ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 14, 2010

On my Min-Max Theorem (1968) and its consequences

Karl Gustafson

Abstract. Central to the origins of my operator trigonometry, a theory in which I initiated the concepts of antieigenvalues and antieigenvectors, is my 1968 Min-Max Theorem. I will discuss its motivation, proof, and consequences. Special emphasis will be given here to a new view that $\sin\phi(A)$ may be viewed as a general optimum which encompasses many other optima of individual interest.

1. The application that gave the Min-Max Theorem

I announced my Min-Max Theorem in 1968 in [1] in connection with a question of multiplicative perturbation of semigroup generators which I was treating at that time in the papers [2], [3], and [4]. Then I was invited to the Third Symposium on Inequalities in Los Angeles in September 1969 and there I presented my Min-Max Theorem and its proof. Also there I first originated the terms antieigenvalue and antieigenvector. On short notice I published that presentation in abbreviated form in [6]. About twenty years went by during which I worked primarily in other domains. But I returned to my antieigenvalue theory in 1992 and strengthened and extended some results (see the books [7] and [14]). In 1999 I turned my attention to applying my antieigenvalue theory to matrix statistics (see the papers [8], [9], [10], and [12]).

Before I detail the original application that induced my Min-Max Theorem, I think it is much easier to describe that theorem in the simpler terms as it was seen by a very conscientious referee of my recent paper [13]. I take the following quote directly from his/her referee report: one of the best, detailed, referee reports that I have ever been privileged to receive! I do

Received March 23, 2010.
2010 *Mathematics Subject Classification.* 47B44, 80M50, 65K10, 15A60.
Key words and phrases. Min-max theorem, operator trigonometry, optimization.

Presented by Invitation under the same title at the 18th International Workshop on Matrices and Statistics, Smolenice Castle, Slovakia, June 23–27, 2009.

not know who the referee was, but it was an expert in operator theory who apparently had not previously been familiar with my antieigenvalue theory. "In the late sixties Karl Gustafson, motivated by a problem in perturbation theory of semi groups noted the following very interesting fact: Let A be a real positive definite $n \times n$ matrix on $\mathbb{R}^n$. Then the two quantities

$$\inf_{x \neq 0} \frac{\langle Ax, x\rangle}{\|x\|\|Ax\|} \quad \text{and} \quad \inf_{\epsilon>0} \sup_{\|x\| \leq 1} \|(\epsilon A - I)x\|$$

satisfying the same relation as the cosine and sine (respectively) of an angle: the sum of their squares yields 1."

The referee had honed in on the fact that for such matrices A with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$, I had shown in 1968 that the left and right quantities above have values

$$\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_n}}{\lambda_1 + \lambda_n} \quad \text{and} \quad \nu_1 = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}.$$

Clearly $\mu_1^2 + \nu_1^2 = 1$. I called μ_1 the first eigenvalue of A. Then setting $\cos \phi(A) = \mu_1$ determines the largest angle $\phi(A)$ through which A may turn any vector. The angle $\phi(A)$ I called the angle of A.

I liked the referee's perception of my operator trigonometry because it nicely brings out the two key elements of my theory for the special case of $n \times n$ symmetric matrices A. If you look at the early papers [2], [3], and [4] you will see those entities appearing in several inequalities therein, although in more general guise of λ_n and λ_1 replaced by lower and upper bounds m and M, respectively. The Min-Max Theorem is, however, more general.

Theorem 1 (Min-Max Theorem)**.** *Let A be a strongly accretive bounded operator on a Hilbert Space. Then*

$$\sup_{\|x\| \leq 1} \inf_{-\infty<\epsilon<\infty} \|(\epsilon A - I)x\|^2 = \inf_{\epsilon>0} \sup_{\|x\| \leq 1} \|(\epsilon A - I)x\|^2.$$

Strongly accretive means: $\mathrm{Re}\langle Ax, x\rangle \geq m_A > 0$ for all $\|x\| = 1$. Thus the Min-Max Theorem holds for all operators A with numerical range $W(A)$ strictly in the right half plane. General accretive operators was the context of the application that motivated my conceiving and proving the Min-Max Theorem. If A is strongly accretive, then $-A$ generates a contraction semigroup T_t of operators for $0 \leq t < \infty$. Such operator semigroup theory often goes under the name Hille–Yosida Theory. I wanted to left multiply the infinitesimal generator $-A$ by a strongly accretive operator B so that the multiplicatively perturbed operator $-BA$ was still the infinitesimal generator of a contraction semigroup. After some elementary functional analysis, this came down to the sufficient condition

$$\inf_{\epsilon>0} \sup_{\|x\| \leq 1} \|(\epsilon B - I)x\| \leq \inf_{x \neq 0} \frac{\mathrm{Re}\langle Ax, x\rangle}{\|Ax\|\|x\|}$$

for $-BA$ to remain a contraction semigroup generator.

