BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 52, Number 1, January 2015, Pages 151–158
S 0273-0979(2014)01473-9
Article electronically published on September 10, 2014

Krylov subspace methods: principles and analysis, by Jörg Liesen and Zdeněk Strakoš, Numerical Methods and Scientific Computation, Oxford University Press, Oxford, 2013, xiv+391 pp., ISBN 978-0-19-965541-0

1. Modern methods for an ancient problem

Over the centuries tremendous energy has been devoted to the basic problem of solving a set of N linear algebraic equations, expressed as $Ax = b$, for the unknown vector x. High school students learn Gaussian elimination to, in principle, resolve all such equations when a unique solution exists, while the criteria for solvability is the centerpiece of first courses in linear algebra. Research interest in this problem derives from the size of the matrices needed in contemporary applications. The challenge is longstanding: in the late 1600s Wallis tackled a 25×25 matrix that arose from his design for a flat roof for Wren's Sheldonian Theatre at Oxford [30]. He could accomplish this feat because most entries in the matrix were zero (reflecting the connectivity of struts in his truss); such a matrix is *sparse*. The same is true for many matrices from current science and engineering problems, but now these systems routinely have millions or billions of unknowns. The nonzero pattern of A, though often quite intricate, enables fast matrix-vector product calculations that form the critical kernel of modern algorithms.

For a generic system of dimension n, Gaussian elimination requires $O(n^3)$ arithmetic operations to produce an exact solution, with no information about x provided until the final stage of computation. In contrast *iterative methods* incrementally refine an approximate solution, allowing a user to quit the process as soon as the solution attains the accuracy demanded by the motivating application (where $Ax = b$ is usually an approximation, such as a discretization of a partial differential equation); see, e.g., [27]. Classical "stationary" iterations, such as the Jacobi, Gauss–Seidel, and Successive Over-Relaxation (SOR) methods, were favorites when computers were humans who could tune parameters to accelerate convergence. These methods remain staples of many numerical analysis courses, but over the past forty years they have been displaced in practice by the more powerful methods that are the subject of the book under review.

Polynomial iterative methods are motivated by a corollary of the Cayley–Hamilton theorem: for any invertible A there exists a polynomial q of degree $n-1$ such that $q(A) = A^{-1}$. This polynomial interpolates $1/z$ at the eigenvalues of A; in many cases a polynomial of much lower degree can give accurate approximations to $1/z$ at these points, and hence to A^{-1}. Even better, one can tailor the approximation to the right-hand side b (which might have larger components in certain eigenvectors than others). Such approximations could be constructed manually, given estimates of the spectrum of A, but better algorithms (implicitly) build a sequence of improving polynomials automatically. The first modern method was the *conjugate gradient* algorithm proposed in 1952 by Hestenes and Stiefel for symmetric positive definite A [16]. Such matrices define the inner product $(x, y)_A = y^*Ax$ and norm $\|x\|_A = \sqrt{(x, x)_A}$. At its kth step the conjugate gradient method uses a three-term recurrence to build the iterate x_k that minimizes the A-norm of the error,

$$\|x - x_k\|_A = \min_{\deg(q)<k} \|x - q(A)b\|_A = \min_{\widehat{x}\in\mathcal{K}_k(A,b)} \|x - \widehat{x}\|_A, \tag{1.1}$$

over the kth *Krylov subspace*

$$\mathcal{K}_k(A,b) = \{q(A)b : \deg(q) < k\} = \operatorname{span}\{b, Ab, \ldots, A^{k-1}b\}.$$

The resulting algorithm is unusual, for like classical iterative methods it produces a sequence of improving approximations, but like Gaussian elimination it gives the exact solution at the nth step, since $x \in \mathcal{K}_n(A,b)$. In some experiments this method converged early, while other cases took more than n iterations to reduce rounding errors. In a history beautifully chronicled by Liesen and Strakoš, this odd behavior kept the conjugate gradient method on the algorithmic sidelines until the 1970s, when application problems grew to sufficient size for its often rapid convergence to be recognized, and needed. Soon the MINRES algorithm followed [20], using a three-term recurrence to minimize the Euclidean norm of the residual,

$$\|b - Ax_k\|_2 = \min_{\widehat{x}\in\mathcal{K}_k(A,b)} \|b - A\widehat{x}\|_2, \tag{1.2}$$

for symmetric A. In a creative burst in the 1980s and early 1990s, many new methods were proposed for nonsymmetric A. These methods are central to science and engineering simulations, consuming many cycles of the world's fastest computers. Despite the popularity of these methods, fundamental mathematical questions remain about their performance, questions that have spurred diverse work in linear algebra and approximation theory for several decades.

