}$ and $\tilde{r}_m = \tilde{r}_0 \in \tilde{\mathscr{S}}$ for all m. Thus $r_m \neq 0$ for all m if $\tilde{r}_0 \neq 0$. We note that $\tilde{r}_0 \neq 0$ in general for arbitrary x_0. Even when we take $x_0 = 0$ so that $\tilde{x}_0 = 0$, we may have $\tilde{r}_0 \neq 0$ as now $r_0 = b$ so that $\tilde{r}_0 = \tilde{b}$, and $Ax = b$ may be consistent even with $\tilde{b} \neq 0$ when $\text{ind}(A) > 1$, as shown in Theorem 2.1 in [18]. (When $Ax = b$ is inconsistent, $\tilde{r}_0 \neq 0$ is always the case. This also follows from Theorem 2.1 in [18].)

We have thus shown that direct application of the projection approach of the previous section has no meaning when the singular system $Ax = b$ is inconsistent, or consistent but $\text{ind}(A) > 1$, and we should look for a suitable modification of this approach.

Invoking now the fact that $\hat{r}_0 = -A(\hat{x}_0 - A^{\mathrm{D}}b)$ (which is implied by $AA^{\mathrm{D}}b = \hat{b}$) in $\hat{r}_m = p_m(A)\hat{r}_0$, we have

$$\hat{r}_m = -Ap_m(A)\hat{e}_0 = -A\hat{e}_m, \tag{3.3}$$

where, we recall, $\hat{e}_m = \hat{x}_m - A^{\mathrm{D}}b,\ m = 0, 1, \ldots$ Thus, the behavior of $\hat{r}_m$, just as that of $\hat{e}_m$, is also determined by the action of $p_m(A)$ on the subspace $\hat{\mathscr{S}}$. If $p_m(A)$ on $\hat{\mathscr{S}}$ is small, then, just like $\hat{e}_m$, $\hat{r}_m$ will be small too.

So far we know that $\hat{e}_m$ and $\hat{r}_m$ will be small simultaneously if $p_m(A)$ on $\hat{\mathscr{S}}$ is small. We also observe that $\hat{r}_m$ will be small if $\hat{e}_m$ is. At this point, it is important to remind ourselves (i) that we aim at making $\hat{e}_m$ small and (ii) that we would like to concentrate on projection methods in which what we are allowed to compute are essentially residuals. We, therefore, ask whether we can make $\hat{e}_m$ small by making $\hat{r}_m$ small. Since $\hat{r}_m = -A\hat{e}_m$ and A is singular, the answer to this question is not immediate. It is in the affirmative, however.

To show this we consider the restriction of the operator A to $\hat{\mathscr{S}}$. Let us denote this restriction by $\hat{A}$. We have $Av = \hat{A}v$ for any $v \in \hat{\mathscr{S}}$. As $\hat{\mathscr{S}}$ is the direct sum of the invariant subspaces of A corresponding only to its *nonzero* eigenvalues λ_j, the

eigenvalues of $\hat{A}$ are also these *nonzero* eigenvalues λ_j. Therefore, $\hat{A}$ is nonsingular, and so are $\hat{A}^k$, $k = 2, 3, \ldots$ Let us order the singular values $\sigma_i^{(k)}$ of $\hat{A}^k$ such that $\sigma_1^{(k)} \geqslant \sigma_2^{(k)} \geqslant \cdots \geqslant \sigma_d^{(k)} > 0$. Here $d = \dim \hat{\mathscr{S}}$. Then

$$\|\hat{r}_m\| = \|A\hat{e}_m\| = \|\hat{A}\hat{e}_m\| \geqslant \sigma_d^{(1)}\|\hat{e}_m\|,$$

which means that $\hat{r}_m \to 0$ implies $\hat{e}_m \to 0$.

Following the discussion given above, we would like to find ways of making $\hat{r}_m$ small. As $\hat{r}_m \in \hat{\mathscr{S}}$, it is clear that we should somehow do all our computational work in $\hat{\mathscr{S}}$. We recall, however, that $r_m = \hat{r}_m + \tilde{r}_m$ with $\hat{r}_m \in \hat{\mathscr{S}}$ and $\tilde{r}_m \in \tilde{\mathscr{S}}$ and with $\tilde{r}_m = \tilde{r}_0 \neq 0$ always when $Ax = b$ is inconsistent, and that $\tilde{r}_0 \neq 0$ is possible also when $Ax = b$ is consistent but $\operatorname{ind}(A) > 1$. This implies that, in general, $r_m \notin \hat{\mathscr{S}}$ holds for all m. We should, therefore, *force* our computations into $\hat{\mathscr{S}}$ in a suitable manner. One way of achieving this is by working with $A^a r_m$ since $A^a r_m = A^a \hat{r}_m \in \hat{\mathscr{S}}$. (Recall that $\tilde{\mathscr{S}} = \mathscr{N}(A^a)$.) Obviously, if $\hat{r}_m$ is small, so is $A^a r_m$. The question is whether $\hat{r}_m$ is small when $A^a r_m$ is. Arguing as in the case of $\hat{r}_m$ versus $\hat{e}_m$, we have

