12
Singular Value Decomposition (SVD) and Polar Form

12.1 Polar Form

In this section we assume that we are dealing with a real Euclidean space E. Let $f\colon E \to E$ be any linear map. In general, it may not be possible to diagonalize a linear map f. However, note that $f^* \circ f$ is self-adjoint, since

$$\langle (f^* \circ f)(u), v\rangle = \langle f(u), f(v)\rangle = \langle u, (f^* \circ f)(v)\rangle.$$

Similarly, $f \circ f^*$ is self-adjoint.

The fact that $f^* \circ f$ and $f \circ f^*$ are self-adjoint is very important, because it implies that $f^* \circ f$ and $f \circ f^*$ can be diagonalized and that they have real eigenvalues. In fact, these eigenvalues are all nonnegative. Indeed, if u is an eigenvector of $f^* \circ f$ for the eigenvalue λ, then

$$\langle (f^* \circ f)(u), u\rangle = \langle f(u), f(u)\rangle$$

and

$$\langle (f^* \circ f)(u), u\rangle = \lambda\langle u, u\rangle,$$

and thus

$$\lambda\langle u, u\rangle = \langle f(u), f(u)\rangle,$$

which implies that $\lambda \geq 0$, since $\langle -, -\rangle$ is positive definite. A similar proof applies to $f \circ f^*$. Thus, the eigenvalues of $f^* \circ f$ are of the form $\mu_1^2, \ldots, \mu_r^2$ or 0, where $\mu_i > 0$, and similarly for $f \circ f^*$. The situation is even better, since we will show shortly that $f^* \circ f$ and $f \circ f^*$ have the same eigenvalues.

Remark: Given any two linear maps $f: E \to F$ and $g: F \to E$, where $\dim(E) = n$ and $\dim(F) = m$, it can be shown that

$$(-\lambda)^m \det(g \circ f - \lambda I_n) = (-\lambda)^n \det(f \circ g - \lambda I_m),$$

and thus $g \circ f$ and $f \circ g$ always have the same nonnull eigenvalues!

The square roots $\mu_i > 0$ of the positive eigenvalues of $f^* \circ f$ (and $f \circ f^*$) are called the *singular values of* f. A self-adjoint linear map $f: E \to E$ whose eigenvalues are nonnegative is called *positive*, and if f is also invertible, *positive definite*. In the latter case, every eigenvalue is strictly positive. We just showed that $f^* \circ f$ and $f \circ f^*$ are positive self-adjoint linear maps.

The wonderful thing about the singular value decomposition is that there exist two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ such that with respect to these bases, f is a diagonal matrix consisting of the singular values of f, or 0. First, we show some useful relationships between the kernels and the images of f, f^*, $f^* \circ f$, and $f \circ f^*$. Recall that if $f: E \to F$ is a linear map, the *image* $\mathrm{Im}\, f$ *of* f is the subspace $f(E)$ of F, and the *rank of* f is the dimension $\dim(\mathrm{Im}\, f)$ of its image. Also recall that

$$\dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(E),$$

and that for every subspace W of E

$$\dim(W) + \dim(W^\perp) = \dim(E).$$

Lemma 12.1.1 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for any linear map $f: E \to F$, we have*

$$\begin{aligned}
\mathrm{Ker}\, f &= \mathrm{Ker}\,(f^* \circ f),\\
\mathrm{Ker}\, f^* &= \mathrm{Ker}\,(f \circ f^*),\\
\mathrm{Ker}\, f &= (\mathrm{Im}\, f^*)^\perp,\\
\mathrm{Ker}\, f^* &= (\mathrm{Im}\, f)^\perp,\\
\dim(\mathrm{Im}\, f) &= \dim(\mathrm{Im}\, f^*),\\
\dim(\mathrm{Ker}\, f) &= \dim(\mathrm{Ker}\, f^*),
\end{aligned}$$

and f, f^*, $f^* \circ f$, and $f \circ f^*$ have the same rank.

Proof. To simplify the notation, we will denote the inner products on E and F by the same symbol $\langle -, - \rangle$ (to avoid subscripts). If $f(u) = 0$, then $(f^* \circ f)(u) = f^*(f(u)) = f^*(0) = 0$, and so $\mathrm{Ker}\, f \subseteq \mathrm{Ker}\,(f^* \circ f)$. By definition of f^*, we have

$$\langle f(u), f(u) \rangle = \langle (f^* \circ f)(u), u \rangle$$

for all $u \in E$. If $(f^* \circ f)(u) = 0$, since $\langle -, - \rangle$ is positive definite, we must have $f(u) = 0$, and so $\mathrm{Ker}\,(f^* \circ f) \subseteq \mathrm{Ker}\, f$. Therefore,

$$\mathrm{Ker}\, f = \mathrm{Ker}\,(f^* \circ f).$$

The proof that $\mathrm{Ker}\, f^* = \mathrm{Ker}\,(f \circ f^*)$ is similar.

By definition of f^*, we have

$$\langle f(u), v \rangle = \langle u, f^*(v) \rangle$$

for all $u \in E$ and all $v \in F$. This immediately implies that

$$\mathrm{Ker}\, f = (\mathrm{Im}\, f^*)^{\perp} \quad \text{and} \quad \mathrm{Ker}\, f^* = (\mathrm{Im}\, f)^{\perp}.$$

Since

$$\dim(\mathrm{Im}\, f) = n - \dim(\mathrm{Ker}\, f)$$

and

$$\dim((\mathrm{Im}\, f^*)^{\perp}) = n - \dim(\mathrm{Im}\, f^*),$$

from

$$\mathrm{Ker}\, f = (\mathrm{Im}\, f^*)^{\perp}$$

we also have

$$\dim(\mathrm{Ker}\, f) = \dim((\mathrm{Im}\, f^*)^{\perp}),$$

from which we obtain

$$\dim(\mathrm{Im}\, f) = \dim(\mathrm{Im}\, f^*).$$

The above immediately implies that $\dim(\mathrm{Ker}\, f) = \dim(\mathrm{Ker}\, f^*)$. From all this we easily deduce that

$$\dim(\mathrm{Im}\, f) = \dim(\mathrm{Im}\,(f^* \circ f)) = \dim(\mathrm{Im}\,(f \circ f^*)),$$

i.e., f, f^*, $f^* \circ f$, and $f \circ f^*$ have the same rank. $\square$

The next lemma shows a very useful property of positive self-adjoint linear maps.

