A new matrix algebra ?

Let $T_j(x)$ be the Chebycev polynomials

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_k(x) = 2T_{k-1}(x)x - T_{k-2}(x), \quad k = 2, 3, \ldots,$$

and recall their alternative representation in $[-1, 1]$

$$T_k(x) = \cos(k \arccos x), \qquad x \in [-1, 1].$$

Some of them: $T_2(x) = 2x^2 - 1$, $T_3(x) = 4x^3 - 3x$, $T_4(x) = 8x^4 - 8x^2 + 1$, ...

Let X be a $n \times n$ matrix with the property that the set $\mathcal{L}$ of all polynomials in X has dimension n (i.e. maximum dimension, since by Cayley-Hamilton theorem if $p_X(\lambda)$ is the characteristic polynomial of X then $p_X(X) = 0$). Usually an X with such property is called non-derogatory. Consider the Chebycev basis of $\mathcal{L}$:

$$\begin{aligned} &J_1 = T_0(X) = I, \quad J_2 = T_1(X) = X, \\ &J_{k+1} = T_k(X) = 2T_{k-1}(X)X - T_{k-2}(X), \quad k = 2, 3, \ldots, n-1. \end{aligned}$$

We are interested in cases where $A = \sum_k a_k J_k$ means $\mathbf{v}^T A = [a_1 \; a_2 \; \cdots \; a_n]$ for some vector $\mathbf{v}$.

For instance, choose

$$X = \begin{bmatrix} 0 & 1 & 0 \\ a & b & c \\ 0 & d & e \end{bmatrix}.$$

Then

$$\begin{aligned} J_1 = I, \; J_2 = X &= \begin{bmatrix} 0 & 1 & 0 \\ a & b & c \\ 0 & d & e \end{bmatrix}, \\ J_3 = T_2(X) &= 2T_1(X)X - T_0(X) = 2X^2 - I \\ &= 2 \begin{bmatrix} a & b & c \\ ba & a + b^2 + cd & bc + ce \\ da & db + ed & dc + e^2 \end{bmatrix} - I. \end{aligned}$$

Note that the first row of J_3 is $[2a - 1 \; 2b \; 2c]$ and thus is equal to $[0 \; 0 \; 1]$ iff $a = \frac{1}{2}$, $b = 0$, $c = \frac{1}{2}$. So, for these particular choices of a, b, c we have $\mathbf{e}_1^T J_k = \mathbf{e}_k^T$, $k = 1, 2, 3$, i.e. $A = \sum_{k=1}^3 a_k J_k$ means $\mathbf{e}_1^T A = [a_1 \; a_2 \; a_3]$. Moreover, since $a = \frac{1}{2}$, $b = 0$, $c = \frac{1}{2}$ imply

$$J_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & d & e \\ d & 2ed & d - 1 + 2e^2 \end{bmatrix},$$

we can say that the counter-identity matrix J is in $\mathcal{L}$ if $d = 1$ and $e = 0$. We rewrite the J_k in this case:

$$J_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; J_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}, \; J_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Note that the eigenvalues of $J_2 = X$ must be real and distinct (see the theory on eigenstructure of tridiagonal matrices). It is easy to obtain them: $-1, 0, 1$.

Let us generalize this example. Let X be a generic non-derogatory tridiagonal matrix with $\mathbf{e}_2^T = [0\ 1\ 0\ \cdots\ 0]$ as first row. Then the following result holds:

If $\mathbf{e}_1^T J_s = \mathbf{e}_s^T$, $s = 1, \ldots, k$, *then* $\mathbf{e}_1^T J_{k+1} = \mathbf{e}_{k+1}^T$ *iff the k-th row of X is* $[\cdots\ 0\ \frac{1}{2}\ 0\ \frac{1}{2}\ 0\ \cdots]$

In fact,

$$\begin{aligned} \mathbf{e}_1^T J_{k+1} &= \mathbf{e}_1^T T_k(X) = \mathbf{e}_1^T(2T_{k-1}(X)X - T_{k-2}(X)) = \mathbf{e}_1^T(2J_kX - J_{k-1}) \\ &= 2\mathbf{e}_k^T X - \mathbf{e}_{k-1}^T = [\cdots\ 0\ 2x_{kk-1}-1\ 2x_{kk}\ 2x_{kk+1}\ 0\ \cdots]. \end{aligned}$$

So, for

$$X = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & & \ddots & \\ & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & a & b \end{bmatrix}$$

we have $\mathbf{e}_1^T J_k = \mathbf{e}_k^T$, $k = 1, \ldots, n$, i.e. $A = \sum_{k=1}^n a_k J_k$ means $\mathbf{e}_1^T A = [a_1\ a_2\ \ldots\ a_n]$. Moreover, J is a polynomial in X or, equivalently, $J_n = J$ iff $a = 1$, $b = 0$ (proof: use the identity $JX = XJ$).

