Numerical Linear Algebra with Applications, Vol. 4(4), 333–347 (1997)

Operator Trigonometry of Iterative Methods

K. Gustafson*

Department of Mathematics, University of Colorado, Boulder, CO 80309–0395, U.S.A.

A new and general approach to the understanding and analysis of widely used iterative methods for the numerical solution of the equation $Ax = b$ is presented. This class of algorithms, which includes CGN, GMRES. ORTHOMIN, BCG, CGS, and others of current importance, utilizes residual norm minimizing procedures, such as those often found under the general names Galerkin method, Arnoldi method, Lanczos method, and so on. The view here is different: the needed error minimizations are seen trigonometrically.

Numer. Linear Algebra Appl., Vol. 4, 333–347 (1997)
(No. of Figures: 0 No. of Tables: 1 No. of Refs: 27)

KEY WORDS iterative methods; operator trigonometry; antieigenvalue

1. Introduction

The goal of this paper is to establish the general connection between trigonometric operator theory [7–9] and iterative algorithms such as GCR, GCR(k), GCG, PCG, Orthomin, CGN, GMRES, CGS, BCG, MinRes, Lanczos, Arnoldi, and others with the same basic characteristics, e.g., residual minimization in selected iterative subspaces and CG-like convergence in roughly n-steps for $n \times n$ system $Ax = b$. This paper is an expansion of the conference lecture [10] and the conference presentation [11]. Section 2 recalls the needed essentials from the operator trigonometry theory, which originated in the period 1967–1969 from problems in operator semigroups. Section 3 contains, with some compression here, the content of the conference paper [11]. In particular, Section 3 presents new trigonometric

* Correspondence to K. Gustafson, Department of Mathematics, University of Colorado, Boulder, CO 80309–0395, U.S.A.

CCC 1070–5325/97/040333–15 $17.50

Received 29 September 1996
Revised 15 January 1997

interpretations of a representative set of the above mentioned important linear solvers. This is done by direct reference to representative papers [4,5,22,23,26,27].

Then in Section 4 we visit the Kantorovich–Wielandt condition number angle, and reveal its natural geometrical meaning in terms of our operator trigonometry. In Sections 5–7 we return to iterative methods and establish new trigonometric meanings of, respectively, the Richardson and Uzawa algorithms, the Chebyshev centers and midwidths, the Jacobi, SOR, and SSOR methods. Section 8 contains a discussion of superlinear convergence for conjugate gradient methods, as seen from the operator trigonometry viewpoint. Section 9 has been added to exhibit operator trigonometry aspects of AMLI (algebraic multilevel iteration) methods which became apparent at the Nijmegen (1996) conference.

For expositional efficiency, we often use the shortcut of direct reference to the excellent books [1,3,17–19,21,24], and when following a specific one of these, we adhere to the notation of that specific book for the reader's convenience.

2. Operator trigonometry

Because recently a rather extensive account [7–9,12,13,16] of the operator trigonometry theory has already been given, including its early development in the period 1967–1969, the discussion here will be truncated to just the needed basic ideas. The angle $\phi(A)$ of an operator A was defined through its cosine

$$\cos A = \inf \frac{\mathrm{Re}[Ax, x]}{\|Ax\|\|x\|}, \quad x \in D(A),\ Ax \neq 0. \tag{2.1}$$

The early work was concerned with arbitrary densely defined operators A in a real or complex Banach space, the brackets $[y, x]$ denoting any specified semi-inner product, the Re taken because the operators of most interest were accretive $A : \mathrm{Re}[Ax, x] \geq 0$. Briefly, the concept of $\phi(A)$, and in particular $\cos A$, grew out of a need to sharpen criteria for BA to generate a contraction semigroup, given that A did. A key question in those considerations was to know when the operator product BA of two accretive operators A and B is itself accretive. See the account in [7]. Note that the notion of accretivity, when applied to real matrices A in a finite dimensional Hilbert space, is the same as A positive real: $\mathrm{Re}\langle Ax, x\rangle \geq 0$, or equivalently, $\mathrm{Re}A = (A + A^{\mathrm{T}})/2$ symmetric positive semidefinite. To answer the question of when BA is accretive, given that A and B are, the notion of $\sin B$ arose naturally, defined by

$$\sin B = \min_{\epsilon > 0} \|\epsilon B - I\| \tag{2.2}$$

The condition for accretivity of BA became

$$\sin B \leq \cos A \tag{2.3}$$

Then an important min-max result, which we may state here in the form

$$\cos^2 A + \sin^2 A = 1 \tag{2.4}$$

for all strongly accretive operators on a Hilbert space, was shown. See the account in [9]. This established the operator trigonometry.

Shortly later the notion $\cos A$ was replaced by the notion of first anti-eigenvalue μ_1. The corresponding minimizing vector in (2.1) was called a first anti-eigenvector. Higher anti-eigenvalues and antieigenvectors were also defined. For positive definite self-adjoint operators A, we know that

$$\mu_1 \equiv \cos A = \frac{2\sqrt{mM}}{M+m}, \quad \sin A = \frac{M-m}{M+m} \tag{2.5}$$

where m and M are the greatest lower bound and least upper bound for the spectrum $\sigma(A)$. The corresponding first anti-eigenvector, which is defined to be that which A "turns" the maximum amount possible, is known for symmetric positive matrices A to be

$$x^1 = \left(\frac{\lambda_n}{\lambda_1+\lambda_n}\right)^{1/2} x_1 + \left(\frac{\lambda_1}{\lambda_1+\lambda_n}\right)^{1/2} x_n \tag{2.6}$$

where now we have introduced the matrix conventions $\lambda_1 = m =$ smallest eigenvalue of A, $\lambda_n = M =$ largest eigenvalue of A, x_1 and x_n being corresponding unit eigenvectors. First anti-eigenvectors come in pairs: you can put a $-$ sign in front of the first coefficient in (2.6). These two first anti-eigenvectors are generally not orthogonal and their linear combinations are not first anti-eigenvectors. Anti-eigenvectors satisfy an interesting non-linear Euler equation, see the account in [8,9].