Lemma 12.1.2 *Given a Euclidean space E of dimension n, for any positive self-adjoint linear map $f: E \to E$ there is a unique positive self-adjoint linear map $h: E \to E$ such that $f = h^2 = h \circ h$. Furthermore,* $\mathrm{Ker}\, f = \mathrm{Ker}\, h$, *and if $\mu_1, \ldots, \mu_p$ are the distinct eigenvalues of h and E_i is the eigenspace associated with μ_i, then $\mu_1^2, \ldots, \mu_p^2$ are the distinct eigenvalues of f, and E_i is the eigenspace associated with μ_i^2.*

Proof. Since f is self-adjoint, by Theorem 11.3.1 there is an orthonormal basis $(u_1, \ldots, u_n)$ consisting of eigenvectors of f, and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f, we know that $\lambda_i \in \mathbb{R}$. Since f is assumed to be positive, we have $\lambda_i \geq 0$, and we can write $\lambda_i = \mu_i^2$, where $\mu_i \geq 0$. If we define $h\colon E \to E$ by its action on the basis $(u_1, \ldots, u_n)$, so that

$$h(u_i) = \mu_i u_i,$$

it is obvious that $f = h^2$ and that h is positive self-adjoint (since its matrix over the orthonormal basis $(u_1, \ldots, u_n)$ is diagonal, thus symmetric). It remains to prove that h is uniquely determined by f. Let $g\colon E \to E$ be any positive self-adjoint linear map such that $f = g^2$. Then there is an orthonormal basis $(v_1, \ldots, v_n)$ of eigenvectors of g, and let $\mu_1, \ldots, \mu_n$ be the eigenvalues of g, where $\mu_i \geq 0$. Note that

$$f(v_i) = g^2(v_i) = g(g(v_i)) = \mu_i^2 v_i,$$

so that v_i is an eigenvector of f for the eigenvalue μ_i^2. If $\mu_1, \ldots, \mu_p$ are the distinct eigenvalues of g and $E_1, \ldots, E_p$ are the corresponding eigenspaces, the above argument shows that each E_i is a subspace of the eigenspace U_i of f associated with μ_i^2. However, we observed (just after Theorem 11.3.1) that

$$E = E_1 \oplus \cdots \oplus E_p,$$

where E_i and E_j are orthogonal if $i \neq j$, and thus we must have $E_i = U_i$. Since $\mu_i, \mu_j \geq 0$ and $\mu_i \neq \mu_j$ implies that $\mu_i^2 \neq \mu_j^2$, the values $\mu_1^2, \ldots, \mu_p^2$ are the distinct eigenvalues of f, and the corresponding eigenspaces are also $E_1, \ldots, E_p$. This shows that $g = h$, and h is unique. Also, as a consequence, $\operatorname{Ker} f = \operatorname{Ker} h$, and if $\mu_1, \ldots, \mu_p$ are the distinct eigenvalues of h, then $\mu_1^2, \ldots, \mu_p^2$ are the distinct eigenvalues of f, and the corresponding eigenspaces are identical. $\square$

There are now two ways to proceed. We can prove directly the singular value decomposition, as Strang does [166, 165], or prove the so-called *polar decomposition* theorem. The proofs are of roughly the same difficulty. We have chosen the second approach, since it is less common in textbook presentations, and since it also yields a little more, namely uniqueness when f is invertible. It is somewhat disconcerting that the next two theorems are given only as an exercise in Bourbaki [20] (Algèbre, Chapter 9, Problem 14, page 127). Yet, the SVD decomposition is of great practical importance. This is probably typical of the attitude of "pure mathematicians." However, the proof hinted at in Bourbaki is quite elegant.

The early history of the singular value decomposition is described in a fascinating paper by Stewart [162]. The SVD is due to Beltrami and Camille Jordan independently (1873, 1874). Gauss is the grandfather of all this, for his work on least squares (1809, 1823) (but Legendre also published a paper on least squares!). Then come Sylvester, Schmidt, and Hermann Weyl. Sylvester's work was apparently "opaque." He gave a computational method to find an SVD. Schimdt's work really has to do with integral equations and symmetric and asymmetric kernels (1907). Weyl's work has to do with perturbation theory (1912). Autonne came up with the polar decomposition (1902, 1915). Eckart and Young extended SVD to rectangular matrices (1936, 1939).

The next three theorems deal with a linear map $f\colon E \to E$ over a Euclidean space E. We will show later on how to generalize these results to linear maps $f\colon E \to F$ between two Euclidean spaces E and F.

Theorem 12.1.3 *Given a Euclidean space E of dimension n, for any linear map $f\colon E \to E$ there are two positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon E \to E$ and an orthogonal linear map $g\colon E \to E$ such that*

$$f = g \circ h_1 = h_2 \circ g.$$

Furthermore, if f has rank r, the maps h_1 and h_2 have the same positive eigenvalues $\mu_1, \ldots, \mu_r$, which are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of both $f^* \circ f$ and $f \circ f^*$. Finally, h_1, h_2 are unique, g is unique if f is invertible, and $h_1 = h_2$ if f is normal.

Proof. By Lemma 12.1.2 there are two (unique) positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon E \to E$ such that $f^* \circ f = h_1^2$ and $f \circ f^* = h_2^2$. Note that

$$\langle f(u), f(v)\rangle = \langle h_1(u), h_1(v)\rangle$$

for all $u, v \in E$, since

$$\langle f(u), f(v)\rangle = \langle u, (f^* \circ f)(v)\rangle = \langle u, (h_1 \circ h_1)(v)\rangle = \langle h_1(u), h_1(v)\rangle,$$

because $f^* \circ f = h_1^2$ and $h_1 = h_1^*$ (h_1 is self-adjoint). From Lemma 12.1.1, $\mathrm{Ker}\, f = \mathrm{Ker}\,(f^* \circ f)$, and from Lemma 12.1.2, $\mathrm{Ker}\,(f^* \circ f) = \mathrm{Ker}\, h_1$. Thus,