2. A projection taxonomy for optimal methods

The Krylov optimization problems in (1.1) and (1.2) produce different approximate solutions x_k, and one might imagine other methods that minimize different measures of error. To organize such algorithms, Saad and Schultz proposed a projection framework [22], and in a similar spirit Liesen and Strakoš classify optimal Krylov algorithms in their Chapter 2. For simplicity we assume the initial approximation $x_0 = 0$. The iterates x_k are drawn from the k-dimensional *approximation space* $\mathcal{S}_k$, leaving the residual $r_k = b - Ax_k \in b - A\mathcal{S}_k$. From $\mathcal{S}_k$ the method selects x_k to be orthogonal in the Euclidean inner product to the *constraint space* $\mathcal{C}_k$. Different algorithms follow from different choices for $\mathcal{S}_k$ and $\mathcal{C}_k$. For conjugate gradients, $\mathcal{S}_k = \mathcal{C}_k = \mathcal{K}_k(A,b)$. MINRES and its nonsymmetric generalization, GMRES [23], take $\mathcal{S}_k = \mathcal{K}_k(A,b)$ and $\mathcal{C}_k = A\mathcal{K}_k(A,b)$. For symmetric A, SYMMLQ [20] uses $\mathcal{S}_k = A\mathcal{K}_k(A,b)$ and $\mathcal{C}_k = \mathcal{K}_k(A,b)$, thus minimizing the Euclidean norm of the error, $\|x - x_k\|_2$. Fletcher's BiCG (biconjugate gradient) algorithm [12] for nonsymmetric A uses $\mathcal{S}_k = \mathcal{K}_k(A,b)$ and $\mathcal{C}_k = \mathcal{K}_k(A^*, \widetilde{b})$, where A^* denotes the adjoint and $\widetilde{b}$ is an auxiliary vector whose role remains somewhat obscure. Like many methods for nonsymmetric A, BiCG does not satisfy a simple optimality property.

3. Arnoldi and Lanczos

When implementing a Krylov subspace method, such as the one suggested by (1.2), a practical problem emerges: as k grows, the basis vectors $A^j b$ for $\mathcal{K}_k(A,b)$ increasingly align with the dominant eigenvector(s) of A, causing the rapid onset of numerical instabilities in any code that builds this power basis. For better behavior construct an orthonormal basis for $\mathcal{K}_k(A,b)$ starting with $v_1 = b/\|b\|$; then at the kth step use Gram–Schmidt to orthogonalize Av_k against the previous basis vectors $v_1, \ldots, v_k$, mimicking the usual construction of orthogonal polynomials. The resulting *Arnoldi process* [3] gives

$$v_{k+1} = \frac{1}{h_{k+1,k}}\big(I - v_1 v_1^* - \cdots - v_k v_k^*\big) A v_k, \tag{3.1}$$

where $h_{k+1,k}$ normalizes the projected vector. (The work increases at every step, as there are more vectors to orthogonalize against.) Rewrite (3.1) as

$$Av_k = h_{1,k}v_1 + h_{2,k}v_2 + \cdots + h_{k,k}v_k + h_{k+1,k}v_{k+1}$$

with $h_{j,k} = v_j^* A v_k$, and stack these relations together in columns to get

$$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^*, \tag{3.2}$$

where

$$V_k = [v_1 \ v_2 \ \cdots \ v_k]$$

and $H_k = V_k^* A V_k$ is *upper Hessenberg*: $h_{j,k} = 0$ when $j > k+1$. When A is symmetric, so too is H_k, implying that $h_{j,k} = 0$ with $j < k-1$. Thus H_k is *tridiagonal*, reducing (3.1) to the efficient three-term recurrence

$$v_{k+1} = \frac{1}{h_{k+1,k}}\big(Av_k - (v_{k-1}^* A v_k)v_{k-1} - (v_k^* A v_k)v_k\big).$$