Now assume $a = 1$, $b = 0$. In this case the J_k have the form:

$$J_k = \begin{bmatrix} & & & & 1 & & & & & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} & & & & \\ & & \iddots & \iddots & \frac{1}{2} & \ddots & \ddots & & & \\ \frac{1}{2} & 0 & \iddots & & & \ddots & \ddots & \frac{1}{2} & & \\ & \frac{1}{2} & & & & & \frac{1}{2} & 0 & \frac{1}{2} & \\ & & \ddots & & & & 0 & \frac{1}{2} & & \\ & & & & & & \iddots & & & \\ & & & & & 1 & & & & \end{bmatrix}.$$

Moreover, the eigenvalues of $J_2 = X$ must be real and distinct (see the theory on eigenstructure of tridiagonal matrices). By using this fact one can obtain them in explicit form. They are $\cos(j\pi/(n-1))$, $j = 0, \ldots, n-1$.

Let us see why. Since $T_{n-1}(X) = J_n = J$ we have the equality $T_{n-1}(X)^2 = I$. Now, if $X\mathbf{v} = \lambda\mathbf{v}$ then $\mathbf{v} = T_{n-1}(X)^2\mathbf{v} = T_{n-1}(\lambda)^2\mathbf{v}$, in other words the eigenvalues λ of X must be zeros of the algebraic equation $T_{n-1}(x)^2 = 1$. For $x \in [-1, 1]$ this equation becomes $\cos^2((n-1)\arccos x) = 1$ whose zeros are: $x = \cos(2k\pi/(n-1))$ and $x = \cos((2k+1)\pi/(n-1))$, $k \in \mathbb{Z}$.

Eigenvectors of X ? Look for x_j such that

$$\begin{aligned} &x_2 = \cos\tfrac{j\pi}{n-1}x_1 \\ &\tfrac{1}{2}(x_1 + x_3) = \cos\tfrac{j\pi}{n-1}x_2 \\ &\cdots \\ &\tfrac{1}{2}(x_{n-2} + x_n) = \cos\tfrac{j\pi}{n-1}x_{n-1} \\ &x_{n-1} = \cos\tfrac{j\pi}{n-1}x_n \end{aligned}$$

The eigenvectors of 1 and -1 are $[1\ 1\ \cdots\ 1]$ and $[1\ \ -1\ 1\ \cdots\ (-1)^{n-1}]$, respectively. For $n=3$ and $n=4$:

$$\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 0 & \\ & & -1 \end{bmatrix},$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \frac{1}{2} & & \\ & & -\frac{1}{2} & \\ & & & -1 \end{bmatrix}$$

For a generic n ?

How Chebycev polynomials arise

Set $y(x) = x^n - p_{n-1}(x)$ where p_{n-1} is the unique degree-$(n-1)$ polynomial solving the minimum problem $\min_{p\in\mathbb{P}_{n-1}} \max_{[-1,1]} |x^n - p(x)|$. If $\mu = \max_{[-1,1]} |y(x)|$ then $y(x)$ assumes the values μ and $-\mu$ alternately in $n+1$ successive points $\{x_i\}_{i=0}^n$ of $[-1,1]$, $-1 \le x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n \le 1$ (see min-max approximation theory []). Obviously $y'(x_i) = 0$, $i = 1, \ldots, n-1$, whereas $y'(x_0)y'(x_n) \neq 0$ since $y'(x)$ is a polynomial of degree $n-1$. Thus $x_0 = -1$, $x_n = 1$. Consider now the function $y(x)^2 - \mu^2$. It is zero in all the x_i and its derivative, $2y(x)y'(x)$, is zero in $x_1, x_2, \ldots, x_{n-1}$. It follows that $y(x)^2 - \mu^2 = c(x^2-1)y'(x)^2$ for some real constant c. Noting that the coefficient of x^{2n} is on the left 1 and on the right cn^2 we conclude that

$$\frac{n^2}{1-x^2} = \frac{y'(x)^2}{\mu^2 - y(x)^2}, \quad \frac{n}{\sqrt{1-x^2}} = \pm\frac{y'(x)}{\sqrt{\mu^2 - y(x)^2}},$$