$$\mathrm{Ker}\, f = \mathrm{Ker}\, h_1.$$

If r is the rank of f, then since h_1 is self-adjoint, by Theorem 11.3.1 there is an orthonormal basis $(u_1, \ldots, u_n)$ of eigenvectors of h_1, and by reordering these vectors if necessary, we can assume that $(u_1, \ldots, u_r)$ are associated with the strictly positive eigenvalues $\mu_1, \ldots, \mu_r$ of h_1 (the singular values of f), and that $\mu_{r+1} = \cdots = \mu_n = 0$. Observe that $(u_{r+1}, \ldots, u_n)$ is an orthonormal basis of $\mathrm{Ker}\, f = \mathrm{Ker}\, h_1$, and that $(u_1, \ldots, u_r)$ is an orthonormal basis of $(\mathrm{Ker}\, f)^\perp = \mathrm{Im}\, f^*$. Note that

$$\langle f(u_i), f(u_j)\rangle = \langle h_1(u_i), h_1(u_j)\rangle = \mu_i\mu_j\langle u_i, u_j\rangle = \mu_i^2\delta_{i\,j}$$

when $1 \leq i, j \leq n$ (recall that $\delta_{i\,j} = 1$ if $i = j$, and $\delta_{i\,j} = 0$ if $i \neq j$). Letting

$$v_i = \frac{f(u_i)}{\mu_i}$$

when $1 \leq i \leq r$, observe that

$$\langle v_i, v_j\rangle = \delta_{i\,j}$$

when $1 \leq i, j \leq r$. Using the Gram–Schmidt orthonormalization procedure, we can extend $(v_1, \ldots, v_r)$ to an orthonormal basis $(v_1, \ldots, v_n)$ of E (even when $r = 0$). Also note that $(v_1, \ldots, v_r)$ is an orthonormal basis of $\operatorname{Im} f$, and $(v_{r+1}, \ldots, v_n)$ is an orthonormal basis of $\operatorname{Im} f^{\perp} = \operatorname{Ker} f^*$.

We define the linear map $g\colon E \to E$ by its action on the basis $(u_1, \ldots, u_n)$ as follows:

$$g(u_i) = v_i$$

for all i, $1 \leq i \leq n$. We have

$$(g \circ h_1)(u_i) = g(h_1(u_i)) = g(\mu_i u_i) = \mu_i g(u_i) = \mu_i v_i = \mu_i \frac{f(u_i)}{\mu_i} = f(u_i)$$

when $1 \leq i \leq r$, and

$$(g \circ h_1)(u_i) = g(h_1(u_i)) = g(0) = 0$$

when $r + 1 \leq i \leq n$ (since $(u_{r+1}, \ldots, u_n)$ is a basis for $\operatorname{Ker} f = \operatorname{Ker} h_1$), which shows that $f = g \circ h_1$. The fact that g is orthogonal follows easily from the fact that it maps the orthonormal basis $(u_1, \ldots, u_n)$ to the orthonormal basis $(v_1, \ldots, v_n)$.

We can show that $f = h_2 \circ g$ as follows. Notice that

$$\begin{aligned}
h_2^2(v_i) &= (f \circ f^*)\left(\frac{f(u_i)}{\mu_i}\right), \\
&= (f \circ (f^* \circ f))\left(\frac{u_i}{\mu_i}\right), \\
&= \frac{1}{\mu_i}(f \circ h_1^2)(u_i), \\
&= \frac{1}{\mu_i} f(h_1^2(u_i)), \\
&= \frac{1}{\mu_i} f(\mu_i^2 u_i), \\
&= \mu_i f(u_i), \\
&= \mu_i^2 v_i
\end{aligned}$$

when $1 \leq i \leq r$, and

$$h_2^2(v_i) = (f \circ f^*)(v_i) = f(f^*(v_i)) = 0$$

when $r + 1 \leq i \leq n$, since $(v_{r+1}, \ldots, v_n)$ is a basis for $\operatorname{Ker} f^* = (\operatorname{Im} f)^{\perp}$. Since h_2 is positive self-adjoint, so is h_2^2, and by Lemma 12.1.2, we must have

$$h_2(v_i) = \mu_i v_i$$

when $1 \leq i \leq r$, and

$$h_2(v_i) = 0$$

when $r+1 \leq i \leq n$. This shows that $(v_1, \ldots, v_n)$ are eigenvectors of h_2 for $\mu_1, \ldots, \mu_n$ (since $\mu_{r+1} = \cdots = \mu_n = 0$), and thus h_1 and h_2 have the same eigenvalues $\mu_1, \ldots, \mu_n$.

As a consequence,

$$(h_2 \circ g)(u_i) = h_2(g(u_i)) = h_2(v_i) = \mu_i v_i = f(u_i)$$

when $1 \leq i \leq n$. Since h_1, h_2, $f^* \circ f$, and $f \circ f^*$ are positive self-adjoint, $f^* \circ f = h_1^2$, $f \circ f^* = h_2^2$, and $\mu_1, \ldots, \mu_r$ are the eigenvalues of both h_1 and h_2, it follows that $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of both $f^* \circ f$ and $f \circ f^*$.

Finally, since

$$f^* \circ f = h_1^2 \quad \text{and} \quad f \circ f^* = h_2^2,$$

by Lemma 12.1.2, h_1 and h_2 are unique and if f is invertible, then h_1 and h_2 are invertible and thus g is also unique, since $g = f \circ h_1^{-1}$. If h is normal, then $f^* \circ f = f \circ f^*$ and $h_1 = h_2$. $\square$

In matrix form, Theorem 12.1.3 can be stated as follows. For every real $n \times n$ matrix A, there is some orthogonal matrix R and some positive symmetric matrix S such that

$$A = RS.$$

Furthermore, R, S are unique if A is invertible. A pair (R, S) such that $A = RS$ is called a *polar decomposition of* A. For example, the matrix

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

is both orthogonal and symmetric, and $A = RS$ with $R = A$ and $S = I$, which implies that some of the eigenvalues of A are negative.