This approach, proposed by Lanczos in 1950 [17], predates Arnoldi's general decomposition. The Lanczos method naturally parallels the three-term recurrence used to construct orthogonal polynomials over a real domain, where the reality of the domain gives the short recurrence in the same way $A = A^*$ does. In their third chapter, Liesen and Strakoš devote nearly 100 pages to the many connections between this three-term recurrence and orthogonal polynomials, Gauss–Christoffel quadrature, continued fractions, moment matching model reduction, and inverse eigenvalue problems for Jacobi matrices. These deep, elegant connections have been often studied (e.g., the 1952 conjugate gradient paper [16] and Fischer's 1996 monograph [11]), but Liesen and Strakoš provide an especially thorough treatment. We briefly sketch the connections to discrete measures and moment matching.

Given symmetric A with distinct eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ and unit eigenvectors $y_1, \ldots, y_n$, and $\|b\|_2 = 1$, construct

$$\omega(\lambda) = \sum_{j=1}^{n} |y_j^* b|^2 \chi_{(-\infty,\lambda_j]}(\lambda),$$

where χ_S is the indicator function for $S \subset \mathbb{R}$. Let $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$ denote the eigenvalues of H_k with corresponding unit eigenvectors $q_1^{(k)}, \ldots, q_k^{(k)}$, and let e_1 be the first column of the $k \times k$ identity matrix. Then

$$\omega_k(\lambda) = \sum_{j=1}^{k} |(q_j^{(k)})^* e_1|^2 \chi_{(-\infty,\theta_j^{(k)}]}(\lambda)$$

approximates ω in the following sense. The points $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$ are the roots of the kth orthogonal polynomial for the inner product $(f,g) = \int f(\lambda)\overline{g(\lambda)}\,\mathrm{d}\omega(\lambda)$, and these roots give nodes of the k-point Gaussian quadrature rule for

$$\int f(\lambda)\,\mathrm{d}\omega(\lambda) = \sum_{j=1}^{n} |y_j^* b|^2 f(\lambda_j) \approx \sum_{j=1}^{k} |(q_j^{(k)})^* e_1|^2 f(\theta_j^{(k)}) = \int f(\lambda)\,\mathrm{d}\omega_k(\lambda),$$

which is exact when f is a polynomial of degree $2k-1$ or less. This approximation has a close connection to model order reduction for dynamical systems [2], for the quadrature result implies

$$b^* A^j b = e_1^* H_k^j e_1, \qquad j = 0, \dots, 2k-1.$$

Thus the transfer function $e_1^*(z - H_k)^{-1} e_1$ for the k-dimensional dynamical system

$$\widehat{x}'(t) = H_k \widehat{x}(t) + e_1 u(t), \qquad \widehat{y}(t) = e_1^* \widehat{x}(t),$$

matches the first $2k$ moments (as $z \to \infty$) of the transfer function $b^*(z - A)^{-1} b$ for the n-dimensional system

$$x'(t) = Ax(t) + bu(t), \qquad y(t) = b^* x(t).$$

This theory enriches our understanding of iterative methods for symmetric matrices. Only a weak shadow of these results extends to the nonsymmetric case, resulting in a theory that is more fragmented and far less complete.

4. No short recurrences for nonsymmetric matrices

In the early 1980s, the burgeoning success of Krylov algorithms for symmetric matrices (with the underlying Lanczos three-term recurrence) led to a quest for a similarly efficient optimal algorithm for all nonsymmetric matrices. To spur research, money was put on the table at the 1981 Gatlinburg Symposium [1]:

> "A prize of \$500 has been offered by Gene Golub for the construction of a 3-term conjugate gradient like descent method for non-symmetric real matrices or a proof that there can be no such method."

In a landmark 1984 paper, Faber and Manteuffel resolved the question in the negative [10] (and subsequently collected the payout); related contemporaneous results were obtained by Voevodin and Tyrtyshnikov. Putting Golub's bet into precise terms turns out to be quite subtle. Liesen, Strakoš, and coauthors have explicated and illuminated the Faber–Mantueffel result in a series of recent papers, and the neat synthesis in Chapter 4 of their book should become the first place students turn to understand this fundamental result. Crucial to the argument is the definition of "gradient like descent method." The iterate x_k should minimize the error $\|x - x_k\|$ over the kth Krylov subspace $\mathcal{K}_k(A, b)$ in a norm induced by an inner product that can depend on A but not b. To build this iterate, one needs a basis $\{v_1, \dots, v_k\}$ for $\mathcal{K}_k(A, b)$, which should be extended at the next step by orthogonalizing Av_k against the previous basis vectors (in the specified inner product). In this context Golub's problem amounts to finding those A that permit such orthogonalization with a three-term recurrence. Aside from a few minor special cases, the Faber–Manteuffel Theorem says that no such recurrence exists unless the matrix is diagonalizable with all its eigenvalues on a line in the complex plane.