$$\|A^a r_m\| = \|A^a \hat{r}_m\| = \|\hat{A}^a \hat{r}_m\| \geqslant \sigma_d^{(a)}\|\hat{r}_m\|.$$

Consequently, we have that $A^a r_m \to 0$ implies $\hat{r}_m \to 0$, which is what we desire.

We are now at the stage where we decide to make $A^a r_m$ small, as this will make $\hat{r}_m$ small, which, in turn, will make $\hat{e}_m$ small. In analogy to projection methods for nonsingular systems, let us now choose a subspace W of some appropriate dimension and require that $A^a r_m$ (and *not* r_m) be orthogonal to every vector in W. Now, due to the fact that $p_m \in \Pi_m^0$, we have $p_m(\lambda) = 1 - \sum_{i=1}^{m-a} c_i \lambda^{a+i}$. Since the number of the unknown constants c_i is $m - a$, the dimension of W should be $m - a$ too.

Let us now express things in mathematical terms. From (2.1)–(2.3) and $p_m(\lambda) = 1 - \sum_{i=1}^{m-a} c_i \lambda^{a+i}$ and $q_{m-1}(\lambda) = \sum_{i=1}^{m-a} c_i \lambda^{a+i-1}$, we have

$$x_m = x_0 + \sum_{i=1}^{m-a} c_i A^{a+i-1} r_0 \quad \text{and} \quad r_m = r_0 - \sum_{i=1}^{m-a} c_i A^{a+i} r_0. \tag{3.4}$$

Let us define the $N \times (m - a)$ matrix V and the $(m - a)$-dimensional column vector c by

$$V = [A^a r_0 | A^{a+1} r_0 | \cdots | A^{m-1} r_0] \quad \text{and} \quad c = [c_1, \ldots, c_{m-a}]^{\mathrm{T}}. \tag{3.5}$$

Then

$$x_m = x_0 + Vc \quad \text{and} \quad r_m = r_0 - AVc. \tag{3.6}$$

Let us denote by W also the $N \times (m - a)$ matrix whose columns form a basis for W. Then orthogonality of $A^a r_m$ to all vectors in W is equivalent to $W^* A^a r_m = 0$, hence to

$$W^* A^{a+1} V c = W^* A^a r_0. \tag{3.7}$$

Assuming that $\det(W^* A^{a+1} V) \neq 0$, (3.7) has a unique solution for c, and the vector x_m is now given by

$$x_m = x_0 + V(W^* A^{a+1} V)^{-1} W^* A^a r_0. \tag{3.8}$$

Note that the columns of the matrix V in (3.5) play the role of a basis for a right subspace V, which is the Krylov subspace $\mathscr{K}_{m-a}(A;\, A^a r_0)$. (We can make the treatment above more general by replacing V in (3.5) by $V = A^a[v_1|v_2|\cdots|v_{m-a}]$ where $v_1, v_2, \ldots,$ are arbitrary linearly independent vectors. In this case too $Vc \in \hat{\mathscr{S}}$ as desired, and (3.7) and (3.8) remain valid.)

Before we end this section we would like to note that the general framework that we have presented for the Drazin-inverse solution of singular systems by projection methods reduces exactly to that pertaining to nonsingular systems when we set $a = 0$ everywhere. Thus, our general framework is a bona fide generalization of that for nonsingular systems in the previous subsection. This is a very pleasant feature of our approach to projection methods for Drazin-inverse solutions.

4. Finite termination property of projection methods

Our aim in this section is to prove that, in the absence of a breakdown, $x_m = A^{\mathrm{D}} b + \tilde{x}_0$ for some finite $m \leqslant N$. Equivalently, subject to certain conditions of regularity, the projection methods of the previous section terminate successfully in a finite number of steps. This property puts our projection methods for the Drazin-inverse solution on even firmer grounds.

For the sake of convenience we shall make the following definition.