Remark: If E is a Hermitian space, Theorem 12.1.3 also holds, but the orthogonal linear map g becomes a unitary map. In terms of matrices, the polar decomposition states that for every complex $n \times n$ matrix A, there is some unitary matrix U and some positive Hermitian matrix H such that

$$A = UH.$$

12.2 Singular Value Decomposition (SVD)

The proof of Theorem 12.1.3 shows that there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$, where $(u_1, \ldots, u_n)$ are eigenvectors of h_1 and $(v_1, \ldots, v_n)$ are eigenvectors of h_2. Furthermore, $(u_1, \ldots, u_r)$ is an orthonormal basis of $\operatorname{Im} f^*$, $(u_{r+1}, \ldots, u_n)$ is an orthonormal basis of $\operatorname{Ker} f$,

$(v_1, \ldots, v_r)$ is an orthonormal basis of $\operatorname{Im} f$, and $(v_{r+1}, \ldots, v_n)$ is an orthonormal basis of $\operatorname{Ker} f^*$. Using this, we immediately obtain the singular value decomposition theorem. Note that the singular value decomposition for linear maps of determinant $+1$ is called the *Cartan decomposition* (after Elie Cartan)!

Theorem 12.2.1 *Given a Euclidean space E of dimension n, for every linear map $f\colon E \to E$ there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ such that if r is the rank of f, the matrix of f w.r.t. these two bases is a diagonal matrix of the form*

$$\begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \end{pmatrix},$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the (positive) square roots of the nonnull eigenvalues of $f^* \circ f$ and $f \circ f^*$, and $\mu_{r+1} = \cdots = \mu_n = 0$. Furthermore, $(u_1, \ldots, u_n)$ are eigenvectors of $f^* \circ f$, $(v_1, \ldots, v_n)$ are eigenvectors of $f \circ f^*$, and $f(u_i) = \mu_i v_i$ when $1 \leq i \leq n$.

Proof. Going back to the proof of Theorem 12.2.1, there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$, where $(u_1, \ldots, u_n)$ are eigenvectors of h_1, $(v_1, \ldots, v_n)$ are eigenvectors of h_2, $f(u_i) = \mu_i v_i$ when $1 \leq i \leq r$, and $f(u_i) = 0$ when $r + 1 \leq i \leq n$. But now, with respect to the orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$, the matrix of f is indeed

$$\begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \end{pmatrix},$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f and $\mu_{r+1} = \cdots = \mu_n = 0$. $\square$

Note that $\mu_i > 0$ for all i ($1 \leq i \leq n$) iff f is invertible. Given an orientation of the Euclidean space E specified by some orthonormal basis $(e_1, \ldots, e_n)$ taken as direct, if $\det(f) \geq 0$, we can always make sure that the two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ are oriented positively. Indeed, if $\det(f) = 0$, we just have to flip u_n to $-u_n$ if necessary, and v_n to $-v_n$ if necessary. If $\det(f) > 0$, since $\mu_i > 0$ for all i, $1 \leq i \leq n$, the orthogonal matrices U and V whose columns are the u_i's and the v_i's have determinants of the same sign. Since $f(u_n) = \mu_n v_n$ and $\mu_n > 0$, we just have to flip u_n to $-u_n$ if necessary, since v_n will also be flipped. Theorem 12.2.1 can be restated in terms of (real) matrices as follows.

Theorem 12.2.2 *For every real $n \times n$ matrix A there are two orthogonal matrices U and V and a diagonal matrix D such that $A = VDU^\top$, where*

D is of the form

$$D = \begin{pmatrix} \mu_1 & & \dots & \\ & \mu_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \mu_n \end{pmatrix},$$

where $\mu_1, \dots, \mu_r$ are the singular values of f, i.e., the (positive) square roots of the nonnull eigenvalues of $A^\top A$ and $A\,A^\top$, and $\mu_{r+1} = \dots = \mu_n = 0$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\,A^\top$. Furthermore, if $\det(A) \geq 0$, it is possible to choose U and V such that $\det(U) = \det(V) = +1$, i.e., U and V are rotation matrices.

A triple (U, D, V) such that $A = V\,D\,U^\top$ is called a *singular value decomposition (SVD) of A*.

Remarks:

(1) In Strang [166] the matrices U, V, D are denoted by $U = Q_2$, $V = Q_1$, and $D = \Sigma$, and an SVD is written as $A = Q_1 \Sigma Q_2^\top$. This has the advantage that Q_1 comes before Q_2 in $A = Q_1 \Sigma Q_2^\top$. This has the disadvantage that A maps the columns of Q_2 (eigenvectors of $A^\top A$) to multiples of the columns of Q_1 (eigenvectors of $A\,A^\top$).

(2) Algorithms for actually computing the SVD of a matrix are presented in Golub and Van Loan [75] and Trefethen and Bau [170], where the SVD and its applications are also discussed quite extensively.

(3) The SVD also applies to complex matrices. In this case, for every complex $n \times n$ matrix A, there are two unitary matrices U and V and a diagonal matrix D such that

$$A = V\,D\,U^*,$$

where D is a diagonal matrix consisting of real entries $\mu_1, \dots, \mu_n$, where $\mu_1, \dots, \mu_r$ are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of A^*A and $A\,A^*$, and $\mu_{r+1} = \ldots = \mu_n = 0$.

It is easy to go from the polar form to the SVD, and conversely. Indeed, given a polar decomposition $A = R_1 S$, where R_1 is orthogonal and S is positive symmetric, there is an orthogonal matrix R_2 and a positive diagonal matrix D such that $S = R_2 D\,R_2^\top$, and thus

$$A = R_1 R_2 D\,R_2^\top = V\,D\,U^\top,$$

where $V = R_1 R_2$ and $U = R_2$ are orthogonal.

Going the other way, given an SVD decomposition $A = V\,D\,U^\top$, let $R = V\,U^\top$ and $S = U\,D\,U^\top$. It is clear that R is orthogonal and that S is positive symmetric, and

$$RS = V\,U^\top U D\,U^\top = V D\,U^\top = A.$$

Note that it is possible to require that $\det(R) = +1$ when $\det(A) \geq 0$.