$y(x) = \mu\cos(n \arccos x + c)$. Finally, $y(1) = \pm\mu \Rightarrow c = k\pi \Rightarrow$

$$y(x) = x^n - p_{n-1}(x) = \pm\mu\cos(n\arccos x) =: \pm\mu T_n(x).$$

Properties of Chebycev polynomials

- $T_k(\lambda)|_{[-1,1]} = \cos(k \arccos \lambda)$
- $T_k(\lambda) = \frac{1}{2}[(\lambda - \sqrt{\lambda^2-1})^k + (\lambda + \sqrt{\lambda^2-1})^k]$
- $|T_k(\lambda)| \le 1$, $\lambda \in [-1,1]$
- $T_k(\cos\frac{i\pi}{k}) = (-1)^i$, $i = 0, 1, \ldots, k$
- $T_k(\cos\frac{(2j+1)\pi}{2k}) = 0$, $j = 0, 1, \ldots, k-1$
- $T_k(\lambda) \ge \lambda^k$, $\lambda \ge 1$
- $T_k(\frac{\lambda+1}{\lambda-1}) = \frac{1}{2}[(\frac{\sqrt{\lambda}+1}{\sqrt{\lambda}-1})^k + (\frac{\sqrt{\lambda}-1}{\sqrt{\lambda}+1})^k]$, $\lambda > 1$
- $T_k(\frac{\lambda+1}{\lambda-1}) > \frac{1}{2}(\frac{\sqrt{\lambda}+1}{\sqrt{\lambda}-1})^k$, $\lambda > 1$
- $0 < a < b$ and $t_k(\lambda) = T_k((b+a-2\lambda)/(b-a))/T_k((b+a)/(b-a))$ imply

  $$\min \max_{[a,b]} |p_k(\lambda)| = \max_{[a,b]} |t_k(\lambda)| = \frac{1}{T_k((b+a)/(b-a))}$$

  where the min is taken on all polynomials of type $a_k\lambda^k + \ldots + a_1\lambda + 1$, $a_k \neq 0$

- $T_{k+j}(\lambda) + T_{|k-j|}(\lambda) = 2T_k(\lambda)T_j(\lambda)$
- $\int_{-1}^{1} \frac{1}{\sqrt{1-\lambda^2}} T_k(\lambda)T_j(\lambda) d\lambda = \pi, \pi/2, 0,\ j = k = 0,\ j = k > 0,\ j \neq k$

Chebycev as characteristic polynomials

Write a semi-infinite matrix $X = [x_{ij}]_{i,j=1}^{+\infty}$ with the property: for all n the characteristic polynomial $p_n(\lambda)$ of the upper left $n \times n$ submatrix of X (X_n) is $T_n^*(\lambda) = T_n(\lambda)/(2^{n-1})$ where $T_n(\lambda)$ is the degree n Chebycev polynomial defined by ().

For the following choices of X,

$$X = \begin{bmatrix} 0 & 1/\sqrt{2} & 0 & 0 & \cdots \\ 1/\sqrt{2} & 0 & 1/2 & 0 & \cdots \\ 0 & 1/2 & 0 & \ddots & \\ 0 & 0 & \ddots & \ddots & \\ \vdots & \vdots & & & \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1/2 & 0 & 1/2 & 0 & \cdots \\ 0 & 1/2 & 0 & \ddots & \\ 0 & 0 & \ddots & \ddots & \\ \vdots & \vdots & & & \end{bmatrix},$$

$$X = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1/2 & 0 & 1 & 0 & \cdots \\ 0 & 1/4 & 0 & \ddots & \\ 0 & 0 & \ddots & \ddots & \\ \vdots & \vdots & & & \end{bmatrix},$$

we have $p_0(\lambda) = 1$, $p_1(\lambda) = \lambda = T_1(\lambda)$, $p_2(\lambda) = \lambda^2 - \frac{1}{2} = \frac{1}{2}(2\lambda^2 - 1) = \frac{1}{2}T_2(\lambda)$, $p_3(\lambda) = \lambda(\lambda^2 - \frac{1}{2}) - \frac{1}{4}\lambda = \lambda^3 - \frac{3}{4}\lambda = \frac{1}{4}(4\lambda^3 - 3\lambda) = \frac{1}{4}T_3(\lambda)$, …, $p_n(\lambda) = \lambda p_{n-1}(\lambda) - \frac{1}{4}p_{n-2}(\lambda) = \frac{1}{2^{n-1}}T_n(\lambda)$, ….