Definition 4.1. Let A be singular with $\operatorname{ind}(A) = a$. We shall call $P(\lambda)$ *the minimal a-incomplete polynomial of A with respect to the vector* $\hat{u} \in \hat{\mathscr{S}} = \mathscr{R}(A^a)$ if $P \in \Pi_m^0$ and m is smallest possible such that $P(A)\hat{u} = 0$.

Definition 4.1 generalizes the concept of *the minimal polynomial of a matrix with respect to a vector*, which can be found, e.g., in [12], and will be of use below.

We start with an existence and uniqueness theorem for $P(\lambda)$.

Theorem 4.1. *$P(\lambda)$ exists and is unique. Furthermore, its degree m satisfies $q \leqslant m \leqslant q + a$, where q is the degree of the minimal polynomial of A with respect to $\hat{u}$, and hence $q \leqslant \dim \hat{\mathscr{S}} \leqslant N - a$. Actually, $P(\lambda)$ is the only polynomial in Π_{q+a}^0 that satisfies $P(A)\hat{u} = 0$.*

Proof. Let $P_1(\lambda)$ be the minimal polynomial of A with respect to $\hat{u}$, and let q be its degree. We know that $P_1(\lambda)$ is unique when normalized to be monic. Furthermore, since $\hat{u} \in \hat{\mathscr{S}}$, $P_1(0) \neq 0$. Let us now set $b_0 = 1/P_1(0)$ and define $b_1, b_2, \ldots, b_a$, recursively by $b_i = -\left[\sum_{s=0}^{i-1} \binom{i}{s} P_1^{(i-s)}(0) b_s\right] \Big/ P_1(0)$, $i = 1, 2, \ldots, a$. As $P_1(0) \neq 0$, all the b_i are well-defined and uniquely fixed by $P_1(\lambda)$. Consider now all polynomials $Q(\lambda)$ that satisfy $Q^{(i)}(0) = b_i$, $i = 0, 1, \ldots, a$. We know that there is at

least one such polynomial, namely, $Q(\lambda) = \sum_{i=0}^{a}(b_i/i!)\lambda^i \equiv P_2(\lambda)$. Consider also the polynomials $f(\lambda) = P_1(\lambda)Q(\lambda)$. As can easily be shown, each such $f(\lambda)$ satisfies $f(0) = 1$ and $f^{(i)}(0) = 0,\ i = 1, \ldots, a$, and hence is in Π_d^0 for some integer $d \geqslant q$, and $f(A)\hat{u} = 0$ since $P_1(A)\hat{u} = 0$. In particular, $P_1(\lambda)P_2(\lambda)$ is one such $f(\lambda)$, and we claim that $P(\lambda) = P_1(\lambda)P_2(\lambda)$.

Obviously, the degree m of $P(\lambda)$ is at least q and at most $q + a$, and thus $P \in \Pi_{q+a}^0$.

We next show that $P(\lambda)$ is the only polynomial in Π_{q+a}^0 that satisfies $P(A)\hat{u} = 0$. For this assume to the contrary that $\bar{P}(\lambda)$ is another polynomial in Π_{q+a}^0 that satisfies $\bar{P}(A)\hat{u} = 0$. We have $P(\lambda) = 1 - \lambda^{a+1}R(\lambda)$ and $\bar{P}(\lambda) = 1 - \lambda^{a+1}\bar{R}(\lambda)$ with $R, \bar{R} \in \Pi_{q-1}$. Since $P(A)\hat{u} = 0$ and $\bar{P}(A)\hat{u} = 0$, it follows that $A^{a+1}[R(A) - \bar{R}(A)]\hat{u} = 0$, which, by $\hat{u} \in \hat{\mathscr{S}}$, implies that $[R(A) - \bar{R}(A)]\hat{u} = 0$. Thus, the polynomial $R_1(\lambda) = R(\lambda) - \bar{R}(\lambda)$ is in Π_{q-1} and satisfies $R_1(A)\hat{u} = 0$. This is impossible as the degree of any polynomial $U(\lambda)$ that satisfies $U(A)\hat{u} = 0$ must be at least q. Therefore, $P(\lambda)$ is unique in Π_{q+a}^0. Obviously, this also implies that $P(\lambda)$ is unique. □

