Is there a Small Skew Cayley Transform with Zero Diagonal ?

§0: 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 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.

Contents Page

§1: Introduction

After Cayley transforms $\$(B) := (I+B)^{-1}\cdot(I–B)$ have been described in §2, a transform with only zeros on its diagonal will be shown to exist because it solves this minimization problem:

Among unitary matrices Q·Ω with a fixed unitary Q and variable unitary diagonal Ω , those matrices Q·Ω "nearest" the identity I in a sense defined in §3 have skew-Hermitian Cayley transforms $S := \$(Q\cdot\Omega) = –S^H$ with zero diagonals and with no element s_{jk} bigger than 1 in magnitude.

Now, why might this interest us? It's a long story … .

Let H be an Hermitian matrix (so $H^H = H$) whose eigenvalues are ordered monotonically (this is crucial) and put into a real column vector v , and whose corresponding eigenvectors can then be chosen to constitute the columns of some unitary matrix Q satisfying the equations

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

(Notational note: We distinguish diagonal matrices Diag(A) and V = Diag(v) from column vectors diag(A) and v = diag(V) , unlike MATLAB whose `diag(diag(A))` is our Diag(A) .

We also distinguish scalar 0 from zero vectors o and zero matrices O . And $Q^H = \overline{Q}^T$ is the complex conjugate transpose of Q ; and $\imath = \sqrt{-1}$; and all identity matrices are called " I ". The word "skew" serves to abbreviate either "skew-Hermitian" or "real skew-symmetric".)

If Q and v are not known yet but H is very near an Hermitian H_o with known eigenvalue-column v_o (also ordered monotonically) and eigenvector matrix Q_o then, as is well known, v must lie very near v_o . This helps us find v during perturbation analyses or curve tracing or iterative refinement. However, two complications can push Q far from Q_o . First, (†) above does not determine Q uniquely: Replacing Q by $Q\cdot\Omega$ for any unitary diagonal Ω leaves the equations still satisfied. To attenuate this first complication we shall seek a $Q\cdot\Omega$ "nearest" Q_o . Still, no $Q\cdot\Omega$ need be very near Q_o unless gaps between adjacent eigenvalues in v and also in v_o are all rather bigger than $\|H-H_o\|$; this second complication is unavoidable for reasons exposed by examples so simple as $H = \begin{bmatrix} 1+\theta & 0 \\ 0 & 1-\theta \end{bmatrix}$ and $H_o = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}$ with tiny θ and ϕ .

To simplify our exposition we assume $Q_o = I$ with no loss of generality; doing so amounts to choosing the columns of Q_o as a new orthonormal basis turning H_o into $\mathrm{Diag}(v_o)$. Now we can seek solutions Q and v of (†) above with v ordered and Q "nearest" I in some sense.

§2: The Cayley Transform $\$(B) := (I+B)^{-1}\cdot(I-B) = (I-B)\cdot(I+B)^{-1}$

On its domain it is an *Involution*: $\$(\$(B)) = B$. However $\$(-\$(B)) = B^{-1}$ if it exists. \$ maps certain unitary matrices Q to skew matrices S (real if Q is real orthogonal) and back thus:

If I+Q is nonsingular the Cayley transform of unitary $Q = Q^{-1\,H}$ is skew $S := \$(Q) = -S^H$;
and then the Cayley transform of skew $S = -S^H$ recovers unitary $Q = \$(S) = Q^{-1\,H}$.

Thus, given an algebraic equation like (†) to solve for Q subject to a nonlinear side-condition like $Q^H = Q^{-1}$, we can solve instead an equivalent algebraic equation for S subject to a near-linear and thus simpler side-condition $S = -S^H$, though doing so risks losing some solution(s) Q for which I+Q is singular and the Cayley transform S is infinite. But no eigenvectors need be lost that way. Instead their unitary matrix Q can appear post–multiplied harmlessly by a diagonal matrix whose diagonal elements are each either +1 or –1 . Here is why: …

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 Cayley transform $S = \$(Q)$.