After completely treating the main result, Liesen and Strakoš explore intriguing questions on the margins of Golub's bet. Relaxing the method for constructing the basis for $\mathcal{K}_k(A, b)$ opens the door to Gragg's isometric Arnoldi process (an early contribution to the theory of orthogonal polynomials on the unit circle [24]), which applies when A is unitary (or in a small class of related matrices); special techniques also exist for low-rank perturbations of symmetric matrices [5]. Alternatively, one could optimize over other subspaces, such as generalized Krylov subspaces generated by products of both A and A^* [8]. Unfortunately, none of these clever variations have yet led to fast algorithms applicable to a broad class of nonsymmetric matrices.

The Faber–Manteuffel result gave license for two developments: optimal methods with *long recurrences* (e.g., GMRES), and *suboptimal* methods with short recurrences (e.g., restarted GMRES, CGS, QMR, TFQMR, BiCGSTAB, IDR).

5. Convergence theory

In the past twenty years the field has moved from rapid algorithmic invention to refined theoretical work for nonsymmetric matrices. While the methods are relatively easy to describe and have proved enormously powerful in important large-scale applications, their convergence properties remain inscrutable, even for the optimal GMRES algorithm. The performance of these methods in finite precision arithmetic adds another layer of complexity. In their final chapter Liesen and Strakoš address key theoretical and practical aspects of convergence.

The behavior of methods for symmetric A is completely understood via polynomial approximation problems on the spectrum. When A is nonsymmetric (and, especially, nonnormal: A lacks an orthonormal basis of eigenvectors), the convergence theory must be more sophisticated. Two striking results make this clear.

Greenbaum, Pták, Strakoš [15] (with some later refinements) proved that any nonincreasing convergence curve is possible for GMRES (1.2). More precisely, given any set $\{\|b - Ax_k\|_2\}_{k=0}^{n-1}$ and *any n nonzero eigenvalues*, one can construct A with the desired spectrum and b for which the GMRES iterates $\{x_k\}$ give the specified residual norms. Thus, no convergence theory can be based on eigenvalues alone.

The first step to analyzing a residual-minimizing method (1.2) is the inequality

$$\|b - Ax_k\|_2 = \min_{\substack{\deg(p)\le k\\ p(0)=1}} \|p(A)b\|_2 \le \min_{\substack{\deg(p)\le k\\ p(0)=1}} \|p(A)\|_2\|b\|_2, \tag{5.1}$$

which separates A from b. Given A, does there always exist some b for which equality is attained? The answer turns out to be "yes" for normal A, but not in general. With a striking family of 4×4 matrices, Toh [26] showed that the left-hand side of (5.1) could be *arbitrarily smaller* than the right-hand side for all b. Given the complexity of incorporating b into convergence analysis and the perceived rarity of examples like Toh's, work has proceeded to bound the "Ideal GMRES" problem

$$\min_{\substack{\deg(p)\le k\\ p(0)=1}} \|p(A)\|_2,$$

typically optimizing p over sets in the complex plane such as the spectrum, pseudospectra, polynomial numerical hulls, and numerical range. A recent contribution is Crouzeix's Theorem: for any f analytic on the numerical range $W(A)$,

$$\|f(A)\| \le 11.08 \max_{z\in W(A)} |f(z)|,$$

with the enticing conjecture that 11.08 can be replaced by 2 [7].

6. Context and outlook

Liesen and Strakoš open and close their book with an essential point that is easy to overlook in theoretical studies: the equation $Ax = b$ arises as part of a larger problem-solving process that begins with mathematical modeling and data collection, and concludes with some answer derived from x. These diverse steps incur all varieties of error: in instrumentation, modeling, linearization, discretization, convergence tolerance, floating-point precision, interpretation. Framing $Ax = b$ in this context informs the norm in which one computes, the convergence criterion to which one iterates, and the nature of b. For example, if A and b arise from discretization of a differential equation, the Fourier coefficients of b are likely to decay, unlike those of a random b, and this will influence convergence.