Theorem 12.2.2 can be easily extended to rectangular $m\times n$ matrices (see Strang [166] or Golub and Van Loan [75], Trefethen and Bau [170]). As a matter of fact, both Theorem 12.1.3 and Theorem 12.2.1 can be generalized to linear maps $f\colon E \to F$ between two Euclidean spaces E and F. In order to do so, we need to define the analogue of the notion of an orthogonal linear map for linear maps $f\colon E \to F$. By definition, the adjoint $f^*\colon F \to E$ of a linear map $f\colon E \to F$ is the unique linear map such that

$$\langle f(u), v\rangle_2 = \langle u, f^*(v)\rangle_1$$

for all $u \in E$ and all $v \in F$. Then we have

$$\langle f(u), f(v)\rangle_2 = \langle u, (f^* \circ f)(v)\rangle_1$$

for all $u, v \in E$. Letting $n = \dim(E)$, $m = \dim(F)$, and $p = \min(m, n)$, if f has rank p and if for every p orthonormal vectors $(u_1, \ldots, u_p)$ in $(\mathrm{Ker}\, f)^\perp$ the vectors $(f(u_1), \ldots, f(u_p))$ are also orthonormal in F, then

$$f^* \circ f = \mathrm{id}$$

on $(\mathrm{Ker}\, f)^\perp$. The converse is immediately proved. Thus, we will say that a linear map $f\colon E \to F$ is *weakly orthogonal* if is has rank $p = \min(m, n)$ and if

$$f^* \circ f = \mathrm{id}$$

on $(\mathrm{Ker}\, f)^\perp$. Of course, $f^* \circ f = 0$ on $\mathrm{Ker}\, f$. In terms of matrices, we will say that a real $m\times n$ matrix A is weakly orthogonal if its first $p = \min(m, n)$ columns are orthonormal, the remaining ones (if any) being null columns. This is equivalent to saying that

$$A^\top A = I_n$$

if $m \geq n$, and that

$$A^\top A = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix}$$

if $n > m$. In this latter case $(n > m)$, it is immediately shown that

$$A\,A^\top = I_m,$$

and $A^\top$ is also weakly orthogonal. The main difference with orthogonal matrices is that $A\,A^\top$ is usually not a nice matrix of the above form when $m \geq n$ (unless $m = n$). Weakly unitary linear maps are defined analogously.

Theorem 12.2.3 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for every linear map $f\colon E \to F$ there* *are two positive self-adjoint linear maps* $h_1: E \to E$ *and* $h_2: F \to F$ *and a weakly orthogonal linear map* $g: E \to F$ *such that*

$$f = g \circ h_1 = h_2 \circ g.$$

Furthermore, if f *has rank* r*, the maps* h_1 *and* h_2 *have the same positive eigenvalues* $\mu_1, \ldots, \mu_r$*, which are the singular values of* f*, i.e., the positive square roots of the nonnull eigenvalues of both* $f^* \circ f$ *and* $f \circ f^*$*. Finally,* h_1, h_2 *are unique,* g *is unique if* $\mathrm{rank}(f) = \min(m, n)$ *and* $h_1 = h_2$ *if* f *is normal.*

Proof. By Lemma 12.1.2 there are two (unique) positive self-adjoint linear maps $h_1: E \to E$ and $h_2: F \to F$ such that $f^* \circ f = h_1^2$ and $f \circ f^* = h_2^2$. As in the proof of Theorem 12.1.3,

$$\mathrm{Ker}\, f = \mathrm{Ker}\, h_1,$$

and letting r be the rank of f, there is an orthonormal basis $(u_1, \ldots, u_n)$ of eigenvectors of h_1 such that $(u_1, \ldots, u_r)$ are associated with the strictly positive eigenvalues $\mu_1, \ldots, \mu_r$ of h_1 (the singular values of f). The vectors $(u_{r+1}, \ldots, u_n)$ form an orthonormal basis of $\mathrm{Ker}\, f = \mathrm{Ker}\, h_1$, and the vectors $(u_1, \ldots, u_r)$ form an orthonormal basis of $(\mathrm{Ker}\, f)^\perp = \mathrm{Im}\, f^*$. Furthermore, letting

$$v_i = \frac{f(u_i)}{\mu_i}$$

when $1 \le i \le r$, using the Gram–Schmidt orthonormalization procedure, we can extend $(v_1, \ldots, v_r)$ to an orthonormal basis $(v_1, \ldots, v_m)$ of F (even when $r = 0$). Also note that $(v_1, \ldots, v_r)$ is an orthonormal basis of $\mathrm{Im}\, f$, and $(v_{r+1}, \ldots, v_m)$ is an orthonormal basis of $\mathrm{Im}\, f^\perp = \mathrm{Ker}\, f^*$.

Letting $p = \min(m, n)$, we define the linear map $g: E \to F$ by its action on the basis $(u_1, \ldots, u_n)$ as follows:

$$g(u_i) = v_i$$

for all i, $1 \le i \le p$, and

$$g(u_i) = 0$$

for all i, $p + 1 \le i \le n$. Note that $r \le p$. Just as in the proof of Theorem 12.1.3, we have

$$(g \circ h_1)(u_i) = f(u_i)$$

when $1 \le i \le r$, and

$$(g \circ h_1)(u_i) = g(h_1(u_i)) = g(0) = 0$$

when $r + 1 \le i \le n$ (since $(u_{r+1}, \ldots, u_n)$ is a basis for $\mathrm{Ker}\, f = \mathrm{Ker}\, h_1$), which shows that $f = g \circ h_1$. The fact that g is weakly orthogonal follows easily from the fact that it maps the orthonormal vectors $(u_1, \ldots, u_p)$ to the orthonormal vectors $(v_1, \ldots, v_p)$.

We can show that $f = h_2 \circ g$ as follows. Just as in the proof of Theorem 12.1.3,

$$h_2^2(v_i) = \mu_i^2 v_i$$

when $1 \leq i \leq r$, and

$$h_2^2(v_i) = (f \circ f^*)(v_i) = f(f^*(v_i)) = 0$$

when $r+1 \leq i \leq m$, since $(v_{r+1}, \ldots, v_m)$ is a basis for $\mathrm{Ker}\, f^* = (\mathrm{Im}\, f)^{\perp}$. Since h_2 is positive self-adjoint, so is h_2^2, and by Lemma 12.1.2, we must have

$$h_2(v_i) = \mu_i v_i$$

when $1 \leq i \leq r$, and

$$h_2(v_i) = 0$$

when $r+1 \leq i \leq m$. This shows that $(v_1, \ldots, v_m)$ are eigenvectors of h_2 for $\mu_1, \ldots, \mu_m$ (letting $\mu_{r+1} = \cdots = \mu_m = 0$), and thus h_1 and h_2 have the same nonnull eigenvalues $\mu_1, \ldots, \mu_r$. As a consequence,

$$(h_2 \circ g)(u_i) = h_2(g(u_i)) = h_2(v_i) = \mu_i v_i = f(u_i)$$

when $1 \leq i \leq m$. Since h_1, h_2, $f^* \circ f$, and $f \circ f^*$ are positive self-adjoint, $f^* \circ f = h_1^2$, $f \circ f^* = h_2^2$, and $\mu_1, \ldots, \mu_r$ are the eigenvalues of both h_1 and h_2, it follows that $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of both $f^* \circ f$ and $f \circ f^*$.