Proof: By induction:

$$\begin{aligned} \frac{1}{2^{n-1}}T_n(\lambda) &= \frac{1}{2^{n-1}}(2T_{n-1}(\lambda)\lambda - T_{n-2}(\lambda)) = \frac{1}{2^{n-1}}(2 \cdot 2^{n-2}p_{n-1}(\lambda)\lambda - 2^{n-3}p_{n-2}(\lambda)) \\ &= p_{n-1}(\lambda)\lambda - 2^{-2}p_{n-2}(\lambda) = p_n(\lambda) \end{aligned}$$

Notice that the third choice of X implies $\mathbf{e}_1^T p_k(X_n) = \mathbf{e}_{k+1}^T$, $k = 0, \ldots, n-1$ … for all three choices of X_n we refer to $\mathcal{L}$, the set of all polynomials in X_n, as Chebycev algebras … Each X is equal to DXD^{-1} for another X; since $p(DXD^{-1}) = Dp(X)D^{-1}$ from the eigenvectors $\mathbf{v}$ of one algebra we have easily the eigenvectors of the other algebras, they are $D\mathbf{v}$

SVD of $A \in \mathbb{C}^{n\times n}$ and how to compute the singular values of $A \in \mathbb{R}^{n\times n}$

If A is a $n \times n$ normal matrix then there exist matrices U, D, U unitary, $D = \text{diag}\,(\lambda_i)$ with $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n| \geq 0$, such that $A = UDU^*$. It follows that A admits the following *singular value* decomposition

$$A = U\,\text{diag}\,(|\lambda_i|)\,\text{diag}\,(e^{\mathbf{i}\arg(\lambda_i)})U^* = U\sigma V^*,$$
$$\sigma = \text{diag}\,(\sigma_i),\ \sigma_i = |\lambda_i|,\ V = U\,\text{diag}\,(e^{-\mathbf{i}\arg(\lambda_i)}).$$

However, *any $n \times n$ matrix A admits a singular value decomposition*, i.e. there exist unitary matrices U, V and $\sigma = \text{diag}\,(\sigma_i)$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ such that $A = U\sigma V^*$. The σ_i are the *singular values* of A.

Example. For $n = 1$ we have $a_{11} = 1 \cdot |a_{11}|(e^{-\mathbf{i}\arg(a_{11})})^*$.

Proof ...

By knowing the SVD of A we can do many things. In particular we have

(0) $|\det(A)| = \prod_{i=1}^n \sigma_i$

(1) $\sigma_{r+1} = \|A - \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^*\|_2 = \min\{\|A - B\|_2 :\ \mathrm{rank}(B) \le r\}$

(2) $\sqrt{\sum_{j=r+1}^n \sigma_j^2} = \|A - \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^*\|_F = \min\{\|A - B\|_F :\ \mathrm{rank}(B) \le r\}$

(3) $\sigma_n = \|A - \sum_{i=1}^{n-1} \sigma_i \mathbf{u}_i \mathbf{v}_i^*\|_2 = \min\{\|A - B\|_2 :\ \mathrm{rank}(B) \le n-1\} = \min\{\|A - B\|_2 :\ \det(B) = 0\}$

(4) $\|A\|_2 = \sigma_1$, $\|A\|_F = \sqrt{\sum_{i=1}^n \sigma_i^2}$

(5) $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are the eigenvalues of A^*A

(6) $\det(A) \ne 0 \Rightarrow \|A^{-1}\|_2 = 1/\sigma_n$, $\mu_2(A) = \sigma_1/\sigma_n$, $\mu_2(A^*A) = \mu_2(A)^2$

(7) If λ_i are the eigenvalues of A, then $\sigma_n \le |\lambda_i| \le \sigma_1$. If A is normal then $\sigma_i = |\lambda_i|$

(8) If $\sigma_1 \ge \ldots \ge \sigma_k > 0 = \sigma_{k+1} = \ldots = \sigma_n$ then the kernel and the image of A can be represented as follows: $\{\mathbf{x} \in \mathbb{C}^n :\ A\mathbf{x} = \mathbf{0}\} = \mathrm{Span}\,\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$, $\{A\mathbf{x} :\ \mathbf{x} \in \mathbb{C}^n\} = \mathrm{Span}\,\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$