It was natural to denote the right hand side of this inequality $\cos\phi(A)$ and to call it the first (real) antieigenvalue μ_1. But how about the left hand side? For B a bounded strongly accretive operator that becomes $\inf_{\epsilon>0}\|\epsilon B - I\|$ and it is easy to show that the infimum is a unique minimum $\|\epsilon_m B - I\|$ attained at some positive ϵ_m. For B a symmetric positive definite $n \times n$ matrix I knew that the situation was

$$\nu_1(B) = \|\epsilon_m B - I\| = \frac{\lambda_1(B) - \lambda_n(B)}{\lambda_1(B) + \lambda_n(B)},$$

and the sufficient condition was

$$\nu_1(B) \leq \cos\phi(A).$$

I formulated and proved my Min-Max Theorem just because I *wanted* the left side to be $\sin\phi(B)$.

The proof of the Min-Max Theorem need not be given here, see the books [7] and [14] for further elaboration of its details. Suffice it to say that first I looked at all the min-max theorems known to me and I could not use any of those to prove mine. So I resorted to the analytic geometry of parabolas, ellipses, and hyperbolas! That approach also yielded a short elementary proof of the Toeplitz–Hausdorff theorem (see [5]).

For $n \times n$ real symmetric positive definite matrices A, the proof of the Min-Max Theorem establishes that the (normed to one) two most-turned vectors are

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

where x_n and x_1 are any norm-one eigenvectors from the λ_n and λ_1 eigenspaces, respectively. I called these the first antieigenvectors. I also defined higher antiegenvalues and higher antieigenvectors on reduced subspaces of A. These are especially of consequence in the application of my antieigenvalues theory to matrix statistics (see the papers [8], [9], [10], and [12]).

There are many consequences of my Min-Max Theorem through the applications of my operator trigonometry to Markov processes, conjugate gradient solvers, relaxation schemes, wavelets, quantum mechanics, among others. See the recent survey [11]. However, in this paper I want to emphasize a new view of $\sin\phi(A)$ as a general optimum. This I have known since the beginning, but I have never highlighted it nor even explicitly presented the details previously.

2. $\sin\phi(A)$ as a general optimum

Let us stay with the simple case of A a real symmetric positive definite $n \times n$ matrix. The results of course hold as well for Hermitian positive definite $n \times n$ matrices.

Theorem 2 (General Optimum Theorem). *Let $x = c_1x_1 + c_nx_n$ where x_1 is any norm-one eigenvector corresponding to λ_1 and x_n is any norm-one eigenvector corresponding to λ_n such that $|c_1|^2 + |c_n|^2 = 1$. Then*

$$\sin\phi(A) = \|(\epsilon_m A - I)x\|, \qquad \epsilon_m = 2/(\lambda_1 + \lambda_n).$$

In contrast,

$$\cos\phi(A) = \min_{x\neq 0} \frac{\langle Ax, x\rangle}{\|Ax\|\|x\|}$$

is attained only for the special weights

$$|c_1|^2 = \lambda_n/(\lambda_1 + \lambda_n) \quad \text{and} \quad |c_n|^2 = \lambda_1/(\lambda_1 + \lambda_n),$$

i.e., only for the antieigenvectors $x^{\pm}$.

Proof. The proof may be obtained from a careful examination of the vectors x of norm one which "work" in the following two important relations:

$$\begin{aligned}\max\nolimits_{\|x\|=1} \min\nolimits_{\epsilon>0} \|(\epsilon A - I)x\|^2 &= \max\nolimits_{\|x\|=1} \left[1 - \left(\tfrac{\mathrm{Re}\langle Ax,x\rangle}{\|Ax\|}\right)^2\right] \\ &= 1 - \min\nolimits_{\|x\|=1} \left(\tfrac{\mathrm{Re}\langle Ax,x\rangle}{\|Ax\|}\right)^2 \\ &= 1 - \cos^2\phi(A),\end{aligned}$$

and

$$\begin{aligned}\min\nolimits_{\epsilon>0} \max\nolimits_{\|x\|=1} \|(\epsilon A - I)x\|^2 &= \min\nolimits_{\epsilon>0} \|\epsilon A - I\|^2 \\ &= \sin^2\phi(A).\end{aligned}$$

The details may be checked by the reader. □

Theorem 2 (General Optimum Theorem) highlights how much more generally the $\sin\phi(A)$ functional is attained, as compared to the $\cos(A)$ functional. One only needs to be on the unit ball in the combined first and last eigenspaces. Then I view the obtaining of the $\cos\phi(A)$ as a special selection from within that general two-component unit ball.