Finally, if $m \geq n$ and $\mathrm{rank}(f) = n$, then $\mathrm{Ker}\, h_1 = \mathrm{Ker}\, f = (0)$ and h_1 is invertible and if $n \geq m$ and $\mathrm{rank}(f) = m$, then $\mathrm{Ker}\, h_2 = \mathrm{Ker}\, f^* = (0)$ and h_2 is invertible. By Lemma 12.1.2 h_1 and h_2 are unique and since

$$f = g \circ h_1 \quad \text{and} \quad f = h_2 \circ g,$$

if h_1 is invertible then $g = f \circ h_1^{-1}$ and if h_2 is invertible then $g = h_2^{-1} \circ f$, and thus g is also unique. If h is normal, then $f^* \circ f = f \circ f^*$ and $h_1 = h_2$. $\square$

In matrix form, Theorem 12.2.3 can be stated as follows. For every real $m \times n$ matrix A, there is some weakly orthogonal $m \times n$ matrix R and some positive symmetric $n \times n$ matrix S such that

$$A = RS.$$

The proof also shows that if $n > m$, the last $n - m$ columns of R are zero vectors. A pair (R, S) such that $A = RS$ is called a *polar decomposition of* A.

Remark: If E is a Hermitian space, Theorem 12.2.3 also holds, but the weakly orthogonal linear map g becomes a weakly unitary map. In terms of matrices, the polar decomposition states that for every complex $m \times n$

matrix A, there is some weakly unitary $m \times n$ matrix U and some positive Hermitian $n \times n$ matrix H such that

$$A = UH.$$

The proof of Theorem 12.2.3 shows that there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ for E and F, respectively, where $(u_1, \ldots, u_n)$ are eigenvectors of h_1 and $(v_1, \ldots, v_m)$ are eigenvectors of h_2. Furthermore, $(u_1, \ldots, u_r)$ is an orthonormal basis of $\operatorname{Im} f^*$, $(u_{r+1}, \ldots, u_n)$ is an orthonormal basis of $\operatorname{Ker} f$, $(v_1, \ldots, v_r)$ is an orthonormal basis of $\operatorname{Im} f$, and $(v_{r+1}, \ldots, v_m)$ is an orthonormal basis of $\operatorname{Ker} f^*$. Using this, we immediately obtain the singular value decomposition theorem for linear maps $f \colon E \to F$, where E and F can have different dimensions.

Theorem 12.2.4 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for every linear map $f \colon E \to F$ there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ such that if r is the rank of f, the matrix of f w.r.t. these two bases is a $m \times n$ matrix D of the form*

$$D = \begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix} \quad \text{or } D = \begin{pmatrix} \mu_1 & & \ldots & & 0 & \ldots & 0 \\ & \mu_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \mu_m & 0 & \ldots & 0 \end{pmatrix},$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the (positive) square roots of the nonnull eigenvalues of $f^* \circ f$ and $f \circ f^*$, and $\mu_{r+1} = \cdots = \mu_p = 0$, where $p = \min(m, n)$. Furthermore, $(u_1, \ldots, u_n)$ are eigenvectors of $f^* \circ f$, $(v_1, \ldots, v_m)$ are eigenvectors of $f \circ f^*$, and $f(u_i) = \mu_i v_i$ when $1 \leq i \leq p = \min(m, n)$.

Even though the matrix D is an $m \times n$ rectangular matrix, since its only nonzero entries are on the descending diagonal, we still say that D is a diagonal matrix. Theorem 12.2.4 can be restated in terms of (real) matrices as follows.

Theorem 12.2.5 *For every real $m \times n$ matrix A, there are two orthogonal matrices U $(n \times n)$ and V $(m \times m)$ and a diagonal $m \times n$ matrix D such* *that $A = V D U^\top$, where D is of the form*

$$D = \begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix} \quad or \quad D = \begin{pmatrix} \mu_1 & & \ldots & & 0 & \ldots & 0 \\ & \mu_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \mu_m & 0 & \ldots & 0 \end{pmatrix},$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the (positive) square roots of the nonnull eigenvalues of $A^\top A$ and $A\, A^\top$, and $\mu_{r+1} = \ldots = \mu_p = 0$, where $p = \min(m, n)$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\, A^\top$.

A triple (U, D, V) such that $A = V D U^\top$ is called a *singular value decomposition (SVD) of A*. The SVD of matrices can be used to define the pseudo-inverse of a rectangular matrix; see Strang [166], Trefethen and Bau [170], or Golub and Van Loan [75] for a thorough presentation.

