7F Consequences of Singular Value Decomposition

Norms of Linear Maps

The singular value decomposition leads to the following upper bound for $\|Tv\|$.

7.82 *upper bound for $\|Tv\|$*

Suppose $T \in \mathcal{L}(V, W)$. Let s_1 be the largest singular value of T. Then

$$\|Tv\| \leq s_1 \|v\|$$

for all $v \in V$.

For a lower bound on $\|Tv\|$, look at Exercise 14 in Section 7E.

Proof Let $s_1, \ldots, s_m$ denote the positive singular values of T, and let $e_1, \ldots, e_m$ be an orthonormal list in V and $f_1, \ldots, f_m$ be an orthonormal list in W that provide a singular value decomposition of T. Thus

7.83
$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

for all $v \in V$. Hence if $v \in V$ then

$$\begin{aligned} \|Tv\|^2 &= s_1^2 \left|\langle v, e_1 \rangle\right|^2 + \cdots + s_m^2 \left|\langle v, e_m \rangle\right|^2 \\ &\leq s_1^2 \Big(\left|\langle v, e_1 \rangle\right|^2 + \cdots + \left|\langle v, e_m \rangle\right|^2 \Big) \\ &\leq s_1^2 \, \|v\|^2, \end{aligned}$$

where the last inequality follows from Bessel's inequality (6.26). Taking square roots of both sides of the inequality above shows that $\|Tv\| \leq s_1 \|v\|$, as desired. ■

Suppose $T \in \mathcal{L}(V, W)$ and s_1 is the largest singular value of T. The result above shows that

7.84
$$\|Tv\| \leq s_1 \text{ for all } v \in V \text{ with } \|v\| \leq 1.$$

Taking $v = e_1$ in 7.83 shows that $Te_1 = s_1 f_1$. Because $\|f_1\| = 1$, this implies that $\|Te_1\| = s_1$. Thus because $\|e_1\| = 1$, the inequality in 7.84 leads to the equation

7.85
$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} = s_1.$$

The equation above is the motivation for the following definition, which defines the norm of T to be the left side of the equation above without needing to refer to singular values or the singular value decomposition.

7.86 definition: *norm of a linear map, $\|\cdot\|$*

Suppose $T \in \mathcal{L}(V, W)$. Then the *norm* of T, denoted by $\|T\|$, is defined by

$$\|T\| = \max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\}.$$

In general, the maximum of an infinite set of nonnegative numbers need not exist. However, the discussion before 7.86 shows that the maximum in the definition of the norm of a linear map T from V to W does indeed exist (and equals the largest singular value of T).

We now have two different uses of the word *norm* and the notation $\|\cdot\|$. Our first use of this notation was in connection with an inner product on V, when we defined $\|v\| = \sqrt{\langle v, v\rangle}$ for each $v \in V$. Our second use of the norm notation and terminology is with the definition we just made of $\|T\|$ for $T \in \mathcal{L}(V, W)$. The norm $\|T\|$ for $T \in \mathcal{L}(V, W)$ does not usually come from taking an inner product of T with itself (see Exercise 21). You should be able to tell from the context and from the symbols used which meaning of the norm is intended.

The properties of the norm on $\mathcal{L}(V, W)$ listed below look identical to properties of the norm on an inner product space (see 6.9 and 6.17). The inequality in (d) is called the *triangle inequality*, thus using the same terminology that we used for the norm on V. For the reverse triangle inequality, see Exercise 1.

7.87 *basic properties of norms of linear maps*

Suppose $T \in \mathcal{L}(V, W)$. Then

(a) $\|T\| \geq 0$;

(b) $\|T\| = 0 \iff T = 0$;

(c) $\|\lambda T\| = |\lambda|\,\|T\|$ for all $\lambda \in \mathbf{F}$;

(d) $\|S + T\| \leq \|S\| + \|T\|$ for all $S \in \mathcal{L}(V, W)$.

Proof

(a) Because $\|Tv\| \geq 0$ for every $v \in V$, the definition of $\|T\|$ implies that $\|T\| \geq 0$.

(b) Suppose $\|T\| = 0$. Thus $Tv = 0$ for all $v \in V$ with $\|v\| \leq 1$. If $u \in V$ with $u \neq 0$, then
$$Tu = \|u\|\, T\Big(\frac{u}{\|u\|}\Big) = 0,$$
where the last equality holds because $u/\|u\|$ has norm 1. Because $Tu = 0$ for all $u \in V$, we have $T = 0$.

Conversely, if $T = 0$ then $Tv = 0$ for all $v \in V$ and hence $\|T\| = 0$.