Proof: I am grateful to Prof. Jean Gallier for pointing out that Richard Bellman published this lemma in 1960 as an exercise; see Exs. 7 - 11, pp. 92-3 in §4 of Ch. 6 of his book *Introduction to Matrix Analysis* (2d ed. 1970 McGraw-Hill, New York). The non-constructive proof hereunder is utterly different. Let n be the dimension of Q , let $m := 2^n - 1$, and for each $k = 0, 1, 2, ..., m$ obtain n-by-n unitary Q_k by reversing the signs of whichever columns of Q have the same positions as have the nonzero bits in the binary representation of k . For example $Q_0 = Q$, $Q_m = -Q$, and Q_1 is obtained by reversing the sign of just the last column of Q . Were the lemma false we would find every $\det(I+Q_k) = 0$. For argument's sake let us suppose all 2^n of these equations to be satisfied.

Recall that det(...) is a linear function of each column separately; whenever n-by-n B and C differ in only one column, $\det(B+C) = 2^{n-1}\cdot(\det(B) + \det(C))$. Therefore our supposition would imply $\det(I+Q_{2i} + I+Q_{2i+1}) = 2^{n-1}\cdot(\det(I+Q_{2i}) + \det(I+Q_{2i+1})) = 0$ whenever $0 \le i \le (m-1)/2$. Similarly $\det((I+Q_{4j} + I+Q_{4j+1}) + (I+Q_{4j+2} + I+Q_{4j+3})) = 0$ whenever $0 \le j \le (m-3)/4$. And so on. Ultimately $\det(I+Q_0 + I+Q_1 + I+Q_2 + ... + I+Q_m) = 0$ would be inferred though the sum amounts to $2^n\cdot I$, whose determinant cannot vanish! This contradiction ends the lemma's proof.

The lemma lets us replace any search for a unitary or real orthogonal matrix Q of eigenvectors by a search for a skew matrix S from which a Cayley transform will recover one of the sought eigenvector matrices $Q := (I+S)^{-1}\cdot(I-S)$. Constraining the search to skew-Hermitian S with diag(S) = o is justified in §3. A further constraint keeping every $|s_{jk}| \le 1$ to render Q easy to compute accurately is justified in §5 for complex S , but maybe not if Q and S must be real.

Substituting Cayley transform Q = \$(S) into (†) turns them into equations more nearly linear:

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

If all off-diagonal elements h_{jk} of H are so tiny compared with differences $h_{jj} - h_{kk}$ between diagonal elements that 3rd-order terms $S\cdot(H-\mathrm{Diag}(H))\cdot S$ can be neglected, equations (‡) have approximate solutions $v \approx \mathrm{diag}(H)$ and $s_{jk} \approx \frac{1}{2} h_{jk}/(h_{jj} - h_{kk})$ for $j \ne k$. 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, as happens more often, off-diagonal elements are too big for the foregoing approximations for v and S to be acceptable, how do we know equations (‡) must still have at least one solution v and S with diag(S) = o and no huge elements in S ?

Now the question that is this work's title has been motivated: Every unitary matrix G of H 's eigenvectors spawns an infinitude of solutions $Q := G\cdot\Omega$ of (†) whose skew-Hermitian Cayley transforms $S := \$(G\cdot\Omega)$ satisfying (‡) sweep out a continuum as Ω runs through all complex unitary diagonal matrices for which $I+G\cdot\Omega$ is nonsingular. This continuum happens to include at least one skew S with diag(S) = o and no huge elements, as we'll see in §3 and §5.

Lacking this continuum, an ostensibly simpler special case turns out not so simple: When H is real symmetric and G is real orthogonal then, whenever Ω is a real diagonal of –1's and/or +1's for which the Cayley transform $\$(G\cdot\Omega)$ exists, it is a real skew matrix with zeros on its diagonal. The Lemma above ensures that some such $\$(G\cdot\Omega)$ exists. Still unknown is whether at least one such $\$(G\cdot\Omega)$ has no element bigger than 1 in magnitude, though it seems likely despite §4's examples on the brink: They are n-by-n real orthogonal matrices G for which *every* off-diagonal element of *every* (there are 2^{n-1} of them) such $\$(G\cdot\Omega)$ is ±1 .

The continuum swept out in the complex case helps us answer our questions. For any given real or complex unitary G , as Ω ranges through all complex unitary diagonal matrices for which $I+G\cdot\Omega$ is nonsingular, the unitary $G\cdot\Omega$ that comes nearest the identity matrix I in a peculiar sense to be explained forthwith has a Cayley transform $\$(G\cdot\Omega)$ with only zeros on its diagonal and no element bigger than 1 in magnitude.

