Is there a Small Skew Cayley Transform with Zero Diagonal ?

Abstract

The eigenvectors of an Hermitian matrix H are the columns of some complex unitary matrix Q . For any diagonal unitary matrix Ω the columns of Q·Ω are eigenvectors too. Among all such Q·Ω at least one has a skew-Hermitian Cayley transform $S := (I+Q\cdot\Omega)^{-1}\cdot(I-Q\cdot\Omega)$ with just zeros on its diagonal. Why? The proof is unobvious, as is the further observation that Ω may also be so chosen that no element of this S need exceed 1 in magnitude. Thus, plausible constraints, easy to satisfy by perturbations of complex eigenvectors when an Hermitian matrix H is perturbed infinitesimally, can be satisfied for discrete perturbations too. But if H is real symmetric, Q real orthogonal and Ω restricted to diagonals of ±1's, then whether at least one real skew-symmetric S must have no element bigger than 1 in magnitude is not known yet.

Full text posted at `http://www.cs.berkeley.edu/~wkahan/SkCayley.pdf`

Hermitian Eigenproblem

Hermitian Matrix $H = H^H = \overline{H}^T$ — complex conjugate transpose.

Real Eigenvalues $v_1 \le v_2 \le v_3 \le \ldots \le v_n$ sorted and put into a column vector

$$v := [\, v_1, v_2, v_3, \ldots, v_n \,]^T$$

Corresponding eigenvector columns $q_1, q_2, q_3, \ldots, q_n$ need not be determined uniquely but can always be chosen to constitute columns of a *Unitary* matrix Q satisfying

$$H\cdot Q = Q\cdot \mathrm{Diag}(v) \quad \text{and} \quad Q^H = Q^{-1} .$$

$Q\cdot\Omega$ is also an eigenvector matrix for every unitary diagonal matrix $\Omega = \overline{\Omega}^{-1}$.

Familiar special case: Real symmetric $H = H^T$, real orthogonal $Q = Q^{-1\,T}$.

$Q\cdot\Omega$ is also an eigenvector matrix for every diagonal matrix $\Omega = \mathrm{Diag}(\pm 1\text{'s})$.

Perturbed Hermitian Eigenproblem

Given Hermitian Matrix $H = H_o + \Delta H$ for small $\|\Delta H\|$.

Suppose H_o has known eigenvalue column v_o and eigenvector matrix Q_o .

Then eigenvalue column v of H must be close to v_o : $\|v - v_o\|_\infty \le \|\Delta H\|$.

But no eigenvector matrix Q of H need be near Q_o unless $\|\Delta H\|$ is rather smaller than gaps between adjacent eigenvalues v_j of H , or of H_o .

Cautionary Examples: For every tiny nonzero θ , no matter how tiny,

$H = \begin{bmatrix} 1+\theta & 0 \\ 0 & 1-\theta \end{bmatrix}$ has eigenvectors rotated through $\pi/2$ from $H_o = \begin{bmatrix} 1-\theta & 0 \\ 0 & 1+\theta \end{bmatrix}$.

$H = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$ has eigenvectors rotated through $\pi/4$ from $H_o = \begin{bmatrix} 1-\theta & 0 \\ 0 & 1+\theta \end{bmatrix}$.

See Parlett’s book and papers by C. Davis & W. Kahan, and by Paige & Wei, on rotations of eigenspaces.

Still, how are tiny perturbations of eigenvector matrices to be represented?

Infinitesimally Perturbed Unitary Matrices

Say $Q = Q^{-1\,H}$ is perturbed to $Q + dQ = (Q + dQ)^{-1\,H}$; then

$$O = dI = d(Q^H \cdot Q) = dQ^H \cdot Q + Q^H \cdot dQ \,, \quad \text{so}$$

$$dQ = -2Q \cdot dS \quad \text{for some } \textit{Skew-Hermitian} \quad dS = -dS^H \,, \quad \text{and}$$

$$Q + dQ = Q \cdot (I - 2dS) \,.$$

This is what brings skew-Hermitian matrices to our attention.

Discretely Perturbed Unitary Matrices

Say $Q = Q^{-1\,H}$ is perturbed to a nearby $Q + \Delta Q = (Q + \Delta Q)^{-1\,H}$; then

either $Q + \Delta Q = Q \cdot e^{-2\Delta Z}$ for some small skew-Hermitian ΔZ ,

or $Q + \Delta Q = \underset{\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim}{Q \cdot (I + \Delta S)^{-1} \cdot (I - \Delta S)}$ for a small skew-Hermitian ΔS .

This is what brings the *Cayley Transform* to our attention.

What is a Cayley Transform $\$(z)$?

$\$(z)$ is an analytic function of a complex variable z on the Riemann sphere,

Closed by one point at ∞ .