Here is another example. In the papers [8], [9], [10], and [12], where I applied my operator trigonometry to matrix statistics, I showed that the Watson statistical efficiency

$$\mathrm{eff}(\widehat{B})^{-1} = \prod_{i=1}^{p} x_i' V x_i x_i' V^{-1} x_i$$

is optimized at what I called the "inefficiency vectors"

$$x_{\pm}^{j+k} = \pm\frac{1}{\sqrt{2}}x_j + \frac{1}{\sqrt{2}}x_k,$$

(see [12, Theorem 3.1] for details). These are all special choices of weights on the indicated subspace two-component unit balls. As I emphasized in [12], these must be distinguished from the higher antieigenvectors which are formed from different weights within those two-component unit balls.

Another example is the Shisha–Mond optimum

$$\max_{\|x\|=1}[\langle Ax, x\rangle - \langle A^{-1}x, x\rangle^{-1}] = (\lambda_1^{1/2} - \lambda_n^{1/2})^2.$$

If we let $x = c_1x_1 + c_nx_n$ with $|c_1|^2 + |c_n|^2 = 1$ and x_1 and x_n any norm-one first and last eigenvectors, we find

$$\langle Ax, x\rangle - \langle A^{-1}x, x\rangle^{-1} = c_1^2\lambda_1 + c_n^2\lambda_n - \frac{1}{c_1^2\lambda_1^{-1} + c_n^2\lambda_n^{-1}}.$$

Somewhat lengthy calculations show this to be maximized at

$$c_1 = \left(\frac{\lambda_1^{1/2}}{\lambda_1^{1/2} + \lambda_n^{1/2}}\right)^{1/2} \quad \text{and} \quad c_n = \left(\frac{\lambda_n^{1/2}}{\lambda_1^{1/2} + \lambda_n^{1/2}}\right)^{1/2}.$$

This represents yet another special choice of weights from within the $\sin\phi(A)$ general optimizing two component x_1, x_n unit ball.

Many other inequalities in the matrix statistics literature will be optimized by special weight choices within the unit ball all of which attain $\sin\phi(A)$. Perhaps in a later paper I may work out a more comprehensive theory.

I would like to close this paper with another consequence of my Min-Max Theorem that has not been highlighted previously.

3. Antieigenvectors from the Min-Max Theorem

As shown above, the weights needed for the antieigenvectors are special choices which minimize the cosine functional

$$\cos\phi(A) = \min \frac{\langle Ax, x\rangle}{\|Ax\|\|x\|}.$$

In my early work [6] (see [7] and [14]) I differentiated this expression to find the Euler equation for the antieigenvectors. Thus the general viewpoint is that you need the cosine functional to determine the antieigenvectors. In other words, my thinking was just analogous to that of Rayleigh–Ritz eigenvalue-eigenvector theory. From the variational characterization

$$\lambda_n = \min_{x\neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$

you obtain by differentiation the Euler equation

$$Ax_n = \lambda_n x_n$$

and similarly for the higher eigenvalue-eigenvector pairs.

However, a close look at my Min-Max Theorem proof reveals that I actually used to prove it a construction of approximate antieigenvectors to attain or at least approximate the convex minimum $\sin\phi(A)$. Therefore we may state

Theorem 3 (Min-Max Antieigenvector Construction Theorem)**.** *You may even get A's antieigenvectors just from the* $\sin\phi(A) = \|\epsilon_m A - I\|$ *convex minimum and the Min-Max Theorem proof.*

Proof. I refer the reader to the proof of the Min-Max Theorem given in the book [14], pp. 53–55. Using the notation there, we take $x = \xi x_1 + \eta x_n$ and find the expression (3.2-8) of [14] for the attainment of the $\sin\phi(A)$ minimum. The proof in [14] is for the more general A strongly accretive situation and I don't want to delineate those details here. However, for A a real $n \times n$ symmetric positive definite matrix, the ϵ_1 and ϵ_2 in (3.2-8) of [14] become $\epsilon_1 = 1/\lambda_n$ and $\epsilon_n = 1/\lambda_1$ and that expression simplifies to

$$\xi^2\{\lambda_n(1 - 2\lambda_n/(\lambda_1 + \lambda_n))\} + \eta^2\{\lambda_1(1 - 2\lambda_1/(\lambda_1 + \lambda_n))\} = 0$$

or seen more trigonometrically, to the expression

$$\sin^2\phi(A)[\lambda_n\xi^2 - \lambda_1\eta^2] = 0.$$

The expression $\|(\epsilon_m A - I)x\|^2$ simplifies to

$$\sin^2\phi(A)[\xi^2 + \eta^2].$$

Thus one has the 2×2 system to solve:

$$\left\{\begin{array}{ll} \xi^2 + \eta^2 & = 1 \\ \lambda_1\xi^2 + \lambda_n\eta^2 & = 0 \end{array}\right\}$$

from which

$$\xi = \pm\left(\frac{\lambda_1}{\lambda_1 + \lambda_n}\right)^{1/2} \quad \text{and} \quad \eta = \pm\left(\frac{\lambda_n}{\lambda_1 + \lambda_n}\right)^{1/2}.$$