Since $P(\lambda) = 1 - \sum_{i=1}^{q} c_i\lambda^{a+i}$ and $P(A)\hat{u} = 0$, we see that the scalars c_i satisfy the $N \times q$ (overdetermined) system $U^{(q)}c^{(q)} = \hat{u}$, where we have defined the matrices $U^{(j)}$ and the vectors $c^{(j)}$ by $U^{(j)} = [A^{a+1}\hat{u}|A^{a+2}\hat{u}|\cdots|A^{a+j}\hat{u}]$ and $c^{(j)} = [c_1, \ldots, c_j]^{\mathrm{T}}$, respectively. Since, by the previous theorem, $P(\lambda)$ exists and is unique, we have that the overdetermined system $U^{(q)}c^{(q)} = \hat{u}$ is consistent and has a unique solution for the c_i. If $c_q \neq 0$, then the degree of $P(\lambda)$ is $q + a$. If $c_{m-a} \neq 0$ and $c_i = 0,\ i = m - a + 1, \ldots, q$, then the degree of $P(\lambda)$ is m. In this case the system $U^{(m)}c^{(m)} = \hat{u}$ is consistent and has a unique solution for the c_i.

The next theorem states that the Drazin-inverse solution $A^{\mathrm{D}}b$ can be obtained in terms of $\hat{x}_0$ and a (finite) linear combination of the vectors $A^a r_0, A^{a+1}r_0, \ldots, A^N r_0$.

Theorem 4.2. *Let $P(\lambda)$ be the minimal a-incomplete polynomial of A with respect to $\hat{e}_0 = \hat{x}_0 - A^{\mathrm{D}}b$, and let m be its degree. Then $P(\lambda) = 1 - \sum_{i=1}^{m-a}\hat{c}_i\lambda^{a+i}$ for some unique $\hat{c}_i$, and $A^{\mathrm{D}}b + \tilde{x}_0 = x_0 + \sum_{i=1}^{m-a}\hat{c}_i A^{a+i-1}r_0$.*

Proof. Let $p_m(\lambda) = P(\lambda)$. Then from (3.1) we have that $x_m = A^{\mathrm{D}}b + \tilde{x}_0$. But with this $p_m(\lambda)$, we have $q_{m-1}(\lambda) = (1 - p_m(\lambda))/\lambda = \sum_{i=1}^{m-a}\hat{c}_i\lambda^{a+i-1}$ in (2.1). The result now follows. □

Theorem 4.3. *Let m be the degree of $P(\lambda)$, the minimal a-incomplete polynomial of A with respect to $\hat{e}_0 = \hat{x}_0 - A^{\mathrm{D}}b$. In addition, let x_m be the vector generated by the projection method described through* (3.4)–(3.8). *Then, provided* $\det(W^*A^{a+1}V) \neq 0$, *we have $x_m = A^{\mathrm{D}}b + \tilde{x}_0$.*

Proof. The projection equations that define x_m are $W^*A^a r_m = 0$, which, by the fact that $A^a r_m = A^a\hat{r}_m$ and by (3.3) are equivalent to $W^*A^{a+1}p_m(A)\hat{e}_0 = 0$. Now, by the

discussion following the proof of Theorem 4.1, the linear system $f(A)\hat{e}_0 = 0, \quad f \in \Pi_m^0$, is overdetermined but consistent and has a unique solution, namely, $f(\lambda) = P(\lambda)$. By the assumption that $\det(W^* A^{a+1} V) \neq 0$, the linear system $W^* A^{a+1} p_m(A) \hat{e}_0 = 0$ has a unique solution for $p_m(\lambda)$ that is simply $P(\lambda)$. Invoking now Theorem 4.2, this implies that $x_m = A^{\mathrm{D}} b + \tilde{x}_0$. $\square$

5. Examples of projection methods of Krylov subspace type

We can now use the general framework developed in Section 3 to propose some concrete projection methods for computing $A^{\mathrm{D}} b$, in which the left subspaces W, just as the right subspaces V, are Krylov subspaces.

Recall that V stands both for the Krylov subspace $\mathscr{K}_{m-a}(A; A^a r_0)$ and for the $N \times (m-a)$ matrix in (3.5). Also, from (3.6), we have that $x_m \in x_0 + V$.

Finally, as the right subspace V is the same for all the methods considered here, namely, $V = \mathscr{K}_{m-a}(A; A^a r_0)$, what distinguishes one method from another is its corresponding left subspace W.

5.1. *Arnoldi type methods*

Let us choose $W = \mathscr{K}_{m-a}(A; A^\mu r_0)$, where μ is a nonnegative integer. With this choice of W we have $V = A^{a-\mu} W$ when $a \geqslant \mu$, where we have defined $B^\nu \mathscr{K}_m (B; u) = \mathscr{K}_m(B; B^\nu u)$.