1) It is *Involutary* : $\$(\$(z)) = z$. ••• so \$ must be *Bilinear Rational*.

2) It swaps *Invert* ↔ *Negate* : $\$(-z) = 1/\$(z)$ and so $\$(1/z) = -\$(z)$.

Inference: Only two choices for $\$(z)$, $\frac{1-z}{1+z}$ or $\frac{z+1}{z-1}$. Our choice is

$$\$(z) := \frac{1-z}{1+z} , \qquad \text{chosen so that } \$(0) = 1 .$$

\$ maps …

Real Axis ↔ Real Axis , Imaginary Axis ↔ Unit Circle ,

Right Half-Plane ↔ Unit Disk ,

Real Orthogonal Matrix $Q = Q^{-1\,T}$ ↔ Real Skew-Symmetric $S = -S^T$,

Complex Unitary Matrix $Q = Q^{-1\,H}$ ↔ Complex Skew-Hermitian $S = -S^H$.

Evading the Cayley Transform's Pole

$$\$(B) := (I + B)^{-1}\cdot(I - B)$$

Unitary $Q = \$(S) = Q^{-1\,H}$ ↔ Skew-Hermitian $S = \$(Q) = -S^H$

provided

$\det(I + Q) \neq 0$ ↔ S is finite.

Every unitary Q has eigenvalues all with magnitude 1 ; but
no Cayley transform $Q = \$(S)$ can have -1 as an eigenvalue.

Will this exclude any eigenvectors ? **No** :

Lemma: If Q is unitary and if $I+Q$ is singular, then reversing signs of aptly chosen columns of Q will make $I+Q$ nonsingular and provide a finite skew Cayley transform $S = \$(Q)$.

Proof: Any of many simple computations. The earliest I know appeared in 1960 ; see Exs. 7 - 11, pp. 92-3 in §4 of Ch. 6 of Richard Bellman's book *Introduction to Matrix Analysis* (2d. ed. 1970, McGraw-Hill). Or see pp. 2-3 of `…~wkahan/SkCayley.pdf` .

Henceforth take $\det(I + Q) \neq 0$ for granted.

Back to Perturbed Hermitian Eigenproblem

Given Hermitian Matrix $H = H_o + \Delta H$ for small $||\Delta H||$.

Suppose H_o has known eigenvalue column v_o and eigenvector matrix Q_o .

W.L.O.G, exposition is simplified by taking the eigenvectors of H_o as a new orthonormal coordinate system, so that $H_o = \mathrm{Diag}(v_o)$. Now we wish to solve

$$H\cdot Q = Q\cdot \mathrm{Diag}(v) \quad \text{and} \quad Q^H\cdot Q = I \qquad (\dagger)$$

for a sorted eigenvalue column v near v_o , and a unitary Q not far from I .

Substituting $Q = \$(S)$ into $(\dagger)$ transforms it into a slightly less nonlinear

$$(I+S)\cdot H\cdot(I-S) = (I-S)\cdot \mathrm{Diag}(v)\cdot(I+S) \quad \text{and} \quad S^H = -S \qquad (\ddagger)$$

If all $h_{jk}/(h_{jj} - h_{kk})$ for $j \neq k$ are so small that 3rd-order $S\cdot(H - \mathrm{Diag}(H))\cdot S$ will be negligible, then equations $(\ddagger)$ have simple approximate solutions

$$v \approx \mathrm{Diag}(H) \quad \text{and} \quad s_{jk} \approx \tfrac{1}{2} h_{jk}/(h_{jj} - h_{kk}) \quad \text{for} \quad j \neq k .$$

Diagonal elements s_{jj} can be arbitrary imaginaries but small lest 3rd-order terms be not negligible. Forcing $s_{jj} := 0$ seems plausible. But if done when off-diagonal elements are too big to yield acceptable simple approximations to v and S , can $(\ddagger)$ still be solved for v and small S with $\mathrm{diag}(S) = o$?

Do a sorted eigenvalue column v and a skew S both satisfying

$$(I+S)\cdot H\cdot(I-S) = (I-S)\cdot \mathrm{Diag}(v)\cdot(I+S) \quad \text{and} \quad S^H = -S \qquad (\ddagger)$$

always exist with $\mathrm{diag}(S) = o$ and S not too big? If so, then $Q := \$(S)$.

Why might we wish to compute v and S , and then Q ?

Iterative Refinement.

The usual way to enhance the accuracy of solutions v and Q of

$$H\cdot Q = Q\cdot \mathrm{Diag}(v) \quad \text{and} \quad Q^H\cdot Q = I \qquad (\dagger)$$

when H is almost diagonal is *Jacobi Iteration*. It converges quadratically if programmed in a straightforward way, cubically if programmed in a tricky way made doubly tricky if available parallelism is to be exploited too.
See its treatment in Golub & Van Loan’s book, and recent papers by Drmac̄ & Veselic̄.