This thoughtful perspective runs throughout the book. The topics within its purview—classification of algorithms; orthogonal polynomials, continued fractions, and moment problems; the Faber–Manteuffel theorem; convergence theory in exact and finite precision—are covered with inspiring clarity, precision, and thoroughness. Of particular note are the rich historical details that illuminate overlooked early contributions, such as Vorobyev's monograph on the moment problem [29].

This book makes an important contribution to the growing monograph literature on Krylov subspace methods. Since the 1990s numerous books have appeared on the subject, often oriented toward practitioners and suitable for broad graduate courses; see, e.g., [4, 13, 19, 21, 28]. Those seeking a survey of algorithms and preconditioners are well served by these texts. The book under review places less emphasis on algorithmic variety, especially suboptimal methods for nonsymmetric matrices. (For example, the popular restarted variant of GMRES is only tangentially mentioned.) This selectivity allows Liesen and Strakoš to cover topics in sufficient detail to reward engaged readers with a great depth of understanding. The book is an excellent resource for established scholars and research students entering the field, and for focused study in a graduate seminar.

To conclude, we mention several problems in the spirit of the book but beyond its contents, which remain challenges going forward.

The eigenvalues of the compression $H_k = V_k^* A V_k$ in (3.2) are Rayleigh–Ritz eigenvalue estimates for A from the Krylov subspace. For symmetric A these Ritz values interlace the spectrum. Progress is just beginning on *inverse numerical range* problems, which study how Ritz values distribute over $W(A)$; see, e.g., [6, 18]. No work yet addresses the important analogous question for the harmonic Rayleigh–Ritz values that are the roots of the GMRES residual polynomial.

At present there are rather more questions than answers about the convergence of suboptimal iterations for nonsymmetric A. Intriguing results hint at the complex behavior still remaining to be understood, such as Greenbaum's theorem on the role of the vector $\widetilde{b}$ in the BiCG algorithm [14], and the exotic dynamics of restarted GMRES [9]. This challenging area would benefit from creative new approaches.

In recent years interest has grown in large-scale Lyapunov and Sylvester matrix equations, $AX + XB = C$, for the unknown matrix X. This equation can be expressed as $(I \otimes A + B^T \otimes I)\mathrm{vec}(X) = \mathrm{vec}(C)$, but the dimensions usually prohibit even *storage* of $\mathrm{vec}(X)$, thus ruling out methods like GMRES before they even begin. The Sylvester equation is only tractable because X often has accurate low-rank approximations. Krylov projection still plays an important role [25].

Krylov subspace methods have a rich history, essential applications, and a multitude of outstanding challenges: Liesen and Strakoš honor the subject admirably.

References

[1] SIGNUM Newsletter, 16 (1984), p. 7.

[2] Athanasios C. Antoulas, *Approximation of large-scale dynamical systems*, Advances in Design and Control, vol. 6, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. With a foreword by Jan C. Willems. MR2155615 (2006c:93001)

[3] W. E. Arnoldi, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math. **9** (1951), 17–29. MR0042792 (13,163e)

[4] Owe Axelsson, *Iterative solution methods*, Cambridge University Press, Cambridge, 1994. MR1276069 (95f:65005)

[5] Bernhard Beckermann and Lothar Reichel, *The Arnoldi process and GMRES for nearly symmetric matrices*, SIAM J. Matrix Anal. Appl. **30** (2008), no. 1, 102–120, DOI 10.1137/060668274. MR2399571 (2009k:65053)

[6] Russell Carden and Derek J. Hansen, *Ritz values of normal matrices and Ceva's theorem*, Linear Algebra Appl. **438** (2013), no. 11, 4114–4129, DOI 10.1016/j.laa.2012.12.030. MR3034520

[7] Michel Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Funct. Anal. **244** (2007), no. 2, 668–690, DOI 10.1016/j.jfa.2006.10.013. MR2297040 (2008f:47020)