When A is hermitian, hence $a = 1$, and we take $\mu = 1$, the projection equations in (3.7) become

$$(AV)^*(AV)c = (AV)^* r_0 \tag{5.1}$$

and these are actually the normal equations of the least squares problem

$$\min_{c_1,\ldots,c_{m-a}} \|r_0 - AVc\|, \tag{5.2}$$

which is the same as

$$\min_{x_m \in x_0 + \mathscr{K}_{m-1}(A; Ar_0)} \|r_m\|, \tag{5.3}$$

whether A is semidefinite or indefinite. Since $r_m = \hat{r}_m + \tilde{r}_0$ and $(\hat{r}_m, \tilde{r}_0) = 0$ for A hermitian, we have that $\|r_m\|^2 = \|\hat{r}_m\|^2 + \|\tilde{r}_0\|^2$, so that minimizing $\|r_m\|$ in (5.3) is the same as minimizing $\|\hat{r}_m\|$, which shows the validity of the approach above. In case A is hermitian positive semidefinite, and we take $\mu = 0$, the projection equations (3.7) become

$$V^* A V c = V^* r_0 \tag{5.4}$$

and these are the normal equations of the least squares problem

$$\min_{x_m \in x_0 + \mathscr{K}_{m-1}(A; Ar_0)} (x_m - A^{\mathrm{D}} b)^* A (x_m - A^{\mathrm{D}} b) \tag{5.5}$$

as can be shown after some lengthy manipulations. Since $y^*Ay = y^*A\hat{y} = (Ay)^*\hat{y} = (A\hat{y})^*\hat{y} = (\hat{A}\hat{y})^*\hat{y} = \hat{y}^*\hat{A}\hat{y}$, we see that we can replace $(x_m - A^{\mathrm{D}}b)^*A(x_m - A^{\mathrm{D}}b)$ in (5.5) by $(\hat{x}_m - A^{\mathrm{D}}b)^*\hat{A}(\hat{x}_m - A^{\mathrm{D}}b)$, which shows the validity of the approach above once again. Recall that $\hat{A}$ is the restriction of A to $\hat{\mathscr{S}}$, and it is hermitian positive definite when A is hermitian positive semidefinite.

The methods defined by (5.3) and (5.5) were proposed in [3], where CG type recursive algorithms for them are also given. We mention that $x_0 = 0$ in [3], which guarantees that $\tilde{x}_m = 0$ for all m.

5.2. *A GCR type method: DGCR*

Let us choose $W = A^{a+1}V$. Then the equations in (3.7) become

$$(A^{a+1}V)^*(A^{a+1}V)c = (A^{a+1}V)^*A^a r_0 \tag{5.6}$$

and these are the normal equations of the least squares problem

$$\min_{x_m \in x_0 + \mathscr{K}_{m-a}(A;A^a r_0)} \|A^a r_m\|. \tag{5.7}$$

We shall denote this method DGCR, where the letter D serves as a reminder that the Drazin-inverse solution is being computed.

Note that if we set $a = 0$ everywhere, (5.7) becomes

$$\min_{x_m \in x_0 + \mathscr{K}_m(A;r_0)} \|r_m\|,$$

which is how GCR for nonsingular systems is defined.

5.3. *Lanczos type methods*

Let us choose $W = \mathscr{K}_{m-a}(A^*; (A^*)^\mu \check{r}_0)$, where $\check{r}_0$ is an arbitrary vector in $\mathbb{C}^N$, and μ is a nonnegative integer. A Bi–CG type algorithm for the case $\mu = 1$ has recently been developed in [20].

Obviously, when A is hermitian and we set $\check{r}_0 = r_0$, the Lanczos and Arnoldi type methods become mathematically equivalent, since in this situation the left subspace W for the former is the same as that for the latter.

6. Error analysis for DGCR

As the vectors x_m produced by DGCR are also the solutions of the l_2 minimization problems of (5.7), they can conveniently be analyzed.

First, from (3.3) and the fact that $A^a\tilde{r}_m = 0$ we have that

$$A^{a+1}\hat{e}_m = A^{a+1}p_m(A)\hat{e}_0 = -A^a\hat{r}_m = -A^a r_m, \tag{6.1}$$

where, we recall, $\hat{e}_m = \hat{x}_m - A^{\mathrm{D}}b,\ m = 0, 1, \ldots$ Next, since $Av = \hat{A}v$ for every $v \in \hat{\mathscr{S}}$, where $\hat{A}$ is the restriction of A to $\hat{\mathscr{S}}$, we have $A^a r_m = -\hat{A}^{a+1}\hat{e}_m$ and $A^a r_m =$