If the simple solution of (‡) is adequate, it converges cubically and is easy to parallelize. Sometimes the simple solution is inadequate, and then we seek a better solution of (‡) by some slightly more complicated method. S should not be too big lest Cayley transform $Q := (I+S)^{-1}\cdot(I-S)$ be too inaccurate.

Thus is the question at the top of this page motivated.

Do a sorted eigenvalue column v and a skew S both satisfying

$$(I+S)\cdot H\cdot(I-S) = (I-S)\cdot \mathrm{Diag}(v)\cdot(I+S) \quad \text{and} \quad S^H = -S \qquad (\ddagger)$$

always exist with diag(S) = o and S not too big?

YES in the Complex Case, when S can be complex skew-Hermitian. And then at least one such S has diag(S) = o and all $|s_{jk}| \le 1$.

This *Existence Theorem* is proved. How best to find that S is not yet known.

In the Real Case, when a real $H = H^T$ entails a real skew-symmetric $S = -S^T$, every diag(S) = o ; but whether some such S has all $|s_{jk}| \le 1$ too is not yet known, though it seems likely.

What follows will be first some examples, and then an outline of the Existence Theorem's proof.

In what follows, one of the unitary or real orthogonal eigenvector matrices of H is G , and all other eigenvector matrices $Q := G\cdot\Omega$ of H are generated by letting diagonal matrix Ω runs through all …

- … diagonal unitary matrices $\Omega = e^{\imath\,\mathrm{Diag}(x)}$ with real columns x , or
- … real diagonals $\Omega = \mathrm{Diag}([\pm1, \pm1, \pm1, \dots, \pm1])$ in the Real Case.

A 3-by-3 Example

Real orthogonal $G := \$\left(\begin{bmatrix} 0 & 2 & -2 \\ -2 & 0 & 2 \\ 2 & -2 & 0 \end{bmatrix}\right) = \begin{bmatrix} -3 & 4 & 12 \\ 12 & -3 & 4 \\ 4 & 12 & -3 \end{bmatrix}/13$. $Q := G\cdot\Omega$; $S := \$(Q)$.

diag(S) = o for six diagonal matrices Ω . Four of them are real, namely $\Omega := I$, Diag([–1, –1, 1]) , Diag([1, –1, –1]) , and Diag([–1, 1, –1]) .

Typical of the last three is $\$(G\cdot\mathrm{Diag}([\,-1, 1, -1\,])) = \begin{bmatrix} 0 & -1 & \frac{1}{2} \\ 1 & 0 & 1 \\ -\frac{1}{2} & -1 & 0 \end{bmatrix}$. $\|\ldots\| = 3/2$.

The two complex unitary diagonals Ω are scalars $\Omega := (-5 \pm 12\imath)\cdot I/13$.

For them $\$(G\cdot\Omega) = \begin{bmatrix} 0 & -1-3\imath & 1-3\imath \\ 1-3\imath & 0 & -1-3\imath \\ -1-3\imath & 1-3\imath & 0 \end{bmatrix}/4$ and its complex conjugate resp.

Note that its every element is strictly smaller than 1 in magnitude though still $\|\ldots\| = 3/2$. Allowing Q and S to be complex lets S have smaller elements.

An Extreme n-by-n Example

Real orthogonal $G :=$

$$\begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & \cdots & \cdots & & & \\ & & & \cdots & \cdots & & \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ (-1)^{n-1} & & & & & & 0 \end{bmatrix} .$$

Let Ω run through all unitary diagonals with $\det(\Omega) \neq -1$. These include 2^{n-1} real orthogonal diagonals of ±1 ’s with an even number of –1 ’s . Every such Ω has $|\det(\Omega)| = 1$. Every $Q := G \cdot \Omega$ is unitary; 2^{n-1} are orthogonal.

For all such Ω , every off-diagonal element of $S := \$(Q)$ has magnitude $2/|1 + \det(\Omega)|$. It is minimized when $\det(\Omega) = +1$; only then is $\mathrm{diag}(S) = o$, and then every off-diagonal element of S has magnitude 1 . This happens in all the Real Cases.

For a detailed explanation see `…/~wkahan/SkCayley.pdf` .

The Existence Theorem

Given a unitary matrix G

(of eigenvectors of an Hermitian matrix H)

let Ω run through unitary diagonal matrices, so $Q := G\cdot\Omega$ is unitary too,

(also a matrix of eigenvectors of that Hermitian matrix H)

and let $S := \$(Q)$ be the skew-Hermitian Cayley transform of $Q = \$(S)$.

Then $\mathrm{diag}(S) = o$ for at least one such S , and its every element has $|s_{jk}| \le 1$.

Proof:

Among all such $Q = G\cdot\Omega$ the one(s) “nearest” the identity I , in a peculiar sense defined hereunder, must turn out to have the desired kind of $S = \$(Q)$.