§3: £(Q) Gauges How "Near" a Unitary Q is to I

The function £(B) := –log(det((2I + B + B^{-1})/4)) = –log(det((I+B^{-1})·(I+B)/4)) will be used to gauge how "near" any unitary matrix Q = $Q^{-1\,H}$ is to I . The closer is £(Q) to 0 , the "nearer" shall Q be deemed to I . The following digression explores properties of £(Q) :

When (I+Q) is nonsingular, every eigenvalue of unitary Q has magnitude 1 but none is –1 , so matrix (2I + Q + Q^{-1})/4 = $(I+Q)^H$·(I+Q)/4 is Hermitian with real eigenvalues all positive and no bigger than 1 . Therefore its determinant, their product, is also positive and no bigger than 1 ; therefore £(Q) ≥ 0 . Only £(I) = 0 . Another way to confirm this is to observe that £(Q) = log(det(I – $(Q)^2)) = log(det(I + $(Q)^H·$(Q))) > 0 (or +∞) for every unitary Q ≠ I .

£(Q) and $(Q) are differentiable functions of Q except at their poles, where $(Q) is infinite and £(Q) = +∞ because det(I+Q) = 0 . The differential of £(Q) is simpler to derive than its derivative is because of Jacobi's formula d log(det(B)) = trace(B^{-1}·dB) and another formula d(B^{-1}) = –B^{-1}·dB·B^{-1} , and because trace(B·C) = trace(C·B) whenever both matrix products B·C and C·B are square. By applying these formulas we find that

$$\begin{aligned} d\,£(B) &= -\mathrm{trace}((2I + B + B^{-1})^{-1}\cdot(dB - B^{-1}\cdot dB\cdot B^{-1})) \\ &= \mathrm{trace}((I+B)^{-1}\cdot(I-B)\cdot B^{-1}\cdot dB) = \mathrm{trace}(\$(B)\cdot B^{-1}\cdot dB)\ . \end{aligned}$$

How does £(Q·Ω) behave for any fixed unitary Q as Ω runs through the set of all diagonal unitary matrices? This set is swept out by Ω := $e^{\imath \mathrm{Diag}(x)}$ as real vector x runs throughout any hypercube with side-lengths bigger than 2π ; and £(Q·$e^{\imath \mathrm{Diag}(x)}$) must assume its minimum value at some real vector(s) x strictly inside such a hypercube. Such a minimizing Q·$e^{\imath \mathrm{Diag}(x)}$ is a unitary Q·Ω "nearest" I . Let's investigate the Cayley transform of a "nearest" Q·Ω .

Abbreviate Diag(x) = X and Diag(dx) = dX ; and note that X and dX commute, so that d Ω = d $e^{\imath X}$ = $\imath e^{\imath X}$·d X = ıΩ·dX , and therefore

$$d\,£(Q\cdot\Omega) = \mathrm{trace}(\ \$(Q\cdot\Omega)\cdot e^{-\imath X}Q^{-1}\cdot Q\,\imath e^{\imath X}\cdot dX\ ) = \imath\ \mathrm{diag}(\$(Q\cdot\Omega))^T dx\ .$$

Since this d £ must vanish at a minimum of £ for every real dx , so diag($(Q·Ω)) = o there. Thus the question that is this work's title must have an affirmative answer, namely …

Theorem: For each unitary Q there exists at least one unitary diagonal Ω for which the skew-Hermitian Cayley transform S := $(I + Q\cdot\Omega)^{-1}$·(I – Q·Ω) = –S^H has diag(S) = o .

