Theoretical Error Bounds and General Analysis of a few Lanczos-Type Algorithms *

Yousef Saad†

Abstract

In this paper we give an overview of the results used to analyze some of the algorithms developed by Lanczos. These include the minimized iteration algorithm (conjugate gradient) for linear systems and the Lanczos algorithm for eigenvalue problems. We also formalize some simple results showing superlinear convergence in a simple framework.

1 Introduction: Projection Methods

In the early 1950's Cornelius Lanczos introduced a class of methods for solving the two major problems in numerical linear algebra, namely linear systems and eigenvalue problems. In both cases, the methods were derived from the an 'optimality' point of view as well as the projection point of view, by imposing orthogonality conditions. In fact the methods introduced by Lanczos are projection techniques onto Krylov subspaces.

There are two types of general projection methods: orthogonal and oblique. An *orthogonal projection method* is a projection method onto a given subspace K and orthogonally to itself. More precisely, for a linear system

$$Au = b \tag{1}$$

the approximate problem is defined as the problem of finding an approximate solution $\tilde{u}$ to the original system such that

$$\tilde{u} \in u_0 + K; \quad b - A\tilde{u} \perp K$$

For an eigenvalue problem of the form

$$Au = \lambda u \tag{2}$$

*This work was supported by NSF under grants number DMR-9217287 and CCR-9214116 and by the Minnesota Supercomputer Institute

†Unversity of Minnesota, Department of Computer Science, and Minnesota Supercomputer Institute, Minneapolis, Minnesota

the approximate problem consists of finding an approximate pair $\tilde{\lambda}, \tilde{u}$ which satisfies the conditions,

$$\tilde{\lambda} \in \mathbb{C}, \tilde{u} \in K; \quad (\tilde{\lambda}I - A)\tilde{u} \perp K \tag{3}$$

To define an oblique projection method we need a pair of subspaces K and L of the same dimension. An oblique projection method is a projection method onto K orthogonally to L. For linear systems, the approximate problem becomes,

$$\tilde{u} \in u_0 + K; \quad b - A\tilde{u} \perp L$$

and, similarly, for eigenvalue problems the formulation (3) becomes

$$\tilde{\lambda} \in \mathbb{C}, \tilde{u} \in K; \quad (\tilde{\lambda}I - A)\tilde{u} \perp L \tag{4}$$

The basic optimality properties of these methods are as follows.

Linear systems

- When the subspaces L and K are identical, i.e., for orthogonal projection methods, and when the matrix A is symmetric and positive definite, then the approximate solution $\tilde{u}$ minimizes the $A-$norm of the error over the affine subspace $u_0 + K$,

$$\|u^* - \tilde{u}\|_A = \min_{u \in u_0+K} \|u^* - u\|_A$$

- When the subspaces L and K are related by $L = AK$, then the approximate solution $\tilde{u}$ minimizes the $2-$norm of the residual vector over the affine subspace $u_0 + K$,

$$\|b - A\tilde{u}\|_2 = \min_{u \in u_0+K} \|b - Au\|_A$$

  This is true provided that A is nonsigular.

The result of the second case has historically provided a natural mechanism for attempting to define methods that will convergence under mild conditions.

Eigenvalue problems

- When the subspaces L and K are identical, i.e., for orthogonal projection methods, and when the matrix A is Hermitian then the set of approximate eigenpairs $\tilde{\lambda}_i, \tilde{u}_i$ satisfy a number of optimality properties that result from the min-max theorem of the Courant characterization [21, 26]. In particuler, if we label all the eigenvalues increasingly then

$$\lambda_1 = \min_K \frac{(Au, u)}{(u, u)} = \frac{(A\tilde{u}_1, \tilde{u}_1)}{(\tilde{u}_1, \tilde{u}_1)} \tag{5}$$

$$\lambda_j = \min_{uK,\ u \perp \tilde{u}_1, \ldots, \tilde{u}_{j-1}} \frac{(Au, u)}{(u, u)} = \frac{(A\tilde{u}_j, \tilde{u}_j)}{(\tilde{u}_j, \tilde{u}_j)} \tag{6}$$

$$\tag{7}$$

  These results allow to establish error bounds for the Lanczos algorithm, and the subspace iteration algorithm, in the Hermitian case.