[8] L. Elsner and Kh. D. Ikramov, *On a condensed form for normal matrices under finite sequences of elementary unitary similarities*, Proceedings of the Fifth Conference of the International Linear Algebra Society (Atlanta, GA, 1995), Linear Algebra Appl. **254** (1997), 79–98, DOI 10.1016/S0024-3795(96)00526-5. MR1436675 (98b:15011)

[9] Mark Embree, *The tortoise and the hare restart GMRES*, SIAM Rev. **45** (2003), no. 2, 259–266 (electronic), DOI 10.1137/S003614450139961. MR2010378 (2004j:65043)

[10] Vance Faber and Thomas Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal. **21** (1984), no. 2, 352–362, DOI 10.1137/0721026. MR736337 (85i:65038)

[11] Bernd Fischer, *Polynomial based iteration methods for symmetric linear systems*, Wiley-Teubner Series Advances in Numerical Mathematics, John Wiley & Sons, Ltd., Chichester; B. G. Teubner, Stuttgart, 1996. MR1449136 (98e:65017)

[12] R. Fletcher, *Conjugate gradient methods for indefinite systems*, Numerical analysis (Proc 6th Biennial Dundee Conf., Univ. Dundee, Dundee, 1975), Springer, Berlin, 1976, pp. 73–89. Lecture Notes in Math., Vol. 506. MR0461857 (57 #1841)

[13] Anne Greenbaum, *Iterative methods for solving linear systems*, Frontiers in Applied Mathematics, vol. 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR1474725 (98j:65023)

[14] Anne Greenbaum, *On the role of the left starting vector in the two-sided Lanczos algorithm and nonsymmetric linear system solvers*, in Numerical Analysis 1997, D. F. Griffiths, D. J. Higham, and G. A. Watson, eds., Addison Wesley Longman, Harlow, Essex, UK, 1998, pp. 124–132.

[15] Anne Greenbaum, Vlastimil Pták, and Zdeněk Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 3, 465–469, DOI 10.1137/S0895479894275030. MR1397238 (97c:65057)

[16] Magnus R. Hestenes and Eduard Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards **49** (1952), 409–436 (1953). MR0060307 (15,651a)

[17] Cornelius Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards **45** (1950), 255–282. MR0042791 (13,163d)

[18] S. M. Malamud, *Inverse spectral problem for normal matrices and the Gauss-Lucas theorem*, Trans. Amer. Math. Soc. **357** (2005), no. 10, 4043–4064, DOI 10.1090/S0002-9947-04-03649-9. MR2159699 (2006e:15018)

[19] G. Meurant, *Computer solution of large linear systems*, Studies in Mathematics and its Applications, vol. 28, North-Holland Publishing Co., Amsterdam, 1999. MR1700774 (2000f:65003)

[20] C. C. Paige and M. A. Saunders, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. **12** (1975), no. 4, 617–629. MR0383715 (52 #4595)

[21] Yousef Saad, *Iterative methods for sparse linear systems*, 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003. MR1990645 (2004h:65002)

[22] Youcef Saad and Martin H. Schultz, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comp. **44** (1985), no. 170, 417–424, DOI 10.2307/2007961. MR777273 (86d:65047)

[23] Youcef Saad and Martin H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[24] Barry Simon, *Orthogonal polynomials on the unit circle. Part 1*, American Mathematical Society Colloquium Publications, vol. 54, American Mathematical Society, Providence, RI, 2005. Classical theory. MR2105088 (2006a:42002a)

[25] V. Simoncini, *Computational methods for linear matrix equations*. Manuscript, January 2014.

[26] Kim-Chuan Toh, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 1, 30–36, DOI 10.1137/S089547989427909X. MR1428196 (97i:65054)

[27] Lloyd N. Trefethen and David Bau III, *Numerical linear algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[28] Henk A. van der Vorst, *Iterative Krylov methods for large linear systems*, Cambridge Monographs on Applied and Computational Mathematics, vol. 13, Cambridge University Press, Cambridge, 2003. MR1990752 (2005k:65075)

[29] Yu. V. Vorobyev, *Method of moments in applied mathematics*, Gordon and Breach Science Publishers, New York-London-Paris, 1965. Translated from the Russian by Bernard Seckler. MR0184400 (32 #1872)

[30] J. Wallis, *Opera Mathematica*, vol. 1, Oxford University Press, 1695.

Mark Embree
Department of Mathematics, Virginia Tech
E-mail address: embree@vt.edu