The theorem's " at least one " tends to understate how many such diagonals Ω exist. To see why, set Ω := $e^{\imath \mathrm{Diag}(x)}$ again and consider the locus of poles of the function £(Q·$e^{\imath \mathrm{Diag}(x)}$) of the real column x . These poles are the zeros x of det(I + Q·$e^{\imath \mathrm{Diag}(x)}$) . Substitution of the Cayley transform Z := $(Q) = –$Z^H$, perhaps after shifting x's origin by applying §2's Lemma, transforms the determinantal equation for the locus of poles into an equivalent equation

$$\det\big(\ \cos(\mathrm{Diag}(x/2)) - \imath\ Z\cdot\sin(\mathrm{Diag}(x/2))\ \big) = 0\ . \qquad (*)$$

Despite first appearances, the left-hand side of this equation is a real function of the real vector x because matrix cot(Diag(x/2)) – ı Z is Hermitian wherever it is finite. Moreover that left-hand side reverses sign somewhere because it takes both positive and negative values at vectors x whose elements are various integer multiples of 2π . Therefore the space of real vectors x is partitioned into cells by the locus of poles of £ ; inside each cell £ is finite and nonnegative, and the left-hand side of (*) takes on a constant nonzero sign probably opposite to the sign in adjacent cells. Inside every cell each local minimum (or any other *critical point* x where $\partial£/\partial x = o^T$) of £ provides another of the theorem's diagonals $\Omega := e^{\imath \mathrm{Diag}(x)}$. These are likely to be numerous, as we shall see next.

§4: Examples

For every integer $n > 1$ examples exist for which the number of the theorem's diagonals Ω is infinite in the general complex case, 2^{n-1} in the restricted-to-real case. All these diagonals Ω minimize £ ; all of them provide skew Cayley transforms S whose $\mathrm{diag}(S) = o$ and whose every off-diagonal element has magnitude 1 . Here is such an example:

Define n-by-n real orthogonal $G := \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & \dots & \dots & & & \\ & & & \dots & \dots & & \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ (-1)^{n-1} & & & & & & 0 \end{bmatrix}$, and

let Ω run through unitary diagonal matrices with $\det(\Omega) \neq -1$. Then unitary $Q := G\cdot\Omega$ has a skew-Hermitian Cayley transform $S = \$(Q) := (I+Q)^{-1}\cdot(I-Q) = -S^T$ which, as we shall show, has off-diagonal elements all of the same magnitude $2/|1 + \det(\Omega)|$. Moreover this magnitude is minimized just when $\det(\Omega) = +1$, the minimized magnitude is 1 , and $\mathrm{diag}(S) = o$. In particular, for every real orthogonal diagonal Ω of ±1's with an even number of -1's , S is a real skew matrix all of whose off-diagonal elements are ±1's . We'll prove these claims next.

First we must confirm that $\$(Q)$ exists; it will follow from $\Omega^{-1} = \overline{\Omega}$ (the complex conjugate):

$$\det(I+Q) = \det(I + G\cdot\Omega) = \det(\overline{\Omega} + G)\cdot\det(\Omega) = (\det(\overline{\Omega}) + 1)\cdot\det(\Omega) = 1 + \det(\Omega) \neq 0 .$$

Next confirm that the powers $Q^0 = I$, $Q, Q^2, Q^3, \dots, Q^{n-1}$ are linearly independent because their nonzero elements occupy non-overlapping positions in the matrix. Just as $G^n = (-1)^{n-1}\cdot I$, so does Q^n turns out to be a scalar multiple of I . Our next task is to determine this scalar.

Start by defining the n-vector $u := \mathrm{diag}(\Omega)$ so that $\Omega = \mathrm{Diag}(u)$ and the elements of u all have magnitude 1 and product $\det(\Omega)$. Next observe that $G\cdot\mathrm{Diag}(v) = \mathrm{Diag}(G\cdot v)\cdot G$ for any n-vector v . Use this to confirm by induction that

$$(G\cdot\Omega)^k = \mathrm{Diag}(G\cdot u)\cdot\mathrm{Diag}(G^2\cdot u)\cdot\mathrm{Diag}(G^3\cdot u)\cdot\ldots\cdot\mathrm{Diag}(G^k\cdot u)\cdot G^k \quad \text{for each } k = 1, 2, 3, \dots \text{ in turn.}$$

In particular, when $k = n$ we find that $Q^n = (G\cdot\Omega)^n = (-1)^{n-1}\cdot\prod_{1\le k\le n}\mathrm{Diag}(G^k\cdot u)$. Each diagonal element of this product includes the product of all the elements of u each once, and their product is $\det(\Omega)$. Factor it out to obtain $Q^n = \det(\Omega)\cdot(G\cdot I)^n = \det(\Omega)\cdot(-1)^{n-1}\cdot I$.