- In the case when the subspaces L and K are related by $L = AK$, then there are no optimality results similar to those of the orthogonal projection case. In fact this option does not seem to be common in practice. The corresponding eigenvalue problem occurs naturally in the context of Krylov supspace methods for solving linear systems by projection methods of the same type, i.e., methods imposing $L = AK$. These methods include GMRES, ORTHOMIN, GCR, for example. In this situation, the approximate eigenvalues $\tilde{\lambda}_i$ are the roots of the residual polynomial obtained by these methods.

2 Krylov subspace methods

Krylov subspace methods are projection methods onto subspaces of the form

$$K_m(A, v) = \text{span}\{v, Av, \ldots, A^{m-1}v\}$$

In addition to the well-known methods for solving linear systems and eigenvalue problems, there are many techniques in verious areas of scientific computing that exploit Krylov subspaces in different forms. These include nonlinear equations [5, 4], Systems of Ordinary Differential Equations [3, 17], Solution of Partial Differential Equations [9, 8], Matrix equations in control [24, 25].

We now give a short survey of some of the recent work on Krylov subspace methods for linear systems as well as eigenvalue problems.

Orthogonal projection methods

- Conjugate Gradient methods due to Hestenes and Stiefel [11] and to Lanczos [16]. This method is designed for symmetric positive deinite matrices.

- Lanczos Algorithm and variants [16], [20], [19] In these methods A need only be symmetric. Another class of methods based on using the Lanczos algorithm for diagonally shifted skew hermitian matrices was developed by [6], and Widlund [29].

- Full Orthogonalization Method developed in [23] and in different forms by [1], and [12]. require only that A be nonsingular. However, there are no known results on convergence in the general case.

- The Lanczos algorithm for Hermitian matrices; [15, 21]. (Orthogonal projection). Also of interest the preconditioned variants (Davidson).

Oblique projection methods

- The nonsymmetric Lanczos algorithm and its variants, Lanczos minimized iteration (Bi-CG) 1952. Flectcher (Bi CG) '75. Saad '81, Freund and Nachtigal (QMR) '91, Freund (TFQMR) '92.... **A nonsingular**

- Minimal residual methods Vinsome (ORTHOMIN) '76, Young & Jea (ORTHODIR) '80, Saad & Schultz (GMRES) '85, **A nonsingular** Some convergence results for A positive definite.
- The Lanczos algorithm for Non-Hermitian matrices; (Oblique projection);
- Subspace Iteration [Bauer 59]
- Block Lanczos [Underwood & Golub '78]

3 Symmetric matrices

Consider the linear system,

$$Ax = b \tag{8}$$

where A is a real symmetric positive definite. Given an initial guess x_0 and the corresponding residual vector,

$$r_0 = b - Ax_0 \tag{9}$$

the conjugate gradient method [11, 16], consists of finding an approximation of the form

$$x = x_0 + p_{m-1}(A)r_0 \tag{10}$$

where p_{m-1} is a polynomial of degree $m-1$. The Galerkin condition

$$b - Ax \perp K_m$$

yields the conjugate gradient approximation. We refer the reader to a standard reference for a description on the most common variant of the algorithm, see for example [10, 2].

A well-known optimality property which is satisfied by the iterates of the conjugate gradient method is that x_m minimizes the A-norm $\|x - x_*\|_A$ over $x_0 + K_m$. An immediate consequence is the standard result that if we denote by x_m the m-th CG iterate, and by x_* the exact solution and

$$\eta = \frac{\lambda_{min}}{\lambda_{max} - \lambda_{min}}$$

Then,

$$\|x_* - x_m\|_A \leq \frac{\|x_* - x_0\|_A}{T_m(1 + 2\eta)}$$

where T_m is the Chebyshev polynomial of the first kind of degree m.

Notice that the approximate solution is of the form

$$x_m = x_0 + q_m(A)r_0$$

where q_m is a polynomial of degree m such that

$$\|(I - Aq_m(A))e_m\|_A = \min_{q \in \mathbf{P}_{m-1}} \|(I - Aq(A))e_0\|_A$$

The eigen-component associated with λ_i is multiplied by $1 - \lambda_i q(\lambda_i)$. We refer to this as the amplification factor for the eigenvalue λ_i. In order to obtain theoretical error bounds for the error in the conjugate gradient iterates, the usual technique is to find a polynomial for which the maximum amplification coefficient is the smallest possible. This is achievable by using the optimality of the Chebyshev polynomials. If $[\alpha, \beta]$ is a non-empty interval in $\mathbb{R}$ and γ is any real scalar such with $\gamma \geq \beta$. Then the minimum

$$\min_{p \in \mathbf{P}_k, p(\gamma)=1} \max_{t \in [\alpha,\beta]} |p(t)|$$

is reached by the polynomial

$$\hat{T}_k(t) \equiv \frac{T_k\left(1 + 2\frac{t-\beta}{\beta-\alpha}\right)}{T_k\left(1 + 2\frac{\gamma-\beta}{\beta-\alpha}\right)} \ .$$