How to compute U, σ, V such that $A = U\sigma V^*$? An algorithm that works for A real (note that in this case U, V can be chosen real unitary (orthonormal)) consists in the following two steps (1) and (2):

Step (1). Transform A into a bidiagonal matrix

$$QAZ = B = \begin{bmatrix} a_1 & b_1 & & & \\ 0 & a_2 & b_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & b_{n-1} \\ & & & & a_n \end{bmatrix}$$

by using orthonormal transforms Q and Z.

For $n = 1$: $1 \cdot a_{11} \cdot 1 = a_{11}$.

For $n = 2$, if $\alpha = a_{11}/\sqrt{a_{11}^2 + a_{21}^2}$, $\beta = -a_{21}/\sqrt{a_{11}^2 + a_{21}^2}$, then

$$\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha a_{11} - \beta a_{21} & \alpha a_{12} - \beta a_{22} \\ \beta a_{11} + \alpha a_{21} & \beta a_{12} + \alpha a_{22} \end{bmatrix}$$
$$= \begin{bmatrix} \sqrt{a_{11}^2 + a_{21}^2} & \frac{a_{11}a_{12} + a_{21}a_{22}}{\sqrt{a_{11}^2 + a_{21}^2}} \\ 0 & \frac{-a_{21}a_{12} + a_{11}a_{22}}{\sqrt{a_{11}^2 + a_{21}^2}} \end{bmatrix}.$$

For $n > 2$ ... example, $n = 4$:

$$A = \begin{bmatrix} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{bmatrix}, \quad S_{12}^T A = \begin{bmatrix} \square & \square & \square & \square \\ 0 & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{bmatrix},$$

$$S_{13}^T(S_{12}^TA) = \begin{bmatrix} \square & \square & \square & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \\ \square & \square & \square & \square \end{bmatrix}, \; S_{14}^T(S_{13}^TS_{12}^TA) = \begin{bmatrix} \square & \square & \square & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \end{bmatrix},$$

$$(S_{14}^TS_{13}^TS_{12}^TA)S_{23} = \begin{bmatrix} \square & \square & 0 & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \end{bmatrix}, \; (S_{14}^TS_{13}^TS_{12}^TAS_{23})S_{24} = \begin{bmatrix} \square & \square & 0 & 0 \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \\ 0 & \square & \square & \square \end{bmatrix},$$

$$S_{23}'^T(S_{14}^TS_{13}^TS_{12}^TAS_{23}S_{24}) = \begin{bmatrix} \square & \square & 0 & 0 \\ 0 & \square & \square & \square \\ 0 & 0 & \square & \square \\ 0 & \square & \square & \square \end{bmatrix}, S_{24}'^T(S_{23}'^TS_{14}^TS_{13}^TS_{12}^TAS_{23}S_{24}) = \begin{bmatrix} \square & \square & 0 & 0 \\ 0 & \square & \square & \square \\ 0 & 0 & \square & \square \\ 0 & 0 & \square & \square \end{bmatrix},$$

$$(S_{24}'^TS_{23}'^TS_{14}^TS_{13}^TS_{12}^TAS_{23}S_{24})S_{34} = \begin{bmatrix} \square & \square & 0 & 0 \\ 0 & \square & \square & 0 \\ 0 & 0 & \square & \square \\ 0 & 0 & \square & \square \end{bmatrix},$$

$$S_{34}'^T(S_{24}'^TS_{23}'^TS_{14}^TS_{13}^TS_{12}^TAS_{23}S_{24}S_{34}) = \begin{bmatrix} \square & \square & 0 & 0 \\ 0 & \square & \square & 0 \\ 0 & 0 & \square & \square \\ 0 & 0 & 0 & \square \end{bmatrix} = B.$$

Thus $B = QAZ$, $Q = S_{34}'^TS_{24}'^TS_{23}'^TS_{14}^TS_{13}^TS_{12}^T$, $Z = S_{23}S_{24}S_{34}$ ($Q^T = Q^{-1}$, $Z^T = Z^{-1}$). Note that the Givens transformations (plane rotations)

$$S_{ij} = S_{ji} = \begin{bmatrix} I & & & & \\ & \alpha & & \beta & \\ & & I & & \\ & -\beta & & \alpha & \\ & & & & I \end{bmatrix}, \quad \alpha^2 + \beta^2 = 1,$$

used to liquidate the not-on-bidiagonal-part entries are chosen so that they leave unchanged the previously posed zeros. Moreover, S_{ij} is used to liquidate (i, j), $i > j$, when multiplying on the left, and S_{i+1j} is used to liquidate (i, j), $i < j-1$, when multiplying on the right.