The last equation figures in the confirmation of an explicit formula for the Cayley transform:

$$\$(Q) = (I+Q)^{-1}\cdot(I-Q) = \big((1-\det(\Omega))\cdot I + 2\textstyle\sum_{1\le k\le n-1} (-1)^k Q^k \big)/(1+\det(\Omega)) .$$

To confirm it multiply by $I+Q$ and collect terms. This formula validates every claim uttered above for $\$(Q)$ because every unitary diagonal Ω has $|\det(\Omega)| = 1$.

£(Q) , the gauge of "nearness" to I , is minimized when $\det(\Omega) = 1$ and $\mathrm{diag}(S) = o$ since £(Q) = $n\cdot\log(4) - 2\cdot\log|1 + \det(\Omega)| \ \ge\ (n-1)\cdot\log(4)$ with equality just when $\det(\Omega) = 1$.

Here is a different example $Q := \$\left(\begin{bmatrix} 0 & 2 & -2 \\ -2 & 0 & 2 \\ 2 & -2 & 0 \end{bmatrix}\right) = \begin{bmatrix} -3 & 4 & 12 \\ 12 & -3 & 4 \\ 4 & 12 & -3 \end{bmatrix}/13$. Six unitary diagonals Ω satisfy the theorem. Four are real: $\Omega = I$, Diag([–1; –1; 1]) , Diag([1; –1; –1]) and Diag([–1; 1; –1]) .

Typical of the last three is $\$(Q\cdot\mathrm{Diag}([-1; 1; -1])) = \begin{bmatrix} 0 & -1 & \frac{1}{2} \\ 1 & 0 & 1 \\ -\frac{1}{2} & -1 & 0 \end{bmatrix}$; none of them minimizes £(Q·Ω) .

It is minimized by two complex scalar diagonals $\Omega := (-5 \pm 12\imath)I/13$ for which respectively

$\$(Q\cdot\Omega) = \begin{bmatrix} 0 & -1-3\imath & 1-3\imath \\ 1-3\imath & 0 & -1-3\imath \\ -1-3\imath & 1-3\imath & 0 \end{bmatrix}/4$ and its complex conjugate. Note that its every element is strictly smaller than 1 in magnitude, unlike the theorem's four real instances.

§5: Why Minimizing £(Q·Ω) Makes $(Q·Ω) Small.

In general, can the theorem's $S := \$(Q\cdot\Omega)$ be huge for a $Q\cdot\Omega$ "nearest" I ? No; here is why: Once again abbreviate $\mathrm{Diag}(x+\Delta x) = X+\Delta X$ for real columns $x+\Delta x$, and set unitary diagonal $\Omega := e^{\imath X}$, and abbreviate $\$(Q\cdot\Omega) = S$. The second term of the Taylor series expansion

$$£(Q\cdot\Omega\cdot e^{\imath\Delta X}) = £(Q\cdot\Omega) + \big(\partial £(Q\cdot\Omega)/\partial x\big)\cdot\Delta x + \big(\partial^2 £(Q\cdot\Omega)/\partial x^2\big)\cdot\Delta x\cdot\Delta x/2 + O(\Delta x)^3$$

must vanish and the third must be nonnegative for all Δx at a local minimum x of £ . We already have $\partial £(Q\cdot\Omega)/\partial x = \imath\,\mathrm{diag}(S)^T$, and next we shall compute $\partial^2 £(Q\cdot\Omega)/\partial x^2$.

The next two paragraphs serve only to introduce my notation to readers unacquainted with it. Others may skip them.

A continuously differentiable scalar function $f(x)$ of a column-vector argument x has a first *derivative* denoted by $f'(x) = \partial f(x)/\partial x$. It must be a row vector since scalar $df(x) = f'(x)\cdot dx$. Sometimes this *differential* is easier to derive than the derivative; it means that, for every differentiable vector-valued function $x(\mu)$ of any scalar variable μ , the chain rule yields a derivative $df(x(\mu))/d\mu = f'(x(\mu))\cdot x'(\mu)$. For any fixed x this $f'(x)$ is a *linear functional* acting linearly upon vectors in the same space as x and represented by a row often called "The Jacobian Array of First partial Derivatives". Such is $\partial £(Q\cdot e^{\imath\,\mathrm{Diag}(x)})/\partial x = \imath\,\mathrm{diag}(S)^T$.