Remark: The matrix form of Theorem 12.2.3 also yields a variant of the singular value decomposition. First, assume that $m \geq n$. Given an $m \times n$ matrix A, there is a weakly orthogonal $m \times n$ matrix R_1 and a positive symmetric $n \times n$ matrix S such that

$$A = R_1 S.$$

Since S is positive symmetric, there is an orthogonal $n \times n$ matrix R_2 and a diagonal $n \times n$ matrix D with nonnegative entries such that

$$S = R_2 D\, R_2^\top.$$

Thus, we can write

$$A = R_1 R_2 D\, R_2^\top.$$

We claim that $R_1 R_2$ is weakly orthogonal. Indeed,

$$(R_1 R_2)^\top (R_1 R_2) = R_2^\top\, (R_1^\top R_1) R_2,$$

and if $m \geq n$, we have

$$R_1^\top R_1 = I_n,$$

so that

$$(R_1 R_2)^\top (R_1 R_2) = I_n.$$

Thus, $R_1 R_2$ is indeed weakly orthogonal. Let us now consider the case $n > m$. From the version of SVD in which

$$A = V D U^\top$$

where U is $n \times n$ orthogonal, V is $m \times m$ orthogonal, and D is $m \times n$ diagonal with nonnegative diagonal entries, letting V' be the $m \times n$ matrix obtained from V by adding $n-m$ zero columns and D' be the $n \times n$ matrix obtained from D by adding $n-m$ zero rows, it is immediately verified that

$$V'D' = VD,$$

and thus when $n > m$, we also have

$$A = V'D'U^\top,$$

where U is $n \times n$ orthogonal, V' is $m \times n$ weakly orthogonal, and D' is $n \times n$ diagonal with nonnegative diagonal entries. As a consequence, in both cases we have shown that there exists a weakly orthogonal $m \times n$ matrix V, an orthogonal $n \times n$ matrix U, and a diagonal $n \times n$ matrix D with nonnegative entries such that

$$A = VDU^\top.$$

There is yet another alternative when $n > m$. Given an $m \times n$ matrix A, there is a positive symmetric $m \times m$ matrix S and a weakly orthogonal $m \times n$ matrix R_1, such that

$$A = SR_1.$$

Since S is positive symmetric, there is an orthogonal $m \times m$ matrix R_2 and a diagonal $m \times m$ matrix D with nonnegative entries such that

$$S = R_2 D R_2^\top.$$

Thus, we can write

$$A = R_2 D R_2^\top R_1.$$

We claim that $R_2^\top R_1$ is weakly orthogonal. Indeed,

$$(R_2^\top R_1)^\top R_2^\top R_1 = R_1^\top (R_2 R_2^\top) R_1 = R_1^\top R_1,$$

since R_2 is orthogonal, and if $n > m$, we have

$$R_1^\top R_1 = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix},$$

so that

$$(R_2^\top R_1)^\top R_2^\top R_1 = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix},$$

and $R_2^\top R_1$ is weakly orthogonal. Since $n > m$, $(R_2^\top R_1)^\top = R_1^\top R_2$ is also weakly orthogonal. As a consequence, we have shown that when $m \geq n$, there exists a weakly orthogonal $m \times n$ matrix V, an orthogonal $n \times n$ matrix U, and a diagonal $n \times n$ matrix D with nonnegative entries such that

$$A = VDU^\top,$$

and when $n > m$, there exists an orthogonal $m \times m$ matrix V, a weakly orthogonal $m \times n$ matrix $U^\top$ (with U also weakly orthogonal), and a diagonal $m \times m$ matrix D with nonnegative entries, such that

$$A = VDU^\top.$$

In both cases,

$$V^\top AU = D.$$

One of the spectral theorems states that a symmetric matrix can be diagonalized by an orthogonal matrix. There are several numerical methods to compute the eigenvalues of a symmetric matrix A. One method consists in *tridiagonalizing* A, which means that there exists some orthogonal matrix P and some symmetric tridiagonal matrix T such that $A = PTP^\top$. In fact, this can be done using Householder transformations. It is then possible to compute the eigenvalues of T using a bisection method based on Sturm sequences. One can also use Jacobi's method. For details, see Golub and Van Loan [75], Chapter 8, Trefethen and Bau [170], Lecture 26, or Ciarlet [33]. Computing the SVD of a matrix A is more involved. Most methods begin by finding orthogonal matrices U and V and a *bidiagonal* matrix B such that $A = VBU^\top$. This can also be done using Householder transformations. Observe that $B^\top B$ is symmetric tridiagonal. Thus, in principle, the previous method to diagonalize a symmetric tridiagonal matrix can be applied. However, it is unwise to compute $B^\top B$ explicitly, and more subtle methods are used for this last step. Again, see Golub and Van Loan [75], Chapter 8, and Trefethen and Bau [170], Lecture 31.

The polar form has applications in continuum mechanics. Indeed, in any deformation it is important to separate stretching from rotation. This is exactly what QS achieves. The orthogonal part Q corresponds to rotation (perhaps with an additional reflection), and the symmetric matrix S to stretching (or compression). The real eigenvalues $\sigma_1, \ldots, \sigma_r$ of S are the stretch factors (or compression factors) (see Marsden and Hughes [118]). The fact that S can be diagonalized by an orthogonal matrix corresponds to a natural choice of axes, the principal axes.

The SVD has applications to data compression, for instance in image processing. The idea is to retain only singular values whose magnitudes are significant enough. The SVD can also be used to determine the rank of a matrix when other methods such as Gaussian elimination produce very small pivots. One of the main applications of the SVD is the computation of the pseudo-inverse. Pseudo-inverses are the key to the solution of various optimization problems, in particular the method of least squares. This topic is discussed in the next chapter (Chapter 13). Applications of the material of this chapter can be found in Strang [166, 165]; Ciarlet [33]; Golub and Van Loan [75], which contains many other references; and Trefethen and Bau [170].

12.3 Problems

Problem 12.1 (1) Given a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

prove that there are Householder matrices G, H such that

$$GAH = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix} = D,$$

where D is a diagonal matrix, iff the following equations hold:

$$\begin{aligned} (b+c)\cos(\theta+\varphi) &= (a-d)\sin(\theta+\varphi), \\ (c-b)\cos(\theta-\varphi) &= (a+d)\sin(\theta-\varphi). \end{aligned}$$

(2) Discuss the solvability of the system. Consider the following cases:

1. $a - d = a + d = 0$.

2a. $a - d = b + c = 0$, $a + d \neq 0$.

2b. $a - d = 0$, $b + c \neq 0$, $a + d \neq 0$.

3a. $a + d = c - b = 0$, $a - d \neq 0$.

3b. $a + d = 0$, $c - b \neq 0$, $a - d \neq 0$.

4. $a + d \neq 0$, $a - d \neq 0$. Show that the solution in this case is

$$\begin{aligned} \theta &= \frac{1}{2}\left[\arctan\left(\frac{b+c}{a-d}\right) + \arctan\left(\frac{c-b}{a+d}\right)\right], \\ \varphi &= \frac{1}{2}\left[\arctan\left(\frac{b+c}{a-d}\right) - \arctan\left(\frac{c-b}{a+d}\right)\right]. \end{aligned}$$

If $b = 0$, show that the discussion is simpler: Basically, consider $c = 0$ or $c \neq 0$.