3. Trigonometric interpretation of iterative methods

In [6] the insertion of $\sin A$ from (2.5) into the Kantorovich [20,21] error bound for steepest descent produced an entirely new (i.e., trigonometric) understanding of that bound: that steepest descent Kantorovich convergence rate is exactly $\sin A$. In [7] it was noted that the well known bound

$$E_A(x_k) \overset{<}{=} 4\left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^{2k} E_A(x_0) \tag{3.1}$$

for conjugate gradient convergence is also trigonometric, namely,

$$\|x_k - x^*\|_{A^{1/2}} \overset{<}{=} 2(\sin(A^{1/2}))^k \|x_0 - x^*\|_{A^{1/2}} \tag{3.2}$$

for A symmetric positive definite. Here κ as usual denotes the condition number $M/m = \lambda_n/\lambda_1$. Let us now note that the sharper error estimate (see Daniel [3]) for CG convergence

$$E_A(x_k) \overset{<}{=} \left[\frac{2(1-\kappa)^k}{(1+\sqrt{\kappa})^{2k} + (1-\sqrt{\kappa})^{2k}}\right]^2 E_A(x_0) \tag{3.3}$$

may also be interpreted trigonometrically. For A symmetric positive definite, this convergence rate bound becomes

$$\begin{aligned}\left[\frac{2(1-\kappa)^k}{(1+\sqrt{\kappa})^{2k}+(1-\sqrt{\kappa})^{2k}}\right]^{1/k} &= \frac{2^{1/k}\left(1-\frac{M}{m}\right)}{\left[\left((1+\frac{M^{1/2}}{m^{1/2}})^2\right)^k\right]^{1/k}} \\ &= \frac{2^{1/k}\left(\frac{m-M}{m+M}\right)}{\left[\left(\frac{m+M+2m^{1/2}M^{1/2}}{m+M}\right)^k+\left(\frac{m+M-2m^{1/2}M^{1/2}}{m+M}\right)^k\right]^{1/k}} \\ &= \sin A\left[\frac{2}{(1+\cos A)^k+(1-\cos A)^k}\right]^{1/k}\end{aligned} \tag{3.4}$$

now stated in trigonometric terms.

Let us next consider the more sophisticated iterative algorithms such as GCR, CGN, GMRES, BCG, CGS, and others. Specifically, let us turn to the representative papers Greenbaum [5], Eisenstat, Elman, Schultz [4], Van der Sluis and Van der Vorst [26], Van der Vorst and Dekker [27], Nachtigal, Reddy, Trefethen [22], Nachtigal, Reichel, Trefethen [23], taken chronologically. This set of papers enables a good span of current iterative linear solvers, and hence is sufficient to show the potential generality of the new trigonometric approach.

Turning first to [5], an earlier paper analyzing splitting methods in the context of the CG algorithm, we note that the estimate [5, (2.6)] obtained by matrix splitting there can be written trigonometrically as

$$\frac{\|e^{(k)}\|_A}{\|e^{(0)}\|_A} \leq 2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k+\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} = \frac{2}{(\sin A^{1/2})^k+(\csc A^{1/2})^k} \tag{3.5}$$

Note also (as may be seen by using the expressions (2.5), or by doing the real trigonometry), that the bound (3.5) is the same as (3.4). Also let us mention that the general expression [5,(2.5)], with its factors $\lambda_i/(\lambda_i-\lambda_j)$ the same as those in anti-eigenvector components as mentioned in Section 2 above, can also be interpreted trigonometrically. Finally, note that the Tchebyshev polynomial [5, just above (2.6)] which is commonly substituted as a somewhat optimal polynomial choice in these considerations, namely

$$P_k(x) = T_k((2x-\lambda_{\max}-\lambda_{\min})/(\lambda_{\max}-\lambda_{\min}))/T_k((-\lambda_{\max}-\lambda_{\min})/(\lambda_{\max}-\lambda_{\min})) \tag{3.6}$$

also has trigonometric content. The optimizing constant there is $\sin A$, e.g., the normalizing denominator is $T_k(-(\sin A)^{-1})$.

Turning next to [4], we note that the fundamental estimate [4, (3.3)] for residual errors of GCR (estimates depending on this particular estimate occur in many of the other papers dealing with some of the other competing iterative methods)

$$\|r_i\|_2 \leq \min_{q_i\in P_i}\|q_i(A)\|_2\|r_0\|_2 \leq \left[1-\frac{(\lambda_{\min}(\mathrm{Re}A))^2}{\lambda_{\max}(A^{\mathrm{T}}A)}\right]^{1/2}\|r_0\|_2 \tag{3.7}$$

may be interpreted and in principle improved trigonometrically. For example, in the A symmetric positive definite case the quantity in the brackets $[1-m^2/M^2]$ can be improved to $[1-\cos^2 A]$ and thus the improved convergence rate $\sin A$. To obtain this, let us choose

$g_1(z) = 1 + \alpha z$ as in [4] so that we arrive at

$$\min_{\alpha<0} \|I + \alpha A\| = \sin A \tag{3.8}$$

the equality in (3.8) by the min-max theorem (see [9]). This establishes the improvement just mentioned.