$\hat{A}^a\hat{r}_m$ from (6.1). By the fact that $\hat{A}$ is nonsingular, $\|A^a r_m\|$ is a true norm for both $\|\hat{e}_m\|$ and $\|\hat{r}_m\|$. Actually, $\|\hat{e}_m\|$ and $\|\hat{r}_m\|$ are related to $\|A^a r_m\|$, respectively, by

$$\sigma_d^{(a+1)}\|\hat{e}_m\| \leqslant \|A^a r_m\| \leqslant \sigma_1^{(a+1)}\|\hat{e}_m\| \tag{6.2}$$

and

$$\sigma_d^{(a)}\|\hat{r}_m\| \leqslant \|A^a r_m\| \leqslant \sigma_1^{(a)}\|\hat{r}_m\|, \tag{6.3}$$

where, we recall, $\sigma_1^{(k)} > 0$ and $\sigma_d^{(k)} > 0$ are the largest and the smallest singular values of $\hat{A}^k$, $k = 1, 2, \ldots$ In view of this discussion, we conclude that it suffices to study the behavior of $\|A^a r_m\|$ for increasing m, to which we now turn.

Since x_m and hence $p_m(\lambda)$ in DGCR are optimal in the sense

$$\|A^a r_m\| = \|A^a p_m(A)\hat{r}_0\| = \min_{p\in\Pi_m^0} \|A^a p(A)\hat{r}_0\|, \tag{6.4}$$

we have

$$\|A^a r_m\| \leqslant \|p(A)(A^a\hat{r}_0)\| \quad \text{for any } p \in \Pi_m^0. \tag{6.5}$$

Since $A^a\hat{r}_0 \in \hat{\mathscr{S}}$, $A^a\hat{r}_0$ is a linear combination of the eigenvectors and principal vectors corresponding to the nonzero eigenvalues λ_j of A. Therefore,

$$p(A)(A^a\hat{r}_0) = \sum_{\lambda_j\in\sigma(A)\setminus\{0\}} \sum_{i=0}^{k_j-1} u_{ji}\, p^{(i)}(\lambda_j) \tag{6.6}$$

for some vectors u_{ji} that lie in the invariant subspace of A corresponding to λ_j. Here $\sigma(A)$ denotes the spectrum of A, and $k_j = \operatorname{ind}(A - \lambda_j I)$ as before. Thus, for any $p \in \Pi_m^0$,

$$\|p(A)A^a\hat{r}_0\| \leqslant K_1 \left(\max_{\lambda_j\in\sigma(A)\setminus\{0\}} \max_{0\leqslant i\leqslant k_j-1} |p^{(i)}(\lambda_j)| \right), \tag{6.7}$$

with some positive constant K_1 that depends only on $A^a\hat{r}_0$. If Ω is a closed domain in the complex plane containing only the nonzero eigenvalues of A, then we can replace (6.7) by

$$\|p(A)A^a\hat{r}_0\| \leqslant K_1 \left(\max_{0\leqslant i\leqslant \hat{k}-1} \|p^{(i)}\|_\Omega \right), \tag{6.8}$$

where

$$\|f\|_\Omega = \max_{\lambda\in\Omega} |f(\lambda)| \tag{6.9}$$

and

$$\hat{k} = \max\{k_j : \lambda_j \in \sigma(A)\setminus\{0\}\}. \tag{6.10}$$

By a theorem of Pommerenke [15], we know that if $Q(z)$ is an arbitrary polynomial in Π_m and E is a connected compact set of the z-plane, then there holds

$$\max_{z\in E}|Q'(z)| \leqslant \frac{e}{2}\frac{m^2}{\operatorname{cap}(E)}\max_{z\in E}|Q(z)|,$$

where $\operatorname{cap}(E)$ denotes the capacity of E. Now since Ω is a connected compact set of the complex λ-plane, Pommerenke's theorem applies and (6.8) becomes

$$\|p(A)A^a\hat{r}_0\| \leqslant K_2\, m^{2(\hat{k}-1)}\|p\|_\Omega \tag{6.11}$$

for some positive constant K_2 independent of m.