If $f(x)$ is continuously twice differentiable its second derivative, denoted by $f''(x) = \partial^2 f(x)/\partial x^2$, is a *symmetric bilinear operator* acting upon pairs of vectors in the same space as x . "Symmetric" means $f''(x)\cdot y\cdot z = f''(x)\cdot z\cdot y$ because of H.A. Schwarz's lemma that tells when the order of differentiation does not matter. The "Hessian Array of Second partial Derivatives" is a symmetric matrix H(x) that yields $f''(x)\cdot y\cdot z = z^T\cdot H(x)\cdot y$. Sometimes we can derive the differential $df'(x)\cdot y = f''(x)\cdot y\cdot dx = dx^T\cdot H(x)\cdot y$ more easily than the derivative. Such will be the case for the second derivative $\partial^2 £(Q\cdot e^{\imath\,\mathrm{Diag}(x)})/\partial x^2$ derived hereunder.

Recall that the differential of the unitary diagonal $\Omega := e^{\imath X}$ is $d\Omega = \imath\,\Omega\cdot dX$. Then rewrite

$$S = \$(Q\cdot\Omega) = (I+Q\cdot\Omega)^{-1}(I-Q\cdot\Omega) = 2(I+Q\cdot\Omega)^{-1} - I$$

to see easily why

$$dS = -2(I+Q\cdot\Omega)^{-1}\cdot Q\cdot d\Omega\cdot(I+Q\cdot\Omega)^{-1} = -2\imath\,(I+Q\cdot\Omega)^{-1}\cdot Q\cdot\Omega\cdot dX\cdot(I+Q\Omega)^{-1}$$
$$= -\imath\,(I+S)\cdot(I+S)^{-1}\cdot(I-S)\cdot dX\cdot(I+S)/2 = -\imath\,(I-S)\cdot dX\cdot(I+S)/2 \ .$$

Next, $\big(\partial£(Q\cdot\Omega)/\partial x\big)\cdot\Delta x = \imath\,\mathrm{diag}(S)^T\cdot\Delta x = \imath\,\mathrm{trace}(S\cdot\Delta X)$ for any fixed column Δx and therefore

$$\big(\partial^2£(Q\cdot\Omega)/\partial x^2\big)\cdot dx\cdot\Delta x = d\,\big(\partial£(Q\cdot\Omega)/\partial x\big)\cdot\Delta x = \imath\,d\,\mathrm{trace}(S\cdot\Delta X) = \imath\,\mathrm{trace}(d\,S\cdot\Delta X)$$
$$= \imath\,\mathrm{trace}(-\imath\,(I-S)\cdot dX\cdot(I+S)\Delta X)/2 = \mathrm{trace}(dX\cdot\Delta X - S\cdot dX\cdot\Delta X + dX\cdot S\cdot\Delta X - S\cdot dX\cdot S\cdot\Delta X)/2$$
$$= \mathrm{trace}\big(dX\cdot\Delta X + (S^H\cdot dX)\cdot(S\Delta X)\big)/2 = dx^T\cdot(I + |S|^2)\cdot\Delta x/2$$

wherein $|S|^2$ is obtained elementwise by substituting $|s_{ij}|^2$ for each element s_{ij} in S .

Thus we have derived the first three terms of the Taylor Series expansion

$$£(Q\cdot\Omega\cdot e^{\imath\Delta X}) = £(Q\cdot\Omega) + \imath\,\mathrm{diag}(S)^T\cdot\Delta x + \Delta x^T\cdot(I + |S|^2)\cdot\Delta x/4 + O(\Delta x)^3 \ .$$

Since $\mathrm{diag}(S) = o$ and $I + |S|^2$ must be a positive (semi)definite matrix at a minimum of £ , every $|s_{ij}| \le 1$ there. Consequently …

Corollary: At least one of the Theorem's complex skew-Hermitian Cayley transforms $S := \$(Q\cdot\Omega)$ with $\mathrm{diag}(S) = o$ also has every element $|s_{ij}| \le 1$.

§6: 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 .

Prof. W. Kahan
Mathematics Dept. #3840
University of California
Berkeley CA 94720-3840

As an old acquaintance since 1959, I proffer this work to Prof. Dr. F.L. Bauer of Munich for his 80th birthday.