For Symmetric eigenvalue problems a similar analysis exists leading to a similar problem for the largest or smallest eigenvalue. Thus, the optimality property for the smallest eigennavlue $\tilde{\lambda}_1$, is that $\tilde{\lambda}_1$ minimizes the Rayleigh quotient

$$(Av, v)/(v, v)$$

over $K_m(A, v)$. From this, one can derive bounds similar to those of the Conjugate Gradient method.

Kaniel [13] proved the following inequality for the first eigenvalue,

$$0 \leq \lambda_1^{(m)} - \lambda_1 \leq (\lambda_N - \lambda_1) \left[\frac{\tan \angle(v_1, u_1)}{T_{m-1}(1 + 2\gamma_1)}\right]^2$$

in which

$$\gamma_1 = \frac{\lambda_2 - \lambda_1}{\lambda_N - \lambda_2}$$

Kaniel and later Paige, and also proved a number of results for the other eigenvalues [13, 18, 21]. Exploiting the Courant characterization property, the following alternative generalization of the above result was shown in [22],

$$0 \leq \lambda_i^{(m)} - \lambda_i \leq (\lambda_N - \lambda_1) \left[\kappa_i^{(m)} \frac{\tan \angle(v_i, u_i)}{T_{m-i}(1 + 2\gamma_i)}\right]^2$$

in which

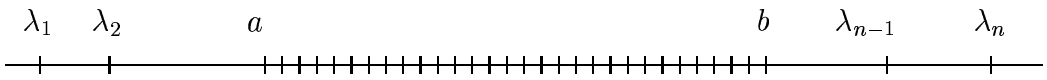
$$\gamma_i = \frac{\lambda_{i+1} - \lambda_i}{\lambda_N - \lambda_{i+1}} \quad \kappa_i^{(m)} = \prod_{j<i} \frac{\lambda_j^{(m)} - \lambda_N}{\lambda_j^{(m)} - \lambda_i}$$

A number of results were also established for eigenvectors [22, 26].

3.1 Improved bounds

One question we may now ask is whether or not the theoretical bounds discussed in the previous section are the best which can be derived. There are cases where the above bounds are very loose, namely when there is a clustering of the eigenvalues.

We consider only the situation of linear systems but eigenvalues problems can be handled in a similar manner. The simplest improved bounds exploit isolated eigenvalues as, for example shown in the following figure,



For this situation, we can resort to polynomials of the form:

$$r_m(\lambda) = \prod_{i=1}^{s} \frac{\lambda_i - \lambda}{\lambda_i} \quad r_{m-s}(\lambda)$$

in which $r_{m-s}(\lambda)$ is any consistent polynomial of degree $m-s$, which is small in interval $[a,b]$. This technique was often exploited in the literature on theoretical error bounds for Krylov methods, see for example, [13, 18, 22, 28, 27]. It yields good results for a small number of isolated eigenvalues but it is not adequate for modeling large 'clusters' of eigenvalues. We would like to explore the above idea a little further and consider specifically *clustered spectra*. 1¡There are many situations where the spectrum consists of one or a few 'clusters' or 'accumulation' points, i.e., locations on the real interval with a high density of eigenvalues. In fact for linear systems this is the norm rather than the exception since the system that is solved is typically preconditioned and the purpose of preconditioning is precisely to cluster the eigenvalues around the vakue one. Certain types of eigenvalue problems also exhibit clustering. In fact preconditioning eigenvalue problems, e.g., by preconditioning has the effect of clustering large eigenvalues of the unpreconditioned system at zero for the preconditioned system.

To model clustering, we will assume, without loss of generality, that the spectrum is clustered around one. We consider the sets:

$$S(r) = \{\lambda_j \ s.t. |\lambda_j - 1| \geq r\}$$

Observe that $|S(r)|$ is the number of eigenvalues outside the interval centered at one and of half-width r, which we denote by $D(1,r)$. In fact, for what follows it convenient and indeed, crucial, to assume that A is a bounded operator on a Hilbert space. Then, $|S(r)|$ will decrease from $|S(0)| = \infty$ to $|S(\infty)| = 0$ as r increases from 0 to infinity.