Substituting (6.11) in (6.5), and keeping in mind that $p \in \Pi_m^0$ is arbitrary, we can now write

$$\|A^a r_m\| \leqslant K_2\, m^{2(\hat{k}-1)}\left(\min_{p\in\Pi_m^0}\|p\|_\Omega\right). \tag{6.12}$$

The problem $\min_{p\in\Pi_m^0}\|p\|_\Omega$ has a unique solution $p^*(z)$ as has been shown by Rivlin and Shapiro [16]. Under the assumption that the boundary of Ω is sufficiently smooth, Eiermann and Starke [7] have shown that $\lim_{m\to\infty}\left(m^{-a}|\Phi(0)|^m\|p^*\|_\Omega\right)$ exists. (The case $a = 0$ of this result was proved earlier by Gutknecht [10].) Here $\Phi(\lambda)$ is the conformal mapping of the exterior of Ω to the exterior of the unit disk, namely, to the set $\{w\colon |w| \geqslant 1\}$. As the boundary of Ω is mapped onto $|w| = 1$, if the point $\lambda = 0$ is in the exterior of Ω, we have $|\Phi(0)| = |w_0| > 1$ and, therefore, $\|p^*\|_\Omega \leqslant Lm^a|\Phi(0)|^{-m}$ for all m where L is some positive constant that depends only on Ω, and also $\lim_{m\to\infty}\|p_m^*\|_\Omega = 0$. Using this fact in (6.12), we finally obtain the following result.

Theorem 6.1. *Choose Ω to be a closed domain that contains $\sigma(A)\backslash\{0\}$ but not $\lambda = 0$, such that its boundary is twice differentiable with respect to arclength. Denote by $\Phi(\lambda)$ the conformal mapping of the exterior of Ω onto the exterior of the unit disk $\{w : |w| \geqslant 1\}$. Then the vector x_m generated by DGCR satisfies*

$$\|A^a r_m\| \leqslant K\, m^{a+2(\hat{k}-1)}\rho^m \quad \textit{for all } m, \tag{6.13}$$

where K is a positive constant independent of m, $\hat{k} = \max\{k_j\colon k_j = \operatorname{ind}(A-\lambda_j I)$, $\lambda_j \in \sigma(A)\backslash\{0\}\}$, and $\rho = 1/|\Phi(0)| < 1$.

What Theorem 6.1 implies is that $\lim_{m\to\infty}\|A^a r_m\| = 0$. We should, of course, bear in mind that $A^a r_m = 0$ for some $m \leqslant N$, as proved already in Theorem 4.3. Therefore, the result of Theorem 6.1 should be understood in the sense that we have an upper bound on $\|A^a r_m\|$ that decreases to zero monotonically for $m \geqslant (a + 2(\hat{k}-1))/|\log\rho|$ essentially exponentially in m.

Theorem 6.1 is also valid for GCR on nonsingular systems when we set $a = 0$ everywhere.

Needless to say, the result in (6.13) would be sharper when Ω is chosen to be as small as possible as this reduces the value of ρ.

Using a standard technique employed in the analysis of CG, it can be shown that if x_m is the solution to (5.3) with A hermitian and possibly indefinite, then

$$\|\hat{r}_m\|/\|\hat{r}_0\| \leqslant \max_{\lambda_j \in \sigma(A)\setminus\{0\}} |p(\lambda_j)| \quad \text{for any } p \in \Pi_m^0 \tag{6.14}$$

and if x_m is the solution to (5.5) with A positive semidefinite, then

$$\|\hat{e}_m\|_A/\|\hat{e}_0\|_A \leqslant \max_{\lambda_j \in \sigma(A)\setminus\{0\}} |p(\lambda_j)| \quad \text{for any } p \in \Pi_m^0, \tag{6.15}$$

where we have defined $\|y\|_A = \sqrt{y^*Ay}$. These results have been obtained in [3]. In (6.14) we should keep in mind that $\sigma(A)\setminus\{0\}$ may be contained in the union $\Omega = [-\gamma, -\delta] \cup [\alpha, \beta]$, with $\alpha, \beta, \gamma, \delta > 0$. In (6.15), however, it is contained in $\Omega = [\alpha, \beta]$ with $\alpha, \beta > 0$, and in this case $\rho = 1/|\Phi(0)| = (\sqrt{\beta} - \sqrt{\alpha})/(\sqrt{\beta} + \sqrt{\alpha})$, which is a familiar quantity in the literature of CG and Chebyshev acceleration.

Concerning $\|p^*\|_\Omega$, in [7] it is mentioned that under more general conditions on Ω there holds

$$\alpha_1 \leqslant m^{-a}|\Phi(0)|^m \|p^*\|_\Omega \leqslant \alpha_2$$

with α_1 and α_2 being positive constants independent of m, as follows from the results of Saff [17].