These indeed happen to be the weights of the antieigenvectors. The additional fact used beyond wanting to be at the $\sin^2\phi(A)$ convex minimum was that in proving the Min-Max Theorem I used a construction that brings you to that minimum by combining a parabolic curve $\|(\epsilon A - I)x_1\|^2$ from the left which achieves its minimum at $\epsilon_1 = \mathrm{Re}\langle Ax_1, x_1\rangle/\|Ax_1\|^2$ and a parabolic curve $\|(\epsilon A - I)x_n\|^2$ from the right which achieves its minimum at $\epsilon_n = \mathrm{Re}\langle Ax_n, x_n\rangle/\|Ax_n\|^2$. The precise way in which you then combine these left and right curves corresponds to constructing an exact or approximate antieigenvector. $\square$

To make this point very clear, let us consider the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Let us plot $\|\epsilon A - I\|$ for $\epsilon \geq 0$. On the left it is the line

$$\ell_1 : \|(\epsilon A - I)x_n\| = 1 - \epsilon\lambda_n, \quad 0 \leq \epsilon \leq \frac{2}{\lambda_1 + \lambda_n}.$$

On the right it is

$$\ell_2 : \|(\epsilon A - I)x_1\| = -1 + \epsilon\lambda_1, \quad \epsilon \geq \frac{2}{\lambda_1 + \lambda_n}.$$

Thus for the given matrix A these are the lines

$$\begin{aligned} \ell_1 &: 1-\epsilon, && 1<\epsilon\le 2/3 \\ \ell_2 &: -1+2\epsilon, && \epsilon\ge 2/3. \end{aligned}$$

The intersection of those lines at $\epsilon = 2/3$ is $\sin\phi(A) = 1/3$. So the norm curve is a "one component" line to the left, a "one component line" to the right, and when you take the ξ and η weights above, you have a "two-component" antieigenvector whose parabolic curve $\|(\epsilon A - I)x\|^2$ lies strictly below the envelope $\|\epsilon A - I\|^2$ but which hits that envelope at its minimum $\|\epsilon_m A - I\|^2 = \sin^2\phi(A)$ exactly at $\epsilon_m = 2/(\lambda_1+\lambda_n)$.

References

[1] K. Gustafson, *A min-max theorem*, Notices Amer. Math. Soc. **15** (1968), 799.

[2] K. Gustafson, *A note on left multiplication of semigroup generators*, Pacific J. Math. **24** (1968), 463–465.

[3] K. Gustafson, *Positive (noncommuting) operator products and semigroups*, Math. Z. **105** (1968), 160–172.

[4] K. Gustafson, *The angle of an operator and positive operator products*, Bull. Amer. Math. Soc. **74** (1968), 488–492.

[5] K. Gustafson, *The Toeplitz-Hausdorff theorem for linear operators*, Proc. Amer. Math. Soc. **25** (1970), 203–204.

[6] K. Gustafson, Antieigenvalue inequalities in operator theory; In: O. Shisha (Ed.) *Inequalities III*, Academic Press, New York, 1972, pp. 115–119.

[7] K. Gustafson, *Lectures on Computational Fluid Dynamics, Mathematical Physics, and Linear Algebra*, World-Scientific, Singapore, 1997.

[8] K. Gustafson, *On geometry of statistical efficiency* (1999), preprint.

[9] K. Gustafson, *Operator trigonometry of statistics and econometrics*, Linear Algebra Appl. **354** (2002), 141–158.

[10] K. Gustafson, *The geometry of statistical efficiency*, Res. Lett. Inform. Math. Sci. **8** (2005), 105–121.

[11] K. Gustafson, *Noncommutative trigonometry*, Oper. Theory Adv. Appl. **167** (2006), 127–155.

[12] K. Gustafson, *The trigonometry of matrix statistics*, Internat. Statist. Rev. **74** (2006), 187–202.

[13] K. Gustafson, *Operator trigonometry of Hotelling correlation, Frobenius condition, Penrose twistor*, Linear Algebra Appl. **430** (2009), 2762–2770.

[14] K. Gustafson and D. K. M. Rao, *Numerical Range*, Springer, Berlin, 1997.

Department of Mathematics, University of Colorado, Boulder, CO 80309-0395, USA

E-mail address: gustafs@euclid.colorado.edu