For any given positive integer μ, define

$$r(\mu) = \min\{r \mid |S(r)| \leq \mu\} \tag{11}$$

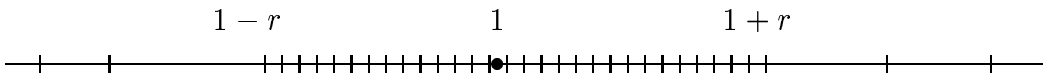
Note that $r(\mu)$ represents the radius of the smallest interval centered at one, which excludes exactly μ eigenvalues, and that we could have replaced inequality in (11). The

function $r(\mu)$ can also be defined for real nonnegative values. We have $r(0) = \rho \geq r(\mu) \geq r(\infty) = 0$.

The main assumption we make is that there exists two positive constants K and β such that

$$r(\mu) \leq \frac{1}{K\mu^{\beta}} \ . \tag{12}$$

with This assumption states that the radius of the interval centered at one, and which excludes μ eigenvalues does not exceed $(K\mu^{\beta})^{-1}$. For example, a spectrum consisting of eigenvalues of the form $\lambda_j = 1 + \frac{1}{\beta^j}$ will satisfy the condition.



Consider the set of polynomials,

$$r_{m,r}(\lambda) = \prod_{i=1}^{\mu} \frac{\lambda_i - \lambda}{\lambda_i} \quad \frac{T_{m-\mu}\left(\frac{1-\lambda}{r}\right)}{T_{m-\mu}\left(\frac{1}{r}\right)}$$

where μ is a parameter which depends on r and which will represent the number of eigenvalues outside the interval $(1-r, 1+r)$. We would like to estimate the maximum of $r_{m,r}$

$$\begin{aligned} \max_{\lambda \in \Sigma(A)} |r_{m,r}(\lambda)| &= \max_{j=\mu+1,N} \prod_{i=1}^{\mu} \left| \frac{\lambda_i - \lambda_j}{\lambda_i} \right| \quad \frac{1}{T_{m-\mu}\left(\frac{1}{r}\right)} \\ &\leq \left(\frac{2\rho}{1-\rho}\right)^{\mu} \frac{1}{T_{m-\mu}\left(\frac{1}{r}\right)} \end{aligned}$$

where $\rho = \max_{j=1,\ldots,N} |1 - \lambda_j|$

$$\max_{\lambda \in \Sigma(A)} |r_{m,r}(\lambda)| \leq \left(\frac{2\rho}{1-\rho}\right)^{\mu} \frac{1}{T_{m-\mu}\left(\frac{1}{r}\right)} \tag{13}$$

We can now exploit the additional degree of freedom given by μ, or equivalently, by r. The idea is to choose r so that μ increases proportionally to m. For example for a given α, with $0<\alpha<1$ we take

$$\mu = \lceil \alpha m \rceil \tag{14}$$

Then, from the relation (12) between r and μ, we have

$$r \leq \frac{1}{K\lceil \alpha m \rceil^{\beta}} \leq \frac{1}{K(\alpha m)^{\beta}}$$

$$\max_{\lambda \in \Sigma(A)} |r_{m,\alpha}(\lambda)| \leq \left(\frac{2\rho}{1-\rho}\right)^{\lceil \alpha m \rceil} \frac{1}{T_{m-\lceil \alpha m \rceil}\left(K(\alpha m)^{\beta}\right)} \tag{15}$$

The above relation, which is a variation of a result shown in [14], estabishes that the method will converge superlinearly in this situation. There are two terms to the right hand side. The first is a term growing with a linear rate m and the second is a term decreasing with a super linear rate in m. Because of the first term, the inequality may not be too sharp for small values of m. It does however show that, asymptotically, the rate of convergence is superlinear. The larger β the faster the asymptotic convergence is guaranteed to be. However, note that there are no specific assumptions regarding β. In fact we do not even need clustered spectra to obtain a similar result. Indeed, we can still exploit (13) for any spectra. Assume again that we exclude μ eigenvalues where μ is defined as in (14). Then, the inequality (13) becomes,

$$\max_{\lambda \in \Sigma(A)} |r_{m,r}(\lambda)| \leq \left(\frac{2\rho}{1-\rho} \right)^{\lceil \alpha m \rceil} \quad \frac{1}{T_{m-\lceil \alpha m \rceil} \left(\frac{1}{r(\alpha m)} \right)} \tag{16}$$