(c) Suppose $\lambda \in \mathbf{F}$. Then
$$\begin{aligned}\|\lambda T\| &= \max\{\|\lambda Tv\| : v \in V \text{ and } \|v\| \leq 1\}\\ &= |\lambda| \max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\}\\ &= |\lambda|\,\|T\|.\end{aligned}$$

(d) Suppose $S \in \mathcal{L}(V, W)$. The definition of $\|S + T\|$ implies that there exists $v \in V$ such that $\|v\| \leq 1$ and $\|S + T\| = \big\|(S + T)v\big\|$. Now
$$\|S + T\| = \big\|(S + T)v\big\| = \|Sv + Tv\| \leq \|Sv\| + \|Tv\| \leq \|S\| + \|T\|,$$
completing the proof of (d). ■

For $S, T \in \mathcal{L}(V, W)$, the quantity $\|S - T\|$ is often called the distance between S and T. Informally, think of the condition that $\|S - T\|$ is a small number as meaning that S and T are close together. For example, Exercise 9 asserts that for every $T \in \mathcal{L}(V)$, there is an invertible operator as close to T as we wish.

7.88 *alternative formulas for $\|T\|$*

Suppose $T \in \mathcal{L}(V, W)$. Then

(a) $\|T\| =$ the largest singular value of T;

(b) $\|T\| = \max\{\|Tv\| : v \in V \text{ and } \|v\| = 1\}$;

(c) $\|T\| =$ the smallest number c such that $\|Tv\| \leq c\|v\|$ for all $v \in V$.

Proof

(a) See 7.85.

(b) Let $v \in V$ be such that $0 < \|v\| \leq 1$. Let $u = v/\|v\|$. Then

$$\|u\| = \left\|\frac{v}{\|v\|}\right\| = 1 \quad \text{and} \quad \|Tu\| = \left\|T\Big(\frac{v}{\|v\|}\Big)\right\| = \frac{\|Tv\|}{\|v\|} \geq \|Tv\|.$$

Thus when finding the maximum of $\|Tv\|$ with $\|v\| \leq 1$, we can restrict attention to vectors in V with norm 1, proving (b).

(c) Suppose $v \in V$ and $v \neq 0$. Then the definition of $\|T\|$ implies that

$$\left\|T\Big(\frac{v}{\|v\|}\Big)\right\| \leq \|T\|,$$

which implies that

$$\|Tv\| \leq \|T\|\,\|v\|. \tag{7.89}$$

Now suppose $c \geq 0$ and $\|Tv\| \leq c\|v\|$ for all $v \in V$. This implies that

$$\|Tv\| \leq c$$

for all $v \in V$ with $\|v\| \leq 1$. Taking the maximum of the left side of the inequality above over all $v \in V$ with $\|v\| \leq 1$ shows that $\|T\| \leq c$. Thus $\|T\|$ is the smallest number c such that $\|Tv\| \leq c\|v\|$ for all $v \in V$. ■

When working with norms of linear maps, you will probably frequently use the inequality 7.89.

For computing an approximation of the norm of a linear map T given the matrix of T with respect to some orthonormal bases, 7.88(a) is likely to be most useful. The matrix of T^*T is quickly computable from matrix multiplication. Then a computer can be asked to find an approximation for the largest eigenvalue of T^*T (excellent numeric algorithms exist for this purpose). Then taking the square root and using 7.88(a) gives an approximation for the norm of T (which usually cannot be computed exactly).

You should verify all assertions in the example below.

7.90 example: *norms*

- If I denotes the usual identity operator on V, then $\|I\| = 1$.
- If $T \in \mathcal{L}(\mathbf{F}^n)$ and the matrix of T with respect to the standard basis of $\mathbf{F}^n$ consists of all 1's, then $\|T\| = n$.
- If $T \in \mathcal{L}(V)$ and V has an orthonormal basis consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\|T\|$ is the maximum of the numbers $|\lambda_1|, \ldots, |\lambda_n|$.
- Suppose $T \in \mathcal{L}(\mathbf{R}^5)$ is the operator whose matrix (with respect to the standard basis) is the 5-by-5 matrix whose entry in row j, column k is $1/(j^2 + k)$. Standard mathematical software shows that the largest singular value of T is approximately 0.8 and the smallest singular value of T is approximately 10^{-6}. Thus $\|T\| \approx 0.8$ and (using Exercise 10 in Section 7E) $\|T^{-1}\| \approx 10^6$. It is not possible to find exact formulas for these norms.

A linear map and its adjoint have the same norm, as shown by the next result.

7.91 *norm of the adjoint*

Suppose $T \in \mathcal{L}(V, W)$. Then $\|T^*\| = \|T\|$.

Proof Suppose $w \in W$. Then

$$\|T^*w\|^2 = \langle T^*w, T^*w \rangle = \langle TT^*w, w \rangle \leq \|TT^*w\| \, \|w\| \leq \|T\| \, \|T^*w\| \, \|w\|.$$

The inequality above implies that

$$\|T^*w\| \leq \|T\| \, \|w\|,$$

which along with 7.88(c) implies that $\|T^*\| \leq \|T\|$.