7. Conclusions

In this work we have developed a unified framework for the construction of Krylov subspace methods for the Drazin-inverse solution of singular linear systems $Ax = b$, $A \in \mathbb{C}^{N\times N}$. The matrices A are assumed to be nonhermitian in general and $a = \text{ind}(A)$ can be arbitrary. In these methods the approximations to $A^{\mathrm{D}}b$, the Drazin-inverse solution of $Ax = b$, are all of the form $x_m = x_0 + \sum_{i=1}^{m-a} c_i A^{a+i-1} r_0$, where x_0 is the initial vector and $r_0 = b - Ax_0$ and c_i are scalars. The c_i are determined by requiring that $A^a r_m$, where $r_m = b - Ax_m$, be orthogonal to an $(m - a)$-dimensional subspace W. After showing the relevance and theoretical validity of this approach, we have proved that, subject to certain regularity assumptions, the methods developed terminate in a finite number of steps. That is, for some finite $m \leqslant N$, there holds $x_m = A^{\mathrm{D}}b + \tilde{x}_0$, where $\tilde{x}_0$ is that part of x_0 that lies in $\mathcal{N}(A^a)$. We have proposed new methods analogous to the Arnoldi, GCR, and Lanczos methods. For one of the methods denoted DGCR we have also provided a rigorous error analysis. A pleasant feature of our approach is that, when ind(A) is set equal to zero, it reduces to the projection approach for nonsingular systems. In this sense, our approach to singular systems is a true generalization of the projection approach relevant to nonsingular systems.

References

[1] W.E. Arnoldi, The principle of minimized iterations in the solution of the matrix eigenvalue problem, Quart. Appl. Math. 9 (1951) 17–29.

[2] A. Ben-Israel, T.N.E. Greville, Generalized Inverses: Theory and Applications, Wiley, New York, 1974.
[3] D. Calvetti, L. Reichel, Q. Zhang, Conjugate gradient algorithms for symmetric inconsistent linear systems, in: J.D. Brown, M.T. Chu, D.C. Ellison, R.J. Plemmons (Eds.), Proceedings of the Cornelius Lanczos International Centenary Conference, SIAM, 1993, pp. 267–272.
[4] S.L. Campbell, C.D. Meyer Jr., Generalized Inverses of Linear Transformations, Pitman, London, 1979.
[5] J.J. Climent, M. Neumann, A. Sidi, A semi-iterative method for singular linear systems with arbitrary index, J. Comp. Appl. Math. 87 (1997) 21–38.
[6] M. Eiermann, I. Marek, W. Niethammer, On the solution of singular linear systems of algebraic equations by semiiterative methods, Numer. Math 53 (1988) 265–283.
[7] M. Eiermann, G. Starke, The near-best solution of a polynomial minimization problem by the Carathéodory-Fejér method, Constr. Approx. 6 (1990) 303–319.
[8] S.C. Eisenstat, H.C. Elman, M.H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, SIAM J. Numer. Anal. 20 (1983) 345–357.
[9] B. Fischer, M. Hanke, M. Hochbruck, A note on conjugate gradient type methods for indefinite and/or inconsistent linear systems, Numer. Algorithms 11 (1996) 181–187.
[10] M.H. Gutknecht, An iterative method for solving linear equations based on minimum norm Pick-Nevanlinna interpolation, in: C.K. Chui, L.L. Schumaker, J.D. Ward (Eds.), Approximation Theory V, Academic Press, New York, 1986, pp. 371–374.
[11] M. Hanke, M. Hochbruck, A Chebyshev-like semiiteration for inconsistent linear systems, Electr. Trans. Numer. Anal. 1 (1993) 89–103.
[12] A.S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964.
[13] E.F. Kaaschieter, Preconditioned conjugate gradients for solving singular systems, J. Comp. Appl. Math. 24 (1988) 265–275.
[14] C. Lanczos, Solution of systems of linear equations by minimized iterations, J. Res. N.B.S. 49 (1952) 33–53.
[15] Ch. Pommerenke, On the derivative of a polynomial, Michigan Math. J. 6 (1959) 373–375.
[16] T.J. Rivlin, H.S. Shapiro, A unified approach to certain problems of approximation and minimization, J. Soc. Indust. Appl. Math. 9 (1961) 670–699.
[17] E.B. Saff, Polynomials of interpolation and approximation to meromorphic functions, Trans. Amer. Math. Soc. 143 (1969) 509–522.
[18] A. Sidi, Application of vector extrapolation methods to consistent singular linear systems, Appl. Numer. Math. 6 (1989–1990) 487–500.
[19] A. Sidi, Development of iterative techniques and extrapolation methods for Drazin inverse solution of consistent or inconsistent singular linear system, Linear Algebra Appl. 167 (1992) 171–203.
[20] A. Sidi, V. Kluzner, A Bi–CG type iterative method for Drazin-inverse solution of singular nonsymmetric linear systems of arbitrary index (submitted).