§0 The Problem:
Given an m-by-n matrix B with $m \ge n$, we seek the nearest m-by-n matrix Q with n orthonormal columns (*i.e.* n-by-n $Q'\cdot Q = I$); it is "nearest" in so far as it minimizes *both*
$\|B-Q\|_F := \sqrt{\text{trace}((B-Q)'\cdot(B-Q))}$ and $\|B-Q\|_2 :=$ Largest Singular Value of B–Q.
For simplicity we assume the columnns of B are linearly independent, as is almost always the case. This nearest $Q = \hat{Q}$ will be compared with Gram-Schmidt's and QR factorization's (re)orthogonalizations Q of B. Then $\hat{Q}$ will be applied to three tasks of which the third, the estimation of angles between subspaces, can be afflicted by inaccuracy unless aided by $\hat{Q}$.

§1 The Solution: The nearest Q is unique; it is the orthogonal factor $\hat{Q} = B\cdot H^{-1}$ of the
Polar Decomposition of $B = \hat{Q}\cdot H$ in which $H = H' := \sqrt{B'\cdot B}$ is positive definite.

Proof:
Let $B = P\cdot\begin{bmatrix} V \\ O \end{bmatrix}\cdot G'$ be the *Singular Value Decomposition* of B; here $P' = P^{-1}$, $G' = G^{-1}$ and V is a square positive diagonal matrix with the singular values of B on its diagonal. The zero matrix O under V is absent when B is square. We claim that *the* orthogonal matrix nearest B is $\hat{Q} := P\cdot\begin{bmatrix} I \\ O \end{bmatrix}\cdot G' = B\cdot(B'\cdot B)^{-1/2}$. To see why this is so consider the difference B–Q for any m-by-n matrix Q with n orthonormal columns; $Q'\cdot Q = I$. Neither norm $\|B-Q\|_{...}$ is changed by unitary pre- or post-multiplication, so $\|B-Q\| = \|P'\cdot(B-Q)\cdot G\|$. Now,

$$P'\cdot B\cdot G = \begin{bmatrix} V \\ O \end{bmatrix}, \quad P'\cdot\hat{Q}\cdot G = \begin{bmatrix} I \\ O \end{bmatrix}, \quad \text{and} \quad P'\cdot Q\cdot G = \begin{bmatrix} C \\ S \end{bmatrix} \text{ satisfying } C'\cdot C + S'\cdot S = I.$$

Consequently $\|B-Q\| = \left\| \begin{bmatrix} V \\ O \end{bmatrix} - \begin{bmatrix} I \\ O \end{bmatrix} + \begin{bmatrix} I \\ O \end{bmatrix} - \begin{bmatrix} C \\ S \end{bmatrix} \right\| = \|Z\|$ wherein $Z := \begin{bmatrix} V-I \\ O \end{bmatrix} - \begin{bmatrix} C-I \\ S \end{bmatrix}$. Hence
$Z'\cdot Z = (V-I)^2 - (V-I)\cdot(C-I) - (C-I)'\cdot(V-I) + (C-I)'\cdot(C-I) + S'\cdot S = (V-I)^2 + V\cdot(I-C) + (I-C')\cdot V$.
Now, $I-C = O$ *only* when $Q = \hat{Q}$. Otherwise $\text{Real}(\text{Diag}(I-C)) \ge O$ because $\text{Diag}(C'\cdot C) \le I$; moreover at least one element of $\text{Real}(\text{Diag}(I-C))$ must be positive, and consequently we find $\|B-Q\|_F^2 = \text{trace}(Z'\cdot Z) > \text{trace}((V-I)^2) = \|B-\hat{Q}\|_F^2$ as was claimed for $Q \ne \hat{Q}$. Next we turn to $\|B-Q\|_2^2 = \max_{\|x\|=1}\|(B-Q)\cdot x\|^2 = \max_{\|e\|=1} e'\cdot Z'\cdot Z\cdot e$. When $Q = \hat{Q}$ we get $Z'\cdot Z = (V-I)^2$, and then the maximizing column vector $e = \hat{e}$, say, can be one of the columns of I because V is diagonal. Otherwise (when $Q \ne \hat{Q}$) we find, because $\text{Real}(\text{Diag}(I-C)) \ge O$, that
$\|B-Q\|_2^2 \ge \hat{e}'\cdot Z'\cdot Z\cdot\hat{e} = \hat{e}'\cdot(V-I)^2\cdot\hat{e} + \hat{e}'\cdot\big(V\cdot(I-C) + (I-C')\cdot V\big)\cdot\hat{e} \ge \hat{e}'\cdot(V-I)^2\cdot\hat{e} = \|B-\hat{Q}\|_2^2$,
as was claimed. END OF PROOF.