Turning next to [26], where the issue of convergence speedup of CG and the roles of Ritz vectors are addressed, we would like to make three comments. First, in as much as we know that the first anti-eigenvector, in the A symmetric or normal case, is composed from just the first and last eigenvectors, it is perhaps too strong to state categorically as in [26] that the convergence of the Ritz vectors plays no role in the analysis of CG convergence. Our view that the anti-eigenvectors control the operator trigonometry and hence all turning angle dynamics as an algorithm progresses toward the solution x^* increases the emphasis on the role of the Ritz vectors as they approximate the eigenvectors, hence the anti-eigenvectors, in that convergence process. Second, the notion of 'second condition number' espoused in [26] is consistent with the idea (see [8,9]) that higher anti-eigenvectors should be thought of as combinatorial selections of eigenvector pairs. Third, the conclusion (for an isolated highest eigenvalue) that the convergence of Ritz values at the upper end of the spectrum will rarely lead to impressive increases in the convergence rate of CG, due to loss of orthogonality, reinforces our conjecture that the correct combinations of eigenvectors, not their orthogonality, is what is essential to a better understanding of the convergence of iterative methods.

Turning next to [27] and to the general question of the design of preconditioners for PCG schemes, we would like to propose that a goal alternate to that of seeking a low condition number $\kappa(HA)$ would be to seek a high first anti-eigenvalue μ. This point was discussed further in [11, Section 4]. We may also suggest that the Manteuffel variation on the Chebyshev method could be interpreted and perhaps sharpened (e.g., [27], (3.19)]). trigonometrically, in as much as the anti-eigenvalues can be expected to generally effect sharper versions of field of values bounds.

Turning next to [22], the authors there make a forceful case for regarding CGN convergence as governed principally by singular values, GMRES convergence as governed principally by eigenvalues, and CGS convergence as governed principally by pseudo-eigenvalues. We would like to assert that all such schemes should be regarded as converging according to, at least in part, anti-eigenvalues or more to the point, anti-eigenvectors. The latter are fundamental to an understanding of all turning of an operator A. The singular values measure only the dilation actions of $A^{\mathrm{T}}A$. The eigenvalues measure only the dilation actions of A. The pseudo-eigenvalues measure essentially growth rates of $(\lambda I - A)^{-1}$ in the field of values sense of the Kreiss theorem. None of these entities measure directly the turning angles of A. All of these dilation actions are combined with the angular actions of A though the anti-eigenvalues and anti-eigenvectors of A. Moreover, the latter theory could explain the breakdown of BCG and CGS, for the cosines of an operator iteration sequence tend to zero as the sequence loses positive definiteness.

Turning next to [23], it may be proposed that we need an 'Arnoldi' method for the estimation of anti-eigenvalues for non-normal A. That is, the Arnoldi methods estimate eigenvalues of A and it is probable [9] that the largest and smallest eigenvalues for non-symmetric A will continue to be important ingredients for the calculation/estimation of the first anti-eigenvalue and corresponding anti-eigenvectors of A.

4. Trigonometric meaning of the Kantorovich–Wielandt condition number angle

The Kantorovich error bound

$$E_A(x_{k+1}) \leq \left(1 - \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}\right) E_A(x_k) \tag{4.1}$$

is fundamental to many gradient methods, see Luenberger [21]. The fact that $\sin A$ is given by (2.5) immediately gave in [6] a new, trigonometric meaning to the Kantorovich error bound, namely

$$E_A(x_{k+1}) \leq (\sin A)^2 E_A(x_k) \tag{4.2}$$

In this section we want to extend this result to the more general Kantorovich–Wielandt inequalities. Excellent treatments of a number of equivalent forms of these inequalities may be found in Horne and Johnson [18] and in the earlier treatment Householder [19].

Theorem 4.1. *(Kantorovich–Wielandt) (a) Let A be an $n \times n$ non-singular matrix with spectral condition number κ, and define the angle θ in the first quadrant by*

$$\cot(\theta/2) = \kappa \tag{4.3}$$

Then

$$|\langle Ax, Ay\rangle| \leq \cos\theta \|Ax\| \|Ay\| \tag{4.4}$$

for every pair of orthogonal vectors x and y. Moreover there exists an orthonormal pair of vectors x, y for which equality holds.

(b) In the case that A is symmetric positive definite,

$$|\langle x, Ay\rangle|^2 \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \langle x, Ax\rangle\langle y, Ay\rangle \tag{4.5}$$

for every pair of orthogonal vectors x and y.

Thus the geometrical interpretation of the condition number angle θ is that of the minimum angle between Ax and Ay as x and y range over all possible orthonormal pairs of vectors. But from (2.5) we notice immediately in (4.5) the presence of $\sin\phi(A)$. From this follows [14] a new meaning of θ.

Theorem 4.2. *The condition number angle θ of the Kantorovich–Wielandt theory, defined by $\cot(\theta/2) = \kappa$, whose geometrical interpretation is that of the minimum angle between Ax and Ay as x and y range over all possible orthonormal pairs of vectors, is precisely related to the operator angle $\phi(A)$ of the operator trigonometry, whose geometrical interpretation is that of the maximum turning angle by A on single vectors x ranging over the whole domain, by*

$$\cos\phi(A^2) = \sin\theta \tag{4.6}$$

Remark 1

One will find various trigonometric discussions of the condition number angle θ in the literature. See in particular Householder [19, pp. 81–85 and pp. 98–102], Horne and Johnson [18, pp. 441–445], Saad [24, pp. 132–135]. Although such discussions circle around the fundamental result of Theorem 4.2, apparently a preoccupation with the condition number κ

precluded seeing the more natural operator-trigonometrical meaning of these inequalities. Also it must be admitted that one needs the 1968 min-max theorem (see [7]), e.g., the relationship (2.4), to obtain Theorem 4.2. Further geometrical connections between the two theories will be given elsewhere [14].