Replacing T with T^* in the inequality $\|T^*\| \leq \|T\|$ and then using the equation $(T^*)^* = T$ shows that $\|T\| \leq \|T^*\|$. Thus $\|T^*\| = \|T\|$, as desired. ■

You may want to construct an alternative proof of the result above using Exercise 9 in Section 7E, which asserts that a linear map and its adjoint have the same positive singular values.

Approximation by Linear Maps with Lower-Dimensional Range

The next result is a spectacular application of the singular value decomposition. It says that to best approximate a linear map by a linear map whose range has dimension at most k, chop off the singular value decomposition after the first k terms. Specifically, the linear map T_k in the next result has the property that $\dim \operatorname{range} T_k = k$ and T_k minimizes the distance to T among all linear maps with range of dimension at most k. This result leads to algorithms for compressing huge matrices while preserving their most important information.

7.92 *best approximation by linear map whose range has dimension $\leq k$*

Suppose $T \in \mathcal{L}(V, W)$ and $s_1 \geq \cdots \geq s_m$ are the positive singular values of T. Suppose $1 \leq k < m$. Then

$$\min\{\|T - S\| : S \in \mathcal{L}(V, W) \text{ and } \dim \operatorname{range} S \leq k\} = s_{k+1}.$$

Furthermore, if

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

is a singular value decomposition of T and $T_k \in \mathcal{L}(V, W)$ is defined by

$$T_k v = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_k \langle v, e_k \rangle f_k$$

for each $v \in V$, then $\dim \operatorname{range} T_k = k$ and $\|T - T_k\| = s_{k+1}$.

Proof If $v \in V$ then

$$\begin{aligned}
\|(T - T_k)v\|^2 &= \|s_{k+1} \langle v, e_{k+1} \rangle f_{k+1} + \cdots + s_m \langle v, e_m \rangle f_m\|^2 \\
&= s_{k+1}{}^2 |\langle v, e_{k+1} \rangle|^2 + \cdots + s_m{}^2 |\langle v, e_m \rangle|^2 \\
&\leq s_{k+1}{}^2 \big(|\langle v, e_{k+1} \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2\big) \\
&\leq s_{k+1}{}^2 \|v\|^2.
\end{aligned}$$

Thus $\|T - T_k\| \leq s_{k+1}$. The equation $(T - T_k)e_{k+1} = s_{k+1} f_{k+1}$ now shows that $\|T - T_k\| = s_{k+1}$.

Suppose $S \in \mathcal{L}(V, W)$ and $\dim \operatorname{range} S \leq k$. Thus $Se_1, \ldots, Se_{k+1}$, which is a list of length $k+1$, is linearly dependent. Hence there exist $a_1, \ldots, a_{k+1} \in \mathbf{F}$, not all 0, such that

$$a_1 Se_1 + \cdots + a_{k+1} Se_{k+1} = 0.$$

Now $a_1 e_1 + \cdots + a_{k+1} e_{k+1} \neq 0$ because $a_1, \ldots, a_{k+1}$ are not all 0. We have

$$\begin{aligned}
\|(T - S)(a_1 e_1 + \cdots + a_{k+1} e_{k+1})\|^2 &= \|T(a_1 e_1 + \cdots + a_{k+1} e_{k+1})\|^2 \\
&= \|s_1 a_1 f_1 + \cdots + s_{k+1} a_{k+1} f_{k+1}\|^2 \\
&= s_1^2 |a_1|^2 + \cdots + s_{k+1}{}^2 |a_{k+1}|^2 \\
&\geq s_{k+1}{}^2 \big(|a_1|^2 + \cdots + |a_{k+1}|^2\big) \\
&= s_{k+1}{}^2 \|a_1 e_1 + \cdots + a_{k+1} e_{k+1}\|^2.
\end{aligned}$$

Because $a_1 e_1 + \cdots + a_{k+1} e_{k+1} \neq 0$, the inequality above implies that

$$\|T - S\| \geq s_{k+1}.$$

Thus $S = T_k$ minimizes $\|T - S\|$ among $S \in \mathcal{L}(V, W)$ with $\dim \operatorname{range} S \leq k$. ■

For other examples of the use of the singular value decomposition in best approximation, see Exercise 22, which finds a subspace of given dimension on which the restriction of a linear map is as small as possible, and Exercise 27, which finds a unitary operator that is as close as possible to a given operator.

Polar Decomposition

Recall our discussion before 7.54 of the analogy between complex numbers z with $|z| = 1$ and unitary operators. Continuing with this analogy, note that every complex number z except 0 can be written in the form

$$\begin{aligned} z &= \left(\frac{z}{|z|}\right)|z| \\ &= \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z}, \end{aligned}$$

where the first factor, namely, $z/|z|$, has absolute value 1.