The same proof works when the columns of B are linearly dependent except that then V has at least one zero on its diagonal, whence a nearest orthonormal $\hat{Q}$ is not unique. Neither need it always be unique if it has to minimize $\|B-Q\|_2$ but not $\|B-Q\|_F$ too; do you see why?

See a slightly more general *Procrustes* problem concerning a nearest matrix with orthonormal columns in §12.4.1 of G.H. Golub & C.F. Van Loan's *Matrix Computations* (2d. ed. 1989, Johns Hopkins Univ. Press). See also p. 385 of Nicholas J. Higham's book *Accuracy and Stability of Numerical Algorithms* 2d. ed. (2002, SIAM, Philadelphia) for related citations.

§2 Approximate Solutions:

If B is already nearly orthogonal then, rather than compute its singular value decomposition, we can compute a residual $Y := B'\cdot B - I$ and then use as many terms as necessary of a series

$$\hat{Q} = B\cdot(B'\cdot B)^{-1/2} = B\cdot(I + Y)^{-1/2} = B - B\cdot Y\cdot(I/2 - 3Y/8 + 5Y^2/16 - 35Y^4/128 + \ldots)$$

of which only the first few terms can be worth using; otherwise the SVD is faster.

If Y is so small that $1 - \|Y\|_{\ldots}^2$ rounds to 1 for practically any norm $\|\ldots\|_{\ldots}$, then $\hat{Q}$ will be approximated adequately by $\overline{Q} := B - \frac{1}{2}B\cdot Y = \hat{Q}\cdot(I - 3Y^2/8 + Y^3/8 - \ldots)$ because its residual $\overline{Q}'\cdot\overline{Q} - I = \frac{1}{4}Y^2\cdot(Y - 3I)$ will be negligible. And if Y is not so small, but still $\|Y\|_{\ldots} << \frac{1}{2}$, then the columns of $\overline{Q}$ will be rather more nearly orthonormal than those of B; and repeating upon $\overline{Q}$ the process performed upon B will yield an approximation $\hat{Q}\cdot(I - 27Y^4/128 + \ldots)$.

The accuracy to which $\overline{Q}$ can approximate $\hat{Q}$ is limited by the accuracy left in the residual Y after cancellation. Y is best computed by extra-precise accumulation of scalar products during matrix multiplication. The MATLAB expression `Y = [B', I}*[B; -I]` does this in version 5.2 on old 680x0-based Apple Macintoshes, and in version 6.5 on Wintel PCs after the command `system_dependent('setprecision', 64)` has been executed.

Occasionally $\hat{Q}$ is best computed as an unconsummated sum $\hat{Q} = Q + \Delta Q$ in which Q and ΔQ are stored separately. Such an occasion arises when roundoff makes the rounded sum $Q + \Delta Q$ materially less accurate than the unrounded sum, and when some other residual like $H\cdot\hat{Q} - \hat{Q}\cdot V$ to be computed extra-precisely is rendered as $(H\cdot Q - Q\cdot V) + (H\Delta Q - \Delta Q\cdot V)$ a little more accurately because the first term $(H\cdot Q - Q\cdot V)$ already enjoys massive cancellation.

§3 Other Reorthogonalization Schemes:

Given a perhaps rectangular matrix B whose columns are nearly orthonormal, Gram-Schmidt and QR factorization are the reorthogonalization schemes that may come to mind first. These schemes change B into $Q := B\cdot R^{-1}$ where $Q'\cdot Q = I$ and R is upper-triangular. To ensure that both schemes produce the same R we insist that its diagonal be positive, thus determing R uniquely as the upper-triangular right Cholesky factor of $B'\cdot B = R'\cdot R$ except for rounding errors that differ among the schemes.

These two schemes suffer from an accidental dependence upon the ordering of B's columns. For instance if the last column is slightly in error but the others are accurately orthonormal, these schemes adjust only the last column. However, if all columns but the first are accurately orthogonal but the first errs a bit, these schemes leave the first column's direction unchanged and infect the others with its error. So, permuting the columns of B can change the Gram-Schmidt's or QR's Q non-trivially while merely permuting the columns of $\hat{Q} := B\cdot(B'\cdot B)^{-1/2}$.