5. Trigonometric meaning of Richardson, Uzawa methods

Let us now visit the recent iterative methods literature. Specifically, in the recent book, Saad [24, p. 106] exposes the classic Richardson iteration

$$x_{k+1} = x_k + \alpha(b - Ax_k) \tag{5.1}$$

with iteration matrix $G_\alpha = I - \alpha A$ and convergence factor $\rho(I - \alpha A)$. The latter is optimized at

$$\alpha_{\text{opt}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}, \quad \rho_{\text{opt}} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \tag{5.2}$$

assuming that A has all its eigenvalues positive real. Let us establish the operator-trigonometric meaning of these optimal parameters.

Theorem 5.1. *In the Richardson iteration for strictly accretive A, the optimal parameters α_{opt} and ρ_{opt} of (5.2) correspond, respectively, to*

$$\epsilon_m = \frac{\text{Re}\langle Ax^1, x^1\rangle}{\|Ax^1\|^2}, \quad \sin A = \inf_{\epsilon>0} \|\epsilon A - I\| \tag{5.3}$$

of the operator trigonometry, where ϵ_m is the minimizing value giving sin A, *and where x^1 is a first anti-eigenvector of A.*

Proof
The figure [24, Fig. 4.4] depicting the spectral radius $\rho(G_\alpha) = \max\{|1 - \alpha\lambda_{\min}|, |1 - \alpha\lambda_{\max}|\}$ is exactly the figure used by the author (independently) in the early (1967–1969) operator trigonometry when developing the theory of sin A from the convexity properties of $\|\epsilon A - I\|$, see the accounts and references in [7–9]. For any unit vector x we have

$$\|(\epsilon A - I)x\|^2 = \epsilon^2\|Ax\|^2 - 2\epsilon\text{Re}\langle Ax, x\rangle + 1 \tag{5.4}$$

which is minimized at

$$\epsilon_m = \frac{\text{Re}\langle Ax, x\rangle}{\|Ax\|^2} \tag{5.5}$$

with minimum value $1 - (\text{Re}\langle Ax, x\rangle/\|Ax\|)^2$. By the min-max theorem (see [9]) the minimum of $\|\epsilon A - I\|$ is attained at the 1st anti-eigenvector x^1, where $\|(\epsilon_m A - I)x^1\|^2 = 1 - \cos^2 A = \sin^2 A$. It is sufficient to consider the symmetric part ReA and from (2.6) a first unit anti-eigenvector $x = x^1$ given by

$$x = (\lambda_{\max}^{1/2}x_1 + \lambda_{\min}^{1/2}x_n)(\lambda_{\max} + \lambda_{\min})^{-1/2}. \tag{5.6}$$

From (5.5) and (5.6) we compute, using the expression (2.5) for cos A and the fact that it is attained at the first antieigenvector,

$$\begin{aligned} \epsilon_m &= \frac{\mathrm{Re}\langle Ax,x\rangle}{\|Ax\|\cdot\|x\|}\cdot\frac{1}{\|Ax\|} \\ &= \frac{2\lambda_{\max}^{1/2}\lambda_{\min}^{1/2}}{(\lambda_{\max}+\lambda_{\min})}\cdot\frac{1}{\|Ax\|} \\ &= \alpha_{\mathrm{opt}}\cdot\frac{\lambda_{\max}^{1/2}\lambda_{\min}^{1/2}}{\|Ax\|} \end{aligned} \tag{5.7}$$

But

$$\begin{aligned} Ax &= \left(\frac{\lambda_{\max}}{\lambda_{\max}+\lambda_{\min}}\right)^{1/2}Ax_1+\left(\frac{\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}\right)^{1/2}Ax_n \\ &= (\lambda_{\max}^{1/2}\lambda_{\min}x_1+\lambda_{\min}^{1/2}\lambda_{\max}x_n)(\lambda_{\max}+\lambda_{min})^{-1/2} \end{aligned} \tag{5.8}$$

and thus

$$\begin{aligned} \|Ax\|^2 &= (\lambda_{\max}\lambda_{\min}^2+\lambda_{\min}\lambda_{\max}^2)(\lambda_{\max}+\lambda_{\min})^{-1} \\ &= \lambda_{\max}\lambda_{\min} \end{aligned} \tag{5.9}$$

from which (5.7) becomes

$$\epsilon_m = \frac{2}{\lambda_{\max}+\lambda_{\min}} = \alpha_{\mathrm{opt}} \tag{5.10}$$

■

Saad [24, p. 240] also considers the Uzawa Saddle Point algorithm from the point of view of Richardson iteration. Specifically, we may establish the following theorem.

Theorem 5.2. *For A symmetric positive definite and B of full rank and $S = B^{\mathrm{T}}A^{-1}B$, then Uzawa's algorithm has optimal convergence parameter*

$$\omega_{\mathrm{opt}} = \frac{\cos S}{(\lambda_{\min}(S)\lambda_{\max}(S))^{1/2}} = \frac{\cos S}{\|Sx^1\|} \tag{5.11}$$

where x^1 is a first anti-eigenvector for S.

Proof
From [24, p. 241] we know

$$\omega_{\mathrm{opt}} = \frac{2}{\lambda_{\min}(S)+\lambda_{\max}(S)} \tag{5.12}$$

from which (5.11) follows by the considerations above since $\omega_{\mathrm{opt}} = \epsilon_m$. ■

Remark 1
The fact that such a basic scheme as the Richardson iteration is optimized trigonometrically has apparently been overlooked up to now.