Since $r(t)$ is a function which decreases from ρ to 0 as μ goes from 0 to ∞, we can state that the superlinear convergence will take place asymptotically not only for clustered spectra, but for *most* distributions. However, rigorously speaking, this is useful only for infinite dimentional operators, since it is only in these cases that one can speak of asymptotic rates of convergence. This asymptotic regime of convergence may take too many steps to reach to be of interest for finite dimensional problems since many of the methods under consideration will deliver the exact solution in a finite number of steps. Superlinear convergence, can however be observed in the finite dimensional case. Several authors, including Van der Sluis and Van der Vorst [28], Concus, Golub, and O'Leary [7] have mentioned this fact and given alternative explanations for it. We also note that in contrast with previous work, the above argument does not exploit weights of the residual vector or the error vector with respect to the eigenvectors. In addition, it may be possible to derive a bound in which the constant α is adjusted to minimize the right-hand-side of the inequalities (16) or (15).

To illustrate superlinear convergence for the CG algorithm, we only show one simple example constructed by taking a diagonal matrix consisting of $\lambda_i = 1 - rand(i)^3, i = 1, \ldots, n$, where $rand(i)$ is a pseudo-random number between -1 and 1. The conjugate gradient algorithm is executed for a solvinv $Ax = b$, where b consists of a vector of all ones and the initial guess is another pseudo-random vector. We plot the numbers

$$\frac{-\log \|x_k - x_*\|_2}{k}$$

versus the step number k. These numbers represent the estimated convergence rates at each step. As can be seen the rates are not only far from being constants, but they also increase, albeit non-monotonically. The resulting plot is shown in Figure 1.

References

[1] O. Axelsson. Conjugate gradient type-methods for unsymmetric and inconsistent systems of linear equations. *Linear Algebra Appl.*, 29:1–16, 1980.

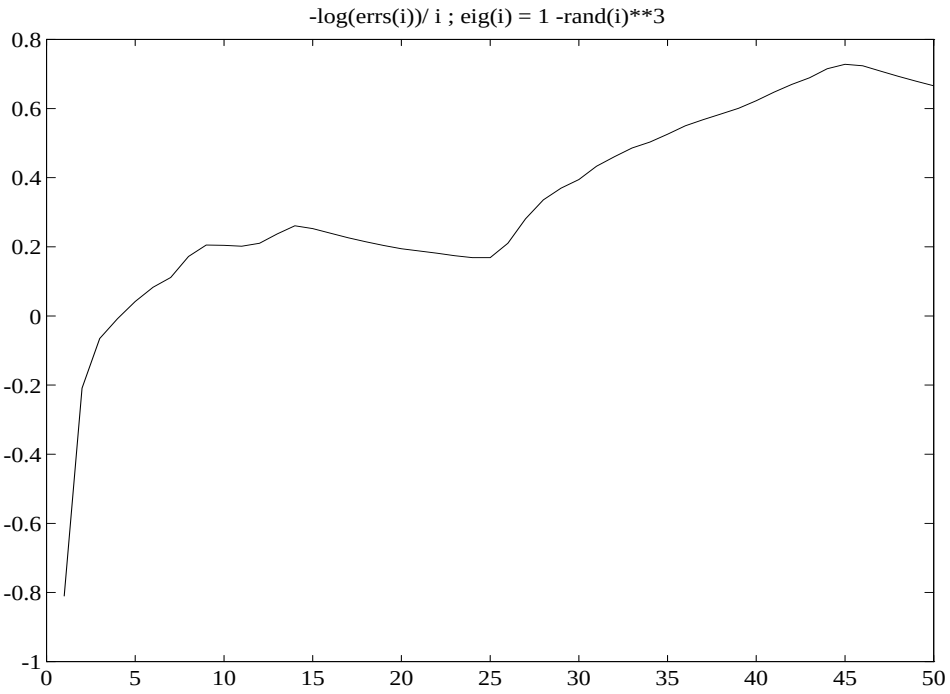


Figure 1: Convergence rate as a function of step for a clustered spectrum.

[2] O. Axelsson and V. A. Barker. *Finite Element Solution of Boundary Value Problems.* Academic Press, Orlando, Florida, 1984.