On the other hand, scaling the columns of B, replacing it by $B\cdot D$ for some positive diagonal D not merely a scalar multiple of I, leaves Q unchanged (and replaces R by $R\cdot D$) but alters $\hat{Q}$ nontrivially. Apparently, alterations to B can alter Q and $\hat{Q}$ rather differently.

§4 Comparisons with Other Reorthogonalization Schemes:

How much closer to B than QR's Q is $\hat{Q}$? Now, $B = \hat{Q}\cdot H = Q\cdot R$ wherein $H = H'$ is the positive definite square root of $H^2 = R'\cdot R = B'\cdot B$ and R is upper-triangular with a positive diagonal. Consequently the distances' ratios are $\|B - Q\|/\|B - \hat{Q}\| = \|R - I\|/\|H - I\|$ for both $\|\ldots\|_2$ and $\|\ldots\|_F$. Having found these ratios to be 1 or more, we wish to estimate how big they can be at most. They cannot exceed 1 much if $\|B\| = \|H\| = \|R\|$ is either very big or very tiny, so the only matrices B worth considering are those restricted in some way that moderates their norms. The restriction chosen hereunder forces each column of B, and hence R, to have norm 1, since it is an easy computation and not too expensive. The chosen restriction turns this paragraph's question into the following:

Suppose n-by-n triangular matrix R has a positive diagonal, and $H = H'$ is the positive definite square root of $H^2 = R'\cdot R = B'\cdot B$ whose diagonal is I. Then how big at most can any of the four ratios $\rho_{\cdot\cdot}(R) := \|R - I\|/\|H - I\| = \|B - Q\|/\|B - \hat{Q}\|$ be?

This question's answers will be complicated by two further questions:

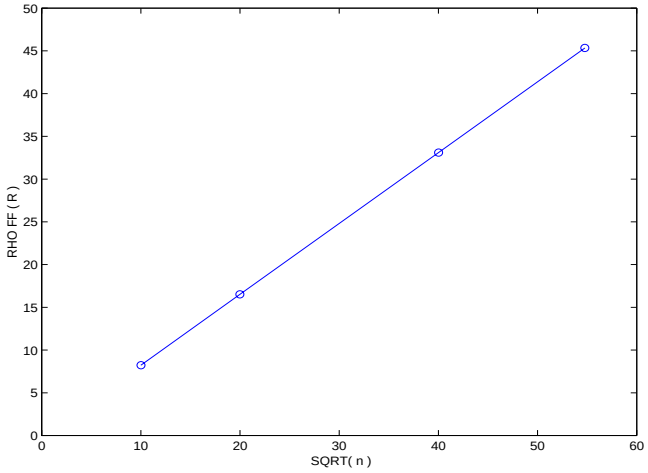
- Are B's columns, and hence R's, already nearly orthonormal, or not?
- Are B's columns, and also R's, real, or complex?

If the given columns are not nearly orthonormal, then the question has an answer buried in §4 of a paper "Backward errors for eigenvalue and singular value decompositions" by Shivkumar Chandrasekaran and Ilse C.F. Ipsen, pp. 215-223 of *Numer. Math.* **68** (1994). They found $\rho_{\cdot 2}(R) := \|R - I\|/\|H - I\|_2 \le 5\sqrt{n}$. Their bound on $\rho_{F2}(R)$ is not too pessimistic since examples exist for which the smaller $\rho_{FF}(R)$ seems to grow proportional to $\sqrt{n}$. An n-by-n example R is obtained by scaling each column of the upper triangular (in MATLAB's notation)

```
toeplitz([1; zeros(n-1,1)], [1, λ.^[0:n-2]])
```

to have norm 1 after choosing $\lambda := (\sqrt{5} - 1)/2 \approx 0.618034\ldots$. In particular …

Dimension n :	100	400	1600	3000
$\rho_{FF}(R) = \|R - I\|_F/\|H - I\|_F$:	8.2218	16.5282	33.0985	45.3310



However, $5\sqrt{n}$ seems far too pessimistic a bound for $\rho_{22}(R) := \|R - I\|_2/\|H - I\|_2$ since, so far as I know, no large n-by-n example R has been found whose ratio $\rho_{22}(R)$ rises within an order of magnitude of their bound. Neither has a much smaller bound been found yet.

What if B's columns, and hence R's, are already nearly orthonormal? To pursue this question we define linear operators $\mathcal{U}$ and $\mathcal{L}$ acting upon square matrices F :

$\mathcal{U}(F)$ keeps the upper triangle and half the diagonal of F , and zeros the lower triangle;

$\mathcal{L}(F)$ keeps the lower triangle and half the diagonal of F , and zeros the upper triangle.