6. Trigonometric meaning of Chebyshev center, midwith

In Section 3 we observed without much elaboration some trigonometric interpretations of Chebyshev polynomial preconditionings. Here we would like to make a further interesting observation.

Saad [24,p. 357] considers the usual expression

$$\min_{p\in P_k,\, p(\gamma)=1} \max_{t\in[\alpha,\beta]} |p(t)| \tag{6.1}$$

where $[\alpha, \beta]$ is an interval containing the spectrum of A, and shows that for any $\gamma \stackrel{<}{=} \alpha$ the value of (6.1) is attained for the shifted and scaled Chebyshev polynomial of the first kind

$$\widehat{C}_k(t) = \frac{C_k\left(1+2\frac{\alpha-t}{\beta-\alpha}\right)}{C_k\left(1+2\frac{\alpha-\gamma}{\beta-\alpha}\right)} \tag{6.2}$$

When $\gamma = 0$, following [24] let

$$\theta = \frac{\beta+\alpha}{2}, \qquad \delta = \frac{\beta-\alpha}{2} \tag{6.3}$$

These are called center and midwidth of the interval $[\alpha, \beta]$, and the minimizing polynomial then may be expressed as

$$\frac{C_k\left(\frac{\theta-t}{\delta}\right)}{C_k\left(\frac{\theta}{\delta}\right)} \tag{6.4}$$

From these expressions a Chebyshev acceleration scheme

$$x_{k+1} = x_k + \rho_k[\rho_{k-1}(x_k - x_{k-1}) - 2\delta^{-1}(b - Ax_k) \tag{6.5}$$

is formulated, where

$$\rho_k = C_k(\theta/\delta)/C_{k+1}(\theta/\delta) \tag{6.6}$$

One can convert this scheme into one with trigonometric meanings, based upon the following lemma. For simplicity we just consider the case of A symmetric positive definite and $[\alpha, \beta] = [\lambda_1, \lambda_n]$ and we will not bother to write out the scheme (6.5) again, trigonometrically.

Theorem 6.1. *Note that*

$$\sin A = \frac{\text{Midwidth}(A)}{\text{Center}(A)} = \frac{\text{Midwidth}(A^{-1})}{\text{Center}(A^{-1})} = \frac{\delta}{\theta} \tag{6.7}$$

Proof
Immediate from the above and (2.5). ∎

Remark 1
An interesting enlarged Chebyshev-trigonometric theory of such accelerated schemes would seem possible but must be developed elsewhere.

7. Trigonometric meaning of Jacobi, SOR, SSOR methods

We now turn to Hackbush [17] where one will find extensive treatment of some of the standard iterative methods and multigrid smoothers, namely the Jacobi, SOR, and SSOR schemes, and various results for them and their variations.

Consider for A symmetric positive definite first the damped Jacobi method

$$x^{m+1} = x^m - \theta D^{-1}(Ax^m - b) \tag{7.1}$$

Then [17, p. 89] the optimal convergence rate is shown to be

$$\rho(M_\theta^{\mathrm{Jac}}) = \frac{\Lambda - \lambda}{\Lambda + \lambda} \tag{7.2}$$

where λ and Λ are the best constants such that $\lambda D \leqq A \leqq \Lambda D$, D the diagonal of A.

For SOR, the Samarskii–Nikolaev result for optimal minimization of the norm of the iteration matrix M_ω^{SOR} is shown [17, p. 94] to be

$$\|M_{\omega'}^{\mathrm{SOR}}\|_A \leqq \sqrt{\frac{\sqrt{\Gamma} - \sqrt{\gamma}}{\sqrt{\Gamma} + \sqrt{\gamma}}} \tag{7.3}$$

where the optimal relaxation parameter is

$$\omega' = \frac{2}{1 + \sqrt{\gamma \Lambda}} \tag{7.4}$$

Here γ and Λ are constants such that

$$\begin{aligned} &0 < \gamma D \leqq A \\ &(\tfrac{1}{2}D - E)D^{-1}(\tfrac{1}{2}D - E^H) \leqq \tfrac{\Gamma}{4}A \end{aligned} \tag{7.5}$$

and it has been assumed that $A = D - E - E^H$ where E is strictly lower triangular, and $0 < \omega < 2$.

For SSOR under the same assumptions a similar result is obtained [17, p. 118]. Namely with γ, Γ and E satisfying (7.5), the same optimal relaxation parameter ω' of (7.4) is obtained, from which

$$\rho(M_{\omega'}^{\mathrm{SSOR}}) \leqq \frac{\sqrt{\Gamma} - \sqrt{\gamma}}{\sqrt{\Gamma} + \sqrt{\gamma}} \tag{7.6}$$

Now, the point is that these expressions all carry trigonometric meaning, although the full theory will have to be worked out elsewhere. Let us formulate this assertion into a proposition, taking the liberty of entering new definitions in order to do so. In as much as the λ and Λ of (7.2) are the lower and upper bounds of A relative to D, we may call $(\Lambda - \lambda)/(\Lambda + \lambda)$ the sin of A relative to D, $\sin(A/D)$. With less intuition as concerns Γ in (7.5), we may similarly call the right hand side of (7.6) the sin of $A^{1/2}$ relative to D and E, or $\sin(A_{\gamma,\Gamma}^{1/2})$. With this terminology we then have the following proposition.

Proposition 7.1. *For the damped Jacobi method the optimal convergence rate is* $\sin(A/D)$. *For the SOR method the optimal convergence rate is bounded by* $(\sin(A_{\gamma,\Gamma}^{1/2}))^{1/2}$. *For the SSOR method the optimal convergence rate is bounded by* $\sin A_{\gamma,\Gamma}^{1/2}$.