Exercise (Elisa Sallicandro). Transform the matrix

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 3 \\ 4 & 0 & 0 & 1 \end{bmatrix}$$

into a bidiagonal matrix B.

Step (2). Set $A_1 := B$ and define a sequence of matrices $\{A_j\}_{j=1}^{+\infty}$ via the rule: for $k = 2, 4, 6, \ldots$

$$A_{k-1} = \begin{bmatrix} a_1^{k-1} & b_1^{k-1} & & \\ 0 & a_2^{k-1} & \ddots & \\ & & \ddots & b_{n-1}^{k-1} \\ & & & a_n^{k-1} \end{bmatrix} \to A_k = A_{k-1}Z_{k-1} = \begin{bmatrix} a_1^k & & & \\ b_1^k & a_2^k & & \\ & \ddots & \ddots & \\ & & b_{n-1}^k & a_n^k \end{bmatrix}$$

(Z_{k-1} is the product of the $n-1$ plane rotations used to liquidate the entries $(i, i+1)$ of A_{k-1}),

$$A_k \;\to\; A_{k+1} = Q_k A_k = \begin{bmatrix} a_1^{k+1} & b_1^{k+1} & & \\ 0 & a_2^{k+1} & \ddots & \\ & & \ddots & b_{n-1}^{k+1} \\ & & & a_n^{k+1} \end{bmatrix}$$

(Q_k is the product of the $n-1$ plane rotations used to liquidate the entries $(i+1, i)$ of A_k).

Then $b_i^j \to 0$ if $j \to +\infty$ ($i = 1, \ldots, n-1$). (Question: one should also prove that the a_i^j, $i = 1, \ldots, n$, have non-negative limit)

Proof: Let us show that $b_{n-1}^j \to 0$ if $j \to +\infty$.

The euclidean norm of the i-th column of A_k is equal to the euclidean norm of the i-th column of A_{k+1}. Thus

$$\begin{array}{l} a_1(k)^2 + b_1(k)^2 = a_1(k+1)^2 \\ b_2(k)^2 + a_2(k)^2 = a_2(k+1)^2 + b_1(k+1)^2 \\ \cdots \\ a_n(k)^2 = a_n(k+1)^2 + b_{n-1}(k+1)^2 \end{array}$$

Moreover, the euclidean norm of the n-th row of A_{k-1} is equal to the euclidean norm of the n-row of A_k. Thus

$$\begin{array}{rcl} \|B\|_F^2 = \|A_{k+1}\|_F^2 \geq a_n(k+1)^2 & = & a_n(k-1)^2 - b_{n-1}(k+1)^2 - b_{n-1}(k)^2 \\ = \quad \ldots & = & a_n(1)^2 - \sum_{j=2}^{k+1} b_{n-1}(j)^2 \geq 0. \end{array}$$

But this implies $\sum_{j=1}^{+\infty} b_{n-1}(j)^2 < +\infty$, and we have the thesis.

($b_{n-2}^j \to 0$: The euclidean norm of the $(n-1)$-th row of A_{k-1} is equal to the euclidean norm of the $(n-1)$-th row of A_k. Thus

$$a_{n-1}(k+1)^2 = a_{n-1}(1)^2 + \sum_{j=1}^{k} b_{n-1}(j)^2 - \sum_{j=2}^{k+1} b_{n-2}(j)^2$$

$\Rightarrow \;\; \sum_{j=1}^{+\infty} b_{n-2}(j)^2 < +\infty \;\;\Rightarrow\;\; b_{n-2}(j) \to 0$ if $j \to \infty$. ...)