Consequently $\mathcal{U}(F) + \mathcal{L}(F) = F$, and $\mathcal{U}(F) = \mathcal{L}(F')'$ in MATLAB's notation, in which

$$\mathcal{U}(F) = \texttt{triu(F) - 0.5*diag(diag(F))} .$$

Now consider an m-by-n example $B = \hat{Q}\cdot(I + \Delta H)$ in which $\Delta H := (B'\cdot B)^{1/2} - I = \Delta H'$ is so tiny, though not negligible, that its square is negligible, and $\hat{Q}$ is the matrix nearest B with orthonormal columns. Then Gram-Schmidt or QR computes the upper-triangular Cholesky factor R of $B'\cdot B = R'\cdot R$ as $R \approx I + 2\mathcal{U}(\Delta H)$ whence follows $\rho_{..}(R) \approx \|2\mathcal{U}(\Delta H)\|/\|\Delta H\|$ approximately, ignoring terms like ΔH^2. Evidently $1 \le \rho_{FF}(R) \le \sqrt{2}$. A bound for $\rho_{22}(R)$ is more complicated:

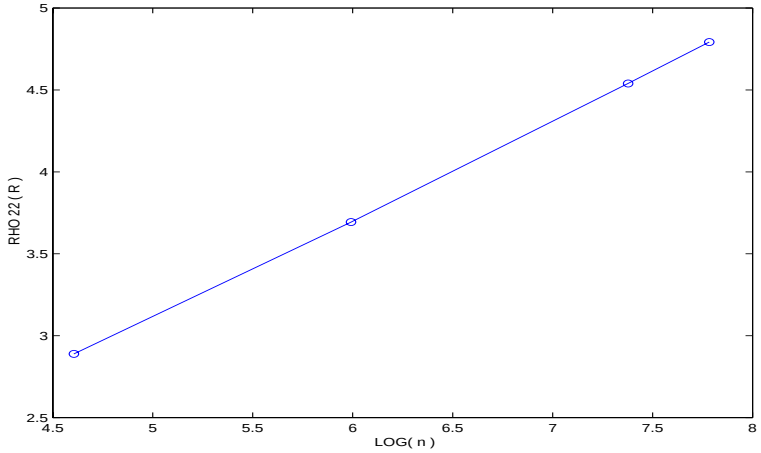
$$\begin{aligned}\rho_{22}(R) &\approx \|\Delta H + \mathcal{U}(\Delta H) - \mathcal{L}(\Delta H)\|_2/\|\Delta H\|_2\\ &\le 1 + \|\mathcal{U}(\Delta H) - \mathcal{U}(\Delta H)'\|_2/\|\mathcal{U}(\Delta H) + \mathcal{U}(\Delta H)'\|_2\\ &< 1 + 4\pi - (2/\pi)\cdot log(\pi) + (2/\pi)\cdot log(n) \approx 12.84 + 0.6366\cdot log(n) ,\end{aligned}$$

according to Corollary 2 in "Spectra of Operators with Fixed Imaginary Parts" by Andrzej Pokrzywa, pp. 359-364 in *Proc. Amer. Math. Soc.* **81** #3 (Mar. 1981). His work also implies that $\max_R \rho_{22}(R)$ must grow like $0.6366\cdot log(n)$ when B's n columns are already almost orthonormal and n is huge. A simple n-by-n example R that appears to illustrate logarithmic growth is obtained by scaling each column of the upper triangular (in MATLAB's notation)

$$\texttt{toeplitz([1; zeros(n-1,1)], [1, λ./[1:n-1]])}$$

to have norm 1 after choosing a small *pure imaginary* $\lambda := \imath/2^{25}$ . . In particular …

Dimension n :	100	400	1600	2400
$\rho_{22}(R) = \|R - I\|_2/\|H - I\|_2$:	2.8885	3.6929	4.5403	4.7923



The data in this example R are *complex numbers*. *Real*-valued examples R whose $\rho_{22}(R)$ grows substantially with dimension n were not found until Andrzej Pokrzywa suggested the use of a familiar map $\mathcal{R}$ from complex m-by-n matrices to a subspace of real 2m-by-2n matrices. $\mathcal{R}(Z)$ replaces each complex element $\zeta = \xi + \imath\eta$ of Z by a 2-by-2 real matrix

$$\begin{bmatrix} \xi & \eta \\ -\eta & \xi \end{bmatrix}.$$

MATLAB's Kronecker product provides an easy implementation of $\mathcal{R}(Z)$ thus:

```
kron(real(Z),  [1, 0;  0, 1])  +  kron(imag(Z),  [0, 1;  -1, 0])
```