Remark 1

A number of similar convergence rates may be found in [17] for methods related to Richardson, Jacobi, SOR, SSOR. Multigrid and Domain Decomposition methods (e.g., see [17,

Chapters 10 and 11]) seem less prone to trigonometric interpretation. See however for example [17, p. 379] for additive Schwarz iteraton with optimal damping factor $\theta = 2(\gamma + \Gamma)$ and [17, p. 399] for multigrid as a subspace decomposition method viewed as a Jacobi method on subspaces associated with a hierarchical basis.

8. Trigonometric meaning of superlinear convergence of conjugate gradient methods

Let us first return to the bounds (3.1), (3.3) and (3.5) given in Section 3 for conjugate gradient convergence rates. As noted there, (3.3), (3.4) and (3.5) are all the same bound, in different guises. In connection with such comparisons one finds interesting operator-trigonometric identities, such as the following.

Lemma 8.1. *For A symmetric positive definite,*

$$\sin(A^{1/2}) = \frac{\sin A}{1 + \cos A} = \left[\frac{1 - \cos A}{1 + \cos A}\right]^{1/2} \tag{8.1}$$

Proof
Using (2.5) and the spectral mapping theorem, we have

$$(\sin A^{1/2})^2 = \left[\frac{M^{1/2} - m^{1/2}}{M^{1/2} + m^{1/2}}\right]^2 = \frac{M + m - 2m^{1/2}M^{1/2}}{M + m + 2m^{1/2}M^{1/2}} \tag{8.2}$$

Thus

$$\begin{aligned} \sin(A^{1/2}) &= \left[\tfrac{1-\cos A}{1+\cos A}\right]^{1/2} \\ &= \tfrac{(1-\cos A)^{1/2}(1-\cos A)^{1/2}}{1+\cos A} \\ &= \tfrac{(1-\cos^2 A)^{1/2}}{1+\cos A} \\ &= \tfrac{\sin A}{1+\cos A} \end{aligned} \tag{8.3}$$

■

Operator trigonometry identities such as (8.1) are useful in visualizing gradient error bounds. For example, the Kantorovich steepest descent bound (4.1) seen trigonometrically (4.2), as compared to the conjugate gradient bound (3.1) seen trigonometrically (3.2), tells us that $E_A(x_k)/E_A(x_0)$ has convergence rate bound $(\sin A)^{2k}$ in steepest descent as compared to $4(\sin(A^{1/2}))^{2k}$. By Lemma 8.1 we know that the latter is actually faster by a factor of $4/(1 + \cos A)^{2k}$. Other operator-trigonometric identities could provide similar insights in other instances.

But, as is well known, whereas steepest descent convergence is slow and for a time at least usually follows the $\sin A$ rate, conjugate gradient schemes converge much faster than the (3.2) $\sin(A^{1/2})$ bound, or even the improved bounds such as (3.3), would indicate.

Why is this so? A good insight is given in Hackbusch [17, p. 272], following the Van der Sluis and Van der Vorst [26] analysis of such superconvergence. One may write the bound (3.3), which is the same as [17, (9.4.10)], as

$$\frac{\|e^k\|_A}{\|e^0\|_A} \leqq \frac{2c^k}{1 + c^{2k}} = \frac{2(\sin(A^{1/2}))^k}{1 + (\sin(A^{1/2}))^{2k}} \tag{8.4}$$

Table 1. Trigonometric superlinear convergence

$E(x_n)$	$c(x_n)$	$\phi(x_n)$	$s(x_n)$	Step size
0.825	1	0	0	0
0.76672	0.97227	13.526	0.23388	0.00297
0.71120	0.93025	21.526	0.36692	0.00914
0.24882	0.74692	41.765	0.66491	0.27400
6.25e-25	0.73424	42.757	0.67889	0.84058
2.77e-28	0.73424	42.757	0.67889	0.00250

As pointed out in [17, pp. 272–273], eventually the ratios $\|e^{k+1}\|_A/\|e^k\|_A$ become smaller than $c \approx 1 - 2/\sqrt{\kappa}$. In terms of our operator trigonometry, we know for the conjugate gradient method these ratios decrease at least as fast as

$$\begin{aligned} \sin(A^{1/2}) &= \frac{\lambda_{\max}^{1/2} - \lambda_{\min}^{1/2}}{\lambda_{\max}^{1/2} + \lambda_{\min}^{1/2}} \\ &= 1 - \frac{2\lambda_{\min}^{1/2}}{\lambda_{\max}^{1/2}} + \frac{2\lambda_{\min}}{\lambda_{\max}} - \frac{2\lambda_{\min}^{3/2}}{\lambda_{\max}^{3/2}} + \cdots \\ &\sim 1 - 2\left(\frac{\lambda_1}{\lambda_n}\right)^{1/2} \end{aligned} \tag{8.5}$$

In the case of gradient descent, the error converges to the subspace $V = \mathrm{sp}\{x_1, x_n\}$. However, this cannot happen for conjugate gradient: if the error e^m was exactly in the subspace V, in two more steps the residual would be exhausted and e^{m+2} would be zero. In fact, the conjugate gradient error moves toward $V^\perp$. On $V^\perp$ the operator A has reduced spectrum $\lambda_2 \leqq \cdots \leqq \lambda_{n-1}$. Hence A's condition number improves to $\kappa(A|_{V^\perp}) = \lambda_{n-1}/\lambda_2$. Then as in (8.5) the convergence rate improves to‘

$$\sin((A|_{V^\perp})^{1/2}) = \frac{\lambda_{n-1}^{1/2} - \lambda_2^{1/2}}{\lambda_{n-1}^{1/2} + \lambda_2^{1/2}} \sim 1 - 2\left(\frac{\lambda_2}{\lambda_{n-1}}\right)^{1/2} \tag{8.6}$$