The peculiar gauge of “nearness” of a unitary Q to I is

$$£(Q) := -\log(\det\big((I+Q^H)\cdot(I+Q)/4\big)) = \log(\det\big(\, I + \$(Q)^H\cdot\$(Q)\,\big)) .$$

$£(Q) > 0$ for every unitary Q except $£(I) = 0$ and

$£(Q) = +\infty$ when $\det(I+Q) = 0$.

What remains of the proof is a characterization of every unitary $Q = G\cdot\Omega$ that minimizes $£(Q)$. For this we need the first two derivatives of $£$.

How to Derive Derivatives, with respect to a real column-vector x , of

$$£(Q) := -\log(\det\big((I+Q^H)\cdot(I+Q)/4\big)) \;=\; \log(\det\big(\,-\log(\det\big((2I+Q^{-1}+Q)/4\big))$$

when unitary $Q = G\cdot\Omega = G\cdot e^{\boldsymbol{\imath}\,\mathrm{Diag}(x)}$.

We shall abbreviate Diag(x) =: X , and then the *Differential* dX := Diag(dx) .

Tools:

$\Omega = e^{\boldsymbol{\imath} X}$ has $d\Omega = de^{\boldsymbol{\imath} X} = \Omega\cdot e^{\boldsymbol{\imath}\, dX}$ since diagonals dX and X commute.

$d(B^{-1}) = -B^{-1}\cdot dB\cdot B^{-1}$.

Jacobi’s formula $d\log(\det(B)) = \mathrm{trace}(B^{-1}\cdot dB)$.

For a derivation see `…/~wkahan/MathH110/jacobi.pdf` .

$\mathrm{trace}(B\cdot C) = \mathrm{trace}(C\cdot B)$.

Using these tools we find first that $d\,£(B) = \mathrm{trace}(\$(B)\cdot B^{-1}\cdot dB)$ in general,

and then that $d\,£(G\cdot\Omega) = \boldsymbol{\imath}\,\mathrm{diag}(\$(G\cdot\Omega))^T dx$, so

$$\partial £(G\cdot\Omega)/\partial x = \boldsymbol{\imath}\,\mathrm{diag}(\$(G\cdot\Omega))^T \;=\; \boldsymbol{\imath}\,\mathrm{diag}(S)^T \;.$$

This must vanish at the minimum (and any other extremum) of $£(G\cdot\Omega)$, so at least one Ω makes $S := \$(G\cdot\Omega)$ have $\mathrm{diag}(S) = o$, as claimed.

The second derivative of $£(G\cdot e^{\imath\,\mathrm{Diag}(x)})$ is representable by a symmetric *Hessian* matrix M of second partial derivatives that figures in

$$\big(\partial^2 £(G\cdot\Omega)/\partial x^2\big)\cdot\Delta x\cdot dx \;=\; dx^T\cdot M\cdot\Delta x\ .$$

For any fixed Δx a lengthy computation of

$$\big(\partial^2 £(G\cdot\Omega)/\partial x^2\big)\cdot\Delta x\cdot dx = d\big(\partial £(G\cdot\Omega)/\partial x\big)\cdot\Delta x = \; d\big(\imath\,\mathrm{diag}(\$(G\cdot\Omega))^T\big)\cdot\Delta x \;=\; \dots$$

yields Hessian $M = (I + |S|^2)/2$ in which $S = \$(G\cdot\Omega)$ and $|S|^2$ is obtained from S elementwise by substituting $|s_{jk}|^2$ for every element s_{jk} .

At the minimum of $£(G\cdot e^{\imath\,\mathrm{Diag}(x)})$ its Hessian $M = (I + |S|^2)/2$ must be positive (semi)definite, and this implies that every $|s_{jk}|^2 \le 1$ since $\mathrm{diag}(S) = o$. Thus is the Existence Theorem's second claim confirmed. And the extreme n-by-n example shows that the upper bound 1 is achievable. END of proof.

• • •

No such proof can work in the Real Case when H is real symmetric, its eigenvector matrix G is real, and Ω is restricted to real orthogonal diagonals. These constitute a discrete set, not a continuum, so derivatives don't matter.

Conclusion:

Perturbing a complex Hermitian matrix H changes its unitary matrix Q of eigenvectors to a perturbed unitary $Q\cdot(I+S)^{-1}\cdot(I-S)$ in which the skew-Hermitian $S = -S^H$ can always be chosen to be small (no element bigger than 1 in magnitude) and to have only zeros on its diagonal. But how to construct this S efficiently and infallibly is not known yet. Neither is it known yet, when H is real symmetric and Q is real orthogonal and S is restricted to be real skew-symmetric, whether S can always be chosen to have no element bigger in magnitude than 1 .