Step (2) for $n = 2$ (convergence).

$$A_{k-1} = \begin{bmatrix} a_1^{k-1} & b_1^{k-1} \\ 0 & a_2^{k-1} \end{bmatrix} \to A_k = A_{k-1} Z_{k-1} = \begin{bmatrix} a_1^k & 0 \\ b_1^k & a_2^k \end{bmatrix} \to A_{k+1} = Q_k A_k = \begin{bmatrix} a_1^{k+1} & b_1^{k+1} \\ 0 & a_2^{k+1} \end{bmatrix},$$

$$\begin{array}{l} \|A_k \mathbf{e}_i\|_2 = \|A_{k+1} \mathbf{e}_i\|_2 \Rightarrow \\ a_1(k)^2 + b_1(k)^2 = a_1(k+1)^2, \\ a_2(k)^2 = b_1(k+1)^2 + a_2(k+1)^2, \end{array}$$

$$\begin{array}{l} \|A_{k-1}^T \mathbf{e}_2\|_2 = \|A_k^T \mathbf{e}_2\|_2 \Rightarrow \\ a_2(k-1)^2 = b_1(k)^2 + a_2(k)^2. \end{array}$$

Thus

$$\begin{array}{rcl} \|B\|_F^2 & = & \|A_{k+1}\|_F^2 \geq a_2(k+1)^2 = a_2(k-1)^2 - b_1(k)^2 - b_1(k+1)^2 \\ & = & a_2(1)^2 - \sum_{j=2}^{k+1} b_1(j)^2 \Rightarrow b_1(j) \to 0. \end{array}$$

Step (2) for $n = 2$ (details and example). Given an upper triangular (bidiagonal) 2×2 matrix

$$A = \begin{bmatrix} a_1 & b_1 \\ 0 & a_2 \end{bmatrix},$$

write an algorithm to compute its singular values σ_1, σ_2. (Notice however that σ_1, σ_2 are simply the squaring roots of the eigenvalues of

$$\begin{bmatrix} \overline{a_1} & 0 \\ \overline{b_1} & \overline{a_2} \end{bmatrix} \begin{bmatrix} a_1 & b_1 \\ 0 & a_2 \end{bmatrix} = \begin{bmatrix} |a_1|^2 & \overline{a_1} b_1 \\ \overline{b_1} a_1 & |b_1|^2 + |a_2|^2 \end{bmatrix}, \ i.e. \ \sqrt{\tfrac{1}{2}(|a_1|^2 + |a_2|^2 + |b_1|^2 \pm \sqrt{(|a_1|^2 + |a_2|^2 + |b_1|^2)^2 - 4|a_1|^2|a_2|^2)}}$$

).

Solution. Set $A_1 = A$ and, for $k = 2, 4, \ldots$

$$Z_{k-1} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \quad \alpha = \frac{a_1}{\sqrt{a_1^2 + b_1^2}}, \ \beta = \frac{-b_1}{\sqrt{a_1^2 + b_1^2}}.$$

Then

$$A_k = A_{k-1} Z_{k-1} = \begin{bmatrix} a_1\alpha - b_1\beta & a_1\beta + b_1\alpha \\ -a_2\beta & a_2\alpha \end{bmatrix} = \begin{bmatrix} \sqrt{a_1^2 + b_1^2} & 0 \\ \frac{a_2 b_1}{\sqrt{a_1^2+b_1^2}} & \frac{a_2 a_1}{\sqrt{a_1^2+b_1^2}} \end{bmatrix}.$$

Now set

$$Q_k = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}, \quad \alpha = \frac{\sqrt{a_1^2 + b_1^2}}{\sqrt{a_1^2 + b_1^2 + \frac{a_2^2 b_1^2}{a_1^2+b_1^2}}}, \ \beta = \frac{\frac{-a_2 b_1}{\sqrt{a_1^2+b_1^2}}}{\sqrt{a_1^2 + b_1^2 + \frac{a_2^2 b_1^2}{a_1^2+b_1^2}}}.$$

Then

$$\begin{aligned} A_{k+1} &= Q_k A_k = Q_k A_{k-1} Z_{k-1} \\ &= \begin{bmatrix} \alpha\sqrt{a_1^2 + b_1^2} - \beta\frac{a_2 b_1}{\sqrt{a_1^2+b_1^2}} & -\beta\frac{a_2 a_1}{\sqrt{a_1^2+b_1^2}} \\ \beta\sqrt{a_1^2 + b_1^2} + \alpha\frac{a_2 b_1}{\sqrt{a_1^2+b_1^2}} & \alpha\frac{a_2 a_1}{\sqrt{a_1^2+b_1^2}} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{a_1^2 + b_1^2 + \frac{a_2^2 b_1^2}{a_1^2+b_1^2}} & \frac{a_2^2 a_1 b_1}{(a_1^2+b_1^2)\sqrt{a_1^2+b_1^2+\frac{a_2^2 b_1^2}{a_1^2+b_1^2}}} \\ 0 & \frac{a_2 a_1}{\sqrt{a_1^2+b_1^2+\frac{a_2^2 b_1^2}{a_1^2+b_1^2}}} \end{bmatrix}. \end{aligned}$$