These observations are consistent with the conjecture [8,9] that higher anti-eigenvectors, viewed as the appropriate linear combinations of eigenvectors, control the convergence of gradient and conjugate gradient methods, and that the resulting higher anti-eigenvalues determine the improving convergence rates. This phenomenon was observed in the gradient example of [9], where A had eigenvalues 1, 2, 10, 20, and where the gradient convergence rate moved from $\sin A = (20-1)/(20+1) \cong 0.905$ to the better rate partial sine $\sin_{2,3} A = (10-2)/(10+2) \cong 0.667$. This corresponds to a move from $V = \mathrm{sp}\{x_1, x_4\}$ to $V^\perp = \mathrm{sp}\{x_2, x_3\}$. Note that the first anti-eigenvectors live in and also span V, and recall that iterating on first anti-eigenvectors corresponds to iterating at both the highest condition number κ and at the largest turning angle $\phi(A)$, whereas both of these are smaller on $V^\perp$.

Table 1 shows the same example [9, Example 5.1], where A is the diagonal 4×4 matrix with 20, 10, 2, 1 as diagonal elements, $b = (1, 1, 1, 1)$, initial guess $x_0 = (0, 0, 0, 0)$, but now $Ax = b$ is solved by conjugate gradient rather than steepest descent. The key entry $s(x_n) = 0.66491 \approx \sin_{2,3} A = 0.667$ signals the onset of superlinear convergence.

Remark 1
More analysis of superlinear convergence in terms of anti-eigenvector behavior will be given elsewhere. Clearly for the case of general sparse but non-symmetric A this will necessitate a better understanding of the operator trigonometry of such operators.

9. Additional remark: trigonometry of AMLI methods

The AMLI (algebraic multilevel iteration) approach, to which the author was introduced at the Nijmegen (1996) AMLI Conference, has interesting connections to the operator trigonometry. Our purpose in this added remark is to quickly exhibit such connections, deferring a more complete development to elsewhere. For efficiency, we will follow the paper [2] presented at the Nijmegen conference. But see also the cited papers in [2]. The reader is also referred to [1, Chapter 9] for a presentation of the essentials of the AMLI approach.

Following [2], let us consider a nested sequence of finite element meshes $\Omega_\ell \supset \Omega_{\ell-1} \supset \cdots \supset \Omega_{k_0}$ and let $V_\ell, V_{\ell-1}, \ldots, V_{k_0}$ be the corresponding finite element spaces for an hp version of a FEM applied to a second order elliptic boundary value problem. In particular, at a chosen discretization level ℓ, let V_1 and V_2 be finite element (e.g., corresponding to a mesh partition) subspaces of V_ℓ such that $V_1 \cap V_2 = 0$. Let $\alpha(u, v)$ be the symmetric positive definite energy inner product on $H^1(\Omega)$ corresponding to the given elliptic operator. Let

$$\gamma = \sup_{u \in V_1, v \in V_2} \frac{a(u, v)}{(a(u, u) \cdot a(v, v))^{1/2}} \tag{9.1}$$

The entity γ is cosine of the angle between the subspaces V_1 and V_2 in the $a(u, v)$ inner product. When $\gamma < 1$, it is called the C.B.S. constant, corresponding to the strengthening acquired by the Cauchy–Bunyakowski–Schwarz inequality when restricted to the subspaces V_1 and V_2. It is known ([2,1]; see also the cited papers by Braess, Maitre and Musy, and others therein) that in the finite element subspace situation, γ can be computed as the maximum of its values (9.1) on subspaces $V_1^{(e)}$ and $V_2^{(e)}$ over single elements e. This is a valuable feature as it means that γ is independent of discretization refinement. In many instances the condition numbers of AMLI methods depend only on γ.

In particular, following [2], for hierarchical finite element basis functions one is led to a finite element symmetric matrix partitioning

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{9.2}$$

and the result [2, Lemma 2.1] that

$$\begin{aligned} \rho(A_{11}^{-1/2} A_{12} A_{22}^{-1/2}) &= \gamma \\ \lambda_{\max}(D^{-1}A) &= 1 + \gamma \\ \lambda_{\min}(D^{-1}A) &= 1 - \gamma \end{aligned} \tag{9.3}$$

where ρ is the spectral radius and where D is the block diagonal portion of A. From (9.3)

the condition number of the block-diagonally preconditioned system is

$$\kappa(D^{-1}A) = \frac{1+\gamma}{1-\gamma} \tag{9.4}$$

The same situation is seen generally for block Jacobi preconditioning of such systems, see [1, Table 9.1, p. 389].

Relevant to [2] and more generally with modification to all block diagonal and Schur complement preconditioning methods, we wish to establish here the following three facts.

Theorem 9.1. *(1) In the hierarchical finite element partitioned systems (9.2) above, the C.B.S. constant* $\gamma = \sin(D^{-1}A)$. *In other words, the cosine of the angle between* V_1 *and* V_2 *is the operator-trigonometric* $\sin(D^{-1/2}AD^{-1/2})$. *(2) In multilevel hierarchical finite element schemes, the interlevel condition number amplification factors are of the form* $(1 + \sin(M^{(k)^{-1}}A^{(k)}))/(1 - \sin(M^{(k)^{-1}}A^{(k)}))$. *(3) For piecewise linear finite element basis functions for the standard elliptic operator, the operator-trigonometric angle* $\phi(D^{-1/2}AD^{-1/2})$ *is less than* 60°, *for any triangle shape.*