[3] P. N. Brown and A. C. Hindmarsh. Matrix-free methods for stiff systems of odes. *SIAM J. Num. Anal.*, 23:610–638, 1986.

[4] P. N. Brown and Y. Saad. Globally convergent techniques in nonlinear Newton-Krylov algorithms. Technical Report 89-57, Research Institute for Advanced Computer Science, 1989.

[5] P. N. Brown and Y. Saad. Hybrid Krylov methods for nonlinear systems of equations. *SIAM J. Sci. Stat. Comp.*, 11:450–481, 1990.

[6] P. Concus and G. H. Golub. A generalized conjugate gradient method for nonsymmetric systems of linear equations. In R. Glowinski and J. L. Lions, editors, *Computing Methods in Applied Sciences and Engineering*, pages 56–65, New York, 1976. Springer Verlag.

[7] P. Concus, G. H. Golub, and D. P. O'Leary. A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations. In James R. Bunch and Donald J. Rose, editors, *Sparse Matrix Computations*, pages 309–332, New York, 1976. Academic Press.

[8] R. A. Friesner, L. S. Tuckerman, B. C. Dornblaser, and T. V. Russo. A method for exponential propagation of large systems of stiff nonlinear differential equations. *Journal of Scientific Computing*, 4:327–354, 1989.

[9] E. Gallopoulos and Y. Saad. Efficient solution of parabolic equations by polynomial approximation methods. *SIAM J. Sci. Stat. Comput.*, 13:1236–1264, 1992.

[10] G. H. Golub and C. Van Loan. *Matrix Computations.* Academic Press, New York, 1981.

[11] M. R. Hestenes and E. L. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, Section B, 49:409–436, 1952.

[12] K. C. Jea and D. M. Young. Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods. *Linear Algebra Appl.*, 34:159–194, 1980.

[13] S. Kaniel. Estimates for some computational techniques in linear algebra. *Math. Comp.*, 20:369–378, 1966.

[14] T. Kerkhoven and Y. Saad. Acceleration techniques for decoupling algorithms in semiconductor simulation. Technical Report 684, University of Illinois, CSRD, Urbana, IL., 1987.

[15] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. of Standards*, 45:255–282, 1950.

[16] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Standards*, 49:33–53, 1952.

[17] B. Nour-Omid. Applications of the lanczos algorithm. *Comput. Phys. Comm*, 53, 1989.

[18] C. C. Paige. *The computation of eigenvalues and eigenvectors of very large sparse matrices.* PhD thesis, London University, Institute of Computer Science, London, England, 1971.

[19] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J Numer. Anal.*, 12:617–624, 1975.

[20] B. N. Parlett. A new look a the Lanczos algorithm for solving symmetric systems of linear equations. *Linear Algebra Appl.*, 29:323–346, 1980.

[21] B. N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice Hall, Englewood Cliffs, 1980.

[22] Y. Saad. On the rates of convergence of the Lanczos and the block Lanczos methods. *SIAM J. Numer. Anal.*, 17:687–706, 1980.

[23] Y. Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Mathematics of Computation*, 37:105–126, 1981.

[24] Y. Saad. Numerical solution of large Lyapunov equations. In M. A. Kaashoek, J. H. van Schuppen, and A. C. Ran, editors, *Signal Processing, Scattering, Operator Theory, and Numerical Methods. Proceedings of the international symposium MTNS-89, vol III*, pages 503–511, Boston, 1990. Birkhauser.

[25] Y. Saad. An overview of Krylov subspace methods with applications to control problems. In M. A. Kaashoek, J. H. van Schuppen, and A. C. Ran, editors, *Signal Processing, Scattering, Operator Theory, and Numerical Methods. Proceedings of the international symposium MTNS-89, vol III*, pages 401–410, Boston, 1990. Birkhauser.

[26] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* Halstead Press, New York, 1992.

[27] G. L. G. Sleijpen and A. van der Sluis. Further results on the convergence behaviour of cg and Ritz values. Technical Report 677, Utrecht University, Department of Mathematics, Utrecht, The Netherlands, 1991.

[28] A. van der Sluis and H. A. van der Vorst. The rate of convergence of conjugate gradients. *Numer. Math.*, 48:543–560, 1986.

[29] O. Widlund. A Lanczos method for a class of non-symmetric systems of linear equations. *SIAM Journal on Numerical Analysis*, 15:801–812, 1978.