Many matrix operations and relations persist after the application of $\mathcal{R}$:

$\mathcal{R}(Z') = \mathcal{R}(Z)'$, $\mathcal{R}(W \pm Z) = \mathcal{R}(W) \pm \mathcal{R}(Z)$, $\mathcal{R}(W\cdot Z) = \mathcal{R}(W)\cdot\mathcal{R}(Z)$, $\mathcal{R}(W^{-1}\cdot Z) = \mathcal{R}(W)^{-1}\cdot\mathcal{R}(Z)$.

The eigenvalues of $\mathcal{R}(Z)$ consist of both the eigenvalues of Z and their complex conjugates.

If R is upper-triangular, so is $\mathcal{R}(R)$; and if R has a positive diagonal, so does $\mathcal{R}(R)$.

If H = H' is Hermitian and positive definite, $\mathcal{R}(H)$ is symmetric and positive definite; and then if Cholesky factorization R'·R = H, so does $\mathcal{R}(R)'\cdot\mathcal{R}(R) = \mathcal{R}(H)$ and so does the positive definite $\sqrt{(\mathcal{R}(H))} = \mathcal{R}(\sqrt{H})$ (except for roundoff).

Finally $\|\mathcal{R}(Z)\|_2 = \|Z\|_2$ but $\|\mathcal{R}(Z)\|_F = \sqrt{2}\cdot\|Z\|_F$.

Consequently $\rho_{22}(\mathcal{R}(R)) = \rho_{22}(R)$; therefore, apparently rare real-valued examples whose ρ_{22} grows like the logarithm of dimension n do exist. All of these have nearly orthonormal columns, so their ρ_{22} cannot grow faster. A more general example, not nearly orthonormal, whose ρ_{22} grows at least as fast as log(n) has not been found yet; perhaps none exists.

§5 Three Applications of the Nearest Orthogonal or Unitary Matrix $\hat{Q}$:

Apparently Gram-Schmidt's or QR's Q is rarely very far from the $\hat{Q}$ nearest B. When is Q far enough from $\hat{Q}$ that the extra work, if any, needed to compute $\hat{Q}$ will be rewarded?

(1): Suppose the columns of B approximate some or all eigenvectors of a real symmetric or Hermitian matrix H = H'. Software that computes partial eigensystems, or computes them in parallel, may sacrifices some orthogonality of eigenvectors belonging to clustered eigenvalues in order to get a result B faster. Algorithms to improve the approximations must start by more nearly (re)orthogonalizing the columns of B to get Q. Consider three ways to compute that Q: First is $B = Q \cdot R$ factorization. Second is the Polar factorization $B = \hat{Q} \cdot H$. Third is §2's approximation $\hat{Q} \approx B - \frac{1}{2} B \cdot Y$ usable when the residual $Y := B' \cdot B - I$ is small enough. No matter which way that Q is computed, it will be used to compute a residual $R := H \cdot Q - Q \cdot A$ for some approximation A to a diagonal matrix of eigenvalues of H. Each eigenvalue of A approximates some eigenvalue of H within $\pm \|R\|_2$; and as many eigenvalues of H are thus approximated as A has. The best $Å := Q' \cdot H \cdot Q$ minimizes $\|R\|_2$ and $\|R\|_F$; if all other eigenvalues of H differ by at least a gap $\gamma >> \|R\|_2$ from the eigenvalues of Å then these approximate their corresponding eigenvalues of H within about $\pm \|R\|_2^2/\gamma$. These assertions about Å are proved in B.N. Parlett's book *The Symmetric Eigenproblem* (1998, SIAM, Philadelphia). Unproved, but supported by some experimental evidence, is an expectation that $Å := \hat{Q}' \cdot H \cdot \hat{Q}$ is most nearly diagonal, so its eigenvalues are easiest to compute accurately, thus perhaps compensating for whatever extra effort it costs to use $\hat{Q}$ in place of QR's Q.

(2): A square nearly unitary matrix B can be computed in the course of obtaining the Schur decomposition $G = B \cdot U \cdot B'$ of a given non-Hermitian square matrix G ; here upper-triangle U has desired eigenvalue approximations on its diagonal. Refinement of these approximations begins with the replacement of B by a more nearly unitary matrix Q before computing the residual $R := G \cdot Q - Q \cdot U$ and then the error $Q' \cdot R = Q' \cdot G \cdot Q - U$ in the computed decomposition. Details of the refinement process are a long story for another day. A part of that story not yet resolved is the advantage, if any, of using $\hat{Q}$ in place of Q from QR factorization.