Proof
For (1), from [2, Lemma 2.1] and using (9.3), (9.4) and (2.5) we have

$$\begin{aligned} \sin(D^{-1/2}AD^{-1/2}) &= \tfrac{\lambda_{\max}(D^{-1}A)-\lambda_{\min}(D^{-1}A)}{\lambda_{\max}(D^{-1}A)+\lambda_{\min}(D^{-1}A)} \\ &= \tfrac{(1+\gamma)-(1-\gamma)}{(1+\gamma)+(1-\gamma)} = \gamma \end{aligned} \tag{9.5}$$

We have made use of the well known fact that the generally non-selfadjoint operator $D^{-1}A$ has the same eigenvalues as $D^{-1/2}AD^{-1/2}$. This, by the way, is also how to treat the $\sin(A/D)$ in Proposition 7.1. Note also that whenever one has an expression $\kappa = (1 + \gamma)/(1 - \gamma)$ for a condition number for a positive operator T, then $\gamma = \sin T$. For (2), we refer to [2, Theorem 3.1], and we are assuming that the same assumptions that rendered (1) valid still prevail. For (3), we refer to [2, Section 5] where it is shown that $\gamma^2 < 3/4$ for any triangulation. Thus $\sin(D^{-1/2}AD^{-1/2}) < \sqrt{3}/2$ and $\phi(D^{-1/2}AD^{-1/2}) < \pi/3$. ■

Remark 1
In a parallel paper [15], the operator trigonometry of the standard Poisson equation model problem has been worked out in detail for finite difference discretizations.

Remark 2
In a previous paper [25] we explored the computational features of highly localized block Jacobi preconditioning on a massively parallel architecture.

REFERENCES

1. O. Axelsson. *Iterative Solution Methods*. Cambridge Press, Cambridge, 1994.
2. O. Axelsson. Stabilization of algebraic multilevel iteration methods. In O. Axelsson and B. Polman, editors. *Proc. Conf. Algebraic Multilevel Iteration Methods with Applications*, pages 49–62. Nijmegen, Netherlands, 1996.
3. J. Daniel. *The Approximate Minimization of Functionals*. Prentice Hall, NJ, 1971.
4. S. Eisenstat, H. Elman and M. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal*., 20, 345–357, 1983.

5. A. Greenbaum. Comparison of splittings used with the Conjugate Gradient Algorithm. *Numer. Math*., 33, 181–194, 1979.
6. K. Gustafson. Antieigenvalues in analysis. In C. Stanojevic and O. Hadzic, editors. *Proceedings of the Fourth International Workshop in Analysis and its Applications*, pages 57–69. Novi Sad, Yugoslavia, 1991.
7. K. Gustafson. Operator trigonometry. *Linear and Multilinear Algebra*, 37, 139–159, 1994.
8. K. Gustafson. Antieigenvalues. *Linear Algebra and Its Applications*, 208/209, 437–454, 1994.
9. K. Gustafson. Matrix trigonometry. *Linear Algebra and Its Applications*, 217, 117–140, 1995.
10. K. Gustafson. Computational trigonometry (abstract). In T. Manteuffel and S. McCormick, editors. *Proc. Colorado Conference on Iterative Methods*, Vol 1. Breckenridge, Colorado, 1994.
11. K. Gustafson. Trigonometric interpretation of iterative methods. In O. Axelsson and Ben Polman, editors, *Proc. Conf. Algebraic Multilevel Iteration Methods with Applications*, pages 23–29. Nijmegen, Netherlands, 1996.
12. K. Gustafson. Operator angles (Gustafson), matrix singular angles (Wielandt), operator deviations (Krein). In B. Huppert and H. Schneider, editors. *Collected Works of Helmut Wielandt*, Vol. 2, pages 356–367. De Gruyters, Berlin, 1996.
13. K. Gustafson. *Lectures on Computational Fluid Dynamics, Mathematical Physics, and Linear Algebra*. Kaigai Publishers, Tokyo, Japan, 1996, World Scientific, Singapore, 1997.
14. K. Gustafson. The geometrical meaning of the Kantorovich–Wielandt inequalities. (To appear).
15. K. Gustafson. Operator trigonometry of the model problem. (To appear).
16. K. Gustafson and D. Rao. *Numerical Range: The Field of Values of Linear Operators and Matrices*. Springer, New York, 1997.
17. W. Hackbusch. *Iterative Solution of Large Sparse Systems of Equations*. Springer, Berlin. 1994.
18. R. Horne and C. Johnson. *Matrix Analysis*. Cambridge Press, Cambridge, 1985.
19. A. Householder. *The Theory of Matrices in Numerical Analysis*. Blaisdell, New York, 1964.
20. L. Kantorovich. Functional analysis and applied mathematics. *Uspehi Mat. Nauk*., 3, 89–185, 1948.
21. D. Luenberger. *Linear and Nonlinear Programming*. Addison–Wesley, Menlo Park, CA, 1984.
22. N. Nachtigal, S. Reddy and L. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Applic*., 13, 778–795, 1992.
23. N. Nachtigal, L. Reichel and L. Trefethen. A Hybrid GMRES algorithm for nonsymmetric linear systems. *SIAM J. Matrix Anal. Applic*., 13, 796–825, 1992.
24. Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing, Boston, 1996.
25. N. Sobh and K. Gustafson. Preconditioned conjugate gradient and finite element methods for massively data-parallel architectures. *Computer Phys. Commun*., 65, 253-267, 1991.
26. A. Van Der Sluis and H. A. Van Der Vorst. The rate of convergence of conjugate gradients. *Numer. Math*., 48, 543–560, 1986.
27. H. A. Van Der Vorst and K. Dekker. Conjugate gradient type methods and preconditioning. *J. of Comp. and Appl. Math*., 24, 73–87, 1988.