(3) Expressing everything in terms of $u = \cot\theta$ and $v = \cot\varphi$, show that the equations of question (1) become

$$\begin{aligned} (b+c)(uv-1) &= (u+v)(a-d), \\ (c-b)(uv+1) &= (-u+v)(a+d). \end{aligned}$$

Remark: I was unable to find an *elegant* solution for this system.

Problem 12.2 The purpose of this problem is to prove that given any linear map $f\colon E \to E$, where E is a Euclidean space of dimension $n \geq 2$ and an orthonormal basis $(e_1, \ldots, e_n)$, there are isometries g_i, h_i, hyperplane reflections or the identity, such that the matrix of

$$g_n \circ \cdots \circ g_1 \circ f \circ h_1 \circ \cdots \circ h_n$$

is a lower bidiagonal matrix, which means that the nonzero entries (if any) are on the main descending diagonal and on the diagonal below it.

(1) Prove that for any isometry $f\colon E \to E$ we have $f = f^* = f^{-1}$ iff $f \circ f = \mathrm{id}$.

(2) Proceed by induction, taking inspiration from the proof of the triangular decomposition given in Chapter 6. Let U_k' be the subspace spanned by $(e_1, \ldots, e_k)$ and U_k'' be the subspace spanned by $(e_{k+1}, \ldots, e_n)$, $1 \leq k \leq n-1$. For the base case, proceed as follows.

Let $v_1 = f^*(e_1)$ and $r_{1,1} = \|v_1\|$. Find an isometry h_1 (reflection or id) such that

$$h_1(f^*(e_1)) = r_{1,1}e_1.$$

Observe that $h_1(f^*(e_1)) \in U_1'$, so that

$$\langle h_1(f^*(e_1)), e_j \rangle = 0$$

for all j, $2 \leq j \leq n$, and conclude that

$$\langle e_1, f \circ h_1(e_j) \rangle = 0$$

for all j, $2 \leq j \leq n$.

Next, let

$$u_1 = f \circ h_1(e_1) = u_1' + u_1'',$$

where $u_1' \in U_1'$ and $u_1'' \in U_1''$, and let $r_{2,1} = \|u_1''\|$. Find an isometry g_1 (reflection or id) such that

$$g_1(u_1'') = r_{2,1}e_2.$$

Show that $g_1(e_1) = e_1$,

$$g_1 \circ f \circ h_1(e_1) = u_1' + r_{2,1}e_2,$$

and that

$$\langle e_1, g_1 \circ f \circ h_1(e_j) \rangle = 0$$

for all j, $2 \leq j \leq n$. At the end of this stage, show that $g_1 \circ f \circ h_1$ has a matrix such that all entries on its first row except perhaps the first are null, and that all entries on the first column, except perhaps the first two, are null.

Assume by induction that some isometries $g_1, \ldots, g_k$ and $h_1, \ldots, h_k$ have been found, either reflections or the identity, and such that

$$f_k = g_k \circ \cdots \circ g_1 \circ f \circ h_1 \circ \cdots \circ h_k$$

has a matrix that is lower bidiagonal up to and including row and column k, where $1 \leq k \leq n-2$.

Let

$$v_{k+1} = f_k^*(e_{k+1}) = v_{k+1}' + v_{k+1}'',$$

where $v'_{k+1} \in U'_k$ and $v''_{k+1} \in U''_k$, and let $r_{k+1,\,k+1} = \|v''_{k+1}\|$. Find an isometry h_{k+1} (reflection or id) such that

$$h_{k+1}(v''_{k+1}) = r_{k+1,\,k+1}e_{k+1}.$$

Show that if h_{k+1} is a reflection, then $U'_k \subseteq H_{k+1}$, where H_{k+1} is the hyperplane defining the reflection h_{k+1}. Deduce that $h_{k+1}(v'_{k+1}) = v'_{k+1}$, and that

$$h_{k+1}(f_k^*(e_{k+1})) = v'_{k+1} + r_{k+1,\,k+1}e_{k+1}.$$

Observe that $h_{k+1}(f_k^*(e_{k+1})) \in U'_{k+1}$, so that

$$\langle h_{k+1}(f_k^*(e_{k+1})), e_j\rangle = 0$$

for all j, $k+2 \leq j \leq n$, and thus

$$\langle e_{k+1}, f_k \circ h_{k+1}(e_j)\rangle = 0$$

for all j, $k+2 \leq j \leq n$.

Next, let

$$u_{k+1} = f_k \circ h_{k+1}(e_{k+1}) = u'_{k+1} + u''_{k+1},$$

where $u'_{k+1} \in U'_{k+1}$ and $u''_{k+1} \in U''_{k+1}$, and let $r_{k+2,\,k+1} = \|u''_{k+1}\|$. Find an isometry g_{k+1} (reflection or id) such that

$$g_{k+1}(u''_{k+1}) = r_{k+2,\,k+1}e_{k+2}.$$

Show that if g_{k+1} is a reflection, then $U'_{k+1} \subseteq G_{k+1}$, where G_{k+1} is the hyperplane defining the reflection g_{k+1}. Deduce that $g_{k+1}(e_i) = e_i$ for all i, $1 \leq i \leq k+1$, and that

$$g_{k+1} \circ f_k \circ h_{k+1}(e_{k+1}) = u'_{k+1} + r_{k+2,\,k+1}e_{k+2}.$$

Since by induction hypothesis

$$\langle e_i, f_k \circ h_{k+1}(e_j)\rangle = 0$$

for all i, j, $1 \leq i \leq k+1$, $k+2 \leq j \leq n$, and since $g_{k+1}(e_i) = e_i$ for all i, $1 \leq i \leq k+1$, conclude that

$$\langle e_i, g_{k+1} \circ f_k \circ h_{k+1}(e_j)\rangle = 0$$

for all i, j, $1 \leq i \leq k+1$, $k+2 \leq j \leq n$. Finish the proof.

Problem 12.3 Write a computer program implementing the method of Problem 12.2 to convert an $n \times n$ matrix to bidiagonal form.