The algorithm

10 $a_1 =?\ a_2 =?\ b_1 =?$

20 $a_1^{new} = \sqrt{a_1^2 + b_1^2 + \frac{a_2^2 b_1^2}{a_1^2+b_1^2}}$

30 $a_2^{new} = \frac{a_2 a_1}{\sqrt{a_1^2+b_1^2+\frac{a_2^2 b_1^2}{a_1^2+b_1^2}}}$

40 $b_1^{new} = \frac{a_2^2 a_1 b_1}{(a_1^2+b_1^2)\sqrt{a_1^2+b_1^2+\frac{a_2^2 b_1^2}{a_1^2+b_1^2}}}$

50 $a_1 = a_1^{new};\ a_2 = a_2^{new};\ b_1 = b_1^{new};$ GOTO 20

should generate a sequence of b_1 convergent to 0, and sequences of a_1 and a_2 convergent to the singular values. (Note that $a_1^{k+1} a_2^{k+1} = |\det(A_{k+1})| = |\det(A_{k-1})| = |\det(A)| = \sigma_1\sigma_2 = a_1 a_2$).

An implementation of the algorithm:

$$
\begin{array}{ll}
 & a_1 =?;\ a_2 =?;\ b_1 =? \\
20 & x = a_1^2 + b_1^2 \\
 & y = a_2 b_1 \\
 & z = a_2 a_1 \\
 & b_1^{new} = y/x \\
 & a_1^{new} = \sqrt{x + y * b_1^{new}} \\
 & a_2^{new} = z/a_1^{new} \\
 & b_1^{new} = b_1^{new} * a_2^{new} \\
 & a_1 = a_1^{new};\ a_2 = a_2^{new};\ b_1 = b_1^{new};\ \text{GOTO } 20
\end{array}
$$

Example. If $a_1 = a_2 = 1$, $b_1 = -1$, then

$$\sigma_1 = \sqrt{\frac{3+\sqrt{5}}{2}}, \quad \sigma_2 = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

Let us do some steps of the proposed algorithm. (Note that $a_1^{k+1} a_2^{k+1} = |\det(A_{k+1})| = |\det(A_{k-1})| = |\det(A)| = \sigma_1\sigma_2 = 1$).

a_1	1	$\sqrt{\frac{5}{2}}$	$\sqrt{\frac{34}{13}}$
a_2	1	$\sqrt{\frac{2}{5}}$	$\sqrt{\frac{13}{34}}$
b_1	-1	$-\frac{1}{\sqrt{10}}$	$-\frac{1}{\sqrt{13\cdot 34}}$
$x = a_1^2 + b_1^2$	2	$\frac{13}{5}$	$\frac{89}{34}$
$y = a_2 b_1$	-1	$-\frac{1}{5}$	$-\frac{1}{34}$
$z = a_2 a_1$	1	1	1
$b_1^{new} = y/x$	$-\frac{1}{2}$	$-\frac{1}{13}$	$-\frac{1}{89}$
$a_1^{new} = \sqrt{x + y * b_1^{new}}$	$\sqrt{\frac{5}{2}}$	$\sqrt{\frac{34}{13}}$	$\sqrt{\frac{233}{89}}$
$a_2^{new} = z/a_1^{new}$	$\sqrt{\frac{2}{5}}$	$\sqrt{\frac{13}{34}}$	$\sqrt{\frac{89}{233}}$
$b_1^{new} = b_1^{new} * a_2^{new}$	$-\frac{1}{\sqrt{10}}$	$-\frac{1}{\sqrt{13\cdot 34}}$	$-\frac{1}{\sqrt{89\cdot 233}}$

We should have $a_1 \to \sigma_1$, $a_2 \to \sigma_2$, $b_1 \to 0$, and this is the case: for instance we have $\frac{5}{2} = 2.5$, $\frac{34}{13} = 2.615$, $\frac{233}{89} = 2.6179, \ldots \to \sigma_1^2 = 2.61803$ ($\sqrt{5} \approx 2.23607$).