(3): The angles between two subspaces, one spanned by E's columns and the other by F's, are often computed by first (re)orthogonalizing E and F, then computing the *column* $c := \text{svd}(F' \cdot E)$ of singular values of $F' \cdot E$, and then the column of angles $\theta := \arccos(c)$. Here QR factorization can be used to (re)orthogonalize E's and F's columns unless they are already so nearly orthonormal that §2's Approximate Solution is accurate enough. On the other hand, if the columns of E, say, are too nearly linearly dependent then the subspace they span must be partially indeterminate in the face of noise (like roundoff) in those columns, and the QR factorization must be partially indeterminate too. The following digression is intended to cope with that indeterminacy in a way roughly similar to what MATLAB's `orth(E)` does.

First scale each column of E to make its estimated noise (or uncertainty) about the same in norm as every other column's. Singular Value Decomposition `[E, V, Discard] = svd(E, 0)` computes a diagonal matrix V of the singular values of E and overwrites E by orthonormal columns spanning the same subspace. Some singular values may lie below the estimated noise level; if so, the corresponding columns of the new E are too uncertain and should be discarded. Suspicion should fall also upon either the noise level's estimate or those columns corresponding to singular values, if any, that exceed the noise level only a little.

Thus (re)orthogonalized, the columns of E span its same subspace as before, and likewise for F. However, roundoff in the formula $\theta := \arccos(c)$ can lose up to half the sig. digits carried by the arithmetic when some angles θ between the two subspaces are tiny, as happens when one subspace is an invariant subspace and the other is an approximation to it provided by an eigensystem program under test. The obvious way to cope with that loss is to compute θ in arithmetic twice as precise as is trusted in the data E and F, and as is desired in θ. When extra-precise arithmetic is unavailable or uneconomical, the following scheme avoids that loss of accuracy:

Suppose E and F have fairly accurately orthonormal columns, and that E has no more columns than F has. (Of course, both matrices must have the same number of rows, at least as many as F has columns.) Compute $B := F'\cdot E$, then its nearest matrix $\hat{Q}$ with orthonormal columns. When all the angles θ are small, $F'\cdot E$ is close enough to $\hat{Q}$ that §2's Approximate Solution is adequate and can be computed quickly without the SVD of $F'\cdot E$. Next compute the *columns* $s := \mathrm{svd}(F\cdot\hat{Q} - E)$ of singular values, and $\theta := 2\arcsin(s/2)$ elementwise. The absolute error in θ is at worst of the order of the roundoff threshold (MATLAB's `eps`) rather than its square root. All this can be proved by means similar to what worked in §1's Proof above. Thus does the accuracy of all angles θ repay the modest extra effort that $\hat{Q}$ costs.

A similar scheme to compute angles θ appeared in "Numerical methods for computing angles between linear subspaces" by Å. Bjorck & G. Golub, pp. 579 - 594 of *Math. Comp.* **27** (1973); it's accurate for tiny angles but can lose almost half the arithmetic's digits at angles θ near $\pi/2$.

. . . . . .

Prof. Nicholas J. Higham has kindly e-mailed a pointer to p. 235 of his book "**Functions of Matrices** *Theory and Computation*" (2008, SIAM, Philadelphia) where he cites an inequality in Lemma 2.4 of Prof. Ji-Guang Sun's "A Note on Backward Perturbations for the Hermitian Eigenvalue Problem" pp. 385-393 of ***BIT* 35** (1995, Springer, Heidelburg) to the effect that if $\|B'\cdot B - I\|_2 < 1$ then $\rho_{FF}(R) = \|B - Q\|_F/\|B - \hat{Q}\|_F \le \sqrt{\frac{1}{2}}\cdot(1 + \|B\|_2)/(1 - \|B'\cdot B - I\|_2)$. This includes my inequality $\rho_{FF}(R) \le \sqrt{2}$ valid only when $\|B'\cdot B - I\|_2$ is infinitesimal, and betters the Chandrasekaran-Ipsen inequality $\rho_{.2}(R) \le 5\sqrt{n}$ when the columns of B are near enough to orthonormal, say $\|B'\cdot B - I\|_2 \le \frac{1}{2}$. Otherwise our example suggests that their $\sqrt{n}$ is deserved.