Our analogy leads us to guess that every operator $T \in \mathcal{L}(V)$ can be written as a unitary operator times $\sqrt{T^*T}$. That guess is indeed correct. The corresponding result is called the polar decomposition, which gives a beautiful description of an arbitrary operator on V.

Note that if $T \in \mathcal{L}(V)$, then T^*T is a positive operator [as was shown in 7.64(a)]. Thus the operator $\sqrt{T^*T}$ makes sense and is well defined as a positive operator on V.

The polar decomposition that we are about to state and prove says that every operator on V is the product of a unitary operator and a positive operator. Thus we can write an arbitrary operator on V as the product of two nice operators, each of which comes from a class that we can completely describe and that we understand reasonably well. The unitary operators are described by 7.55 if $\mathbf{F} = \mathbf{C}$; the positive operators are described by the real and complex spectral theorems (7.29 and 7.31).

Specifically, consider the case $\mathbf{F} = \mathbf{C}$, and suppose

$$T = S\sqrt{T^*T}$$

is a polar decomposition of an operator $T \in \mathcal{L}(V)$, where S is a unitary operator. Then there is an orthonormal basis of V with respect to which S has a diagonal matrix, and there is an orthonormal basis of V with respect to which $\sqrt{T^*T}$ has a diagonal matrix. **Warning:** There may not exist an orthonormal basis that simultaneously puts the matrices of both S and $\sqrt{T^*T}$ into these nice diagonal forms—S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

However (still assuming that $\mathbf{F} = \mathbf{C}$), if T is normal, then an orthonormal basis of V can be chosen such that both S and $\sqrt{T^*T}$ have diagonal matrices with respect to this basis—see Exercise 31. The converse is also true: If $T \in \mathcal{L}(V)$ and $T = S\sqrt{T^*T}$ for some unitary operator $S \in \mathcal{L}(V)$ such that S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis of V, then T is normal. This holds because T then has a diagonal matrix with respect to this same orthonormal basis, which implies that T is normal [by the equivalence of (c) and (a) in 7.31].

The polar decomposition below is valid on both real and complex inner product spaces and for all operators on those spaces.

7.93 *polar decomposition*

Suppose $T \in \mathcal{L}(V)$. Then there exists a unitary operator $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Proof Let $s_1, \dots, s_m$ be the positive singular values of T, and let $e_1, \dots, e_m$ and $f_1, \dots, f_m$ be orthonormal lists in V such that

7.94 $$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$. Extend $e_1, \dots, e_m$ and $f_1, \dots, f_m$ to orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of V.

Define $S \in \mathcal{L}(V)$ by

$$Sv = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n$$

for each $v \in V$. Then

$$\begin{aligned}\|Sv\|^2 &= \big\|\langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n\big\|^2\\ &= \big|\langle v, e_1\rangle\big|^2 + \cdots + \big|\langle v, e_n\rangle\big|^2\\ &= \|v\|^2.\end{aligned}$$

Thus S is a unitary operator.

Applying T^* to both sides of 7.94 and then using the formula for T^* given by 7.77 shows that

$$T^*Tv = s_1{}^2\langle v, e_1\rangle e_1 + \cdots + s_m{}^2\langle v, e_m\rangle e_m$$

for every $v \in V$. Thus if $v \in V$, then

$$\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m$$

because the operator that sends v to the right side of the equation above is a positive operator whose square equals T^*T. Now

$$\begin{aligned}S\sqrt{T^*T}v &= S\big(s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m\big)\\ &= s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m\\ &= Tv,\end{aligned}$$

where the last equation follows from 7.94. ∎

Exercise 27 shows that the unitary operator S produced in the proof above is as close as a unitary operator can be to T.

Alternative proofs of the polar decomposition directly use the spectral theorem, avoiding the singular value decomposition. However, the proof above seems cleaner than those alternative proofs.

Operators Applied to Ellipsoids and Parallelepipeds

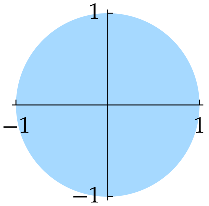
7.95 definition: *ball, B*

The *ball* in V of radius 1 centered at 0, denoted by B, is defined by

$$B = \{v \in V : \|v\| < 1\}.$$

If $\dim V = 2$, the word *disk* is sometimes used instead of *ball*. However, using *ball* in all dimensions is less confusing. Similarly, if $\dim V = 2$, then the word *ellipse* is sometimes used instead of the word *ellipsoid* that we are about to define. Again, using *ellipsoid* in all dimensions is less confusing.

You can think of the ellipsoid defined below as obtained by starting with the ball B and then stretching by a factor of s_k along each f_k-axis.



The ball B in $\mathbf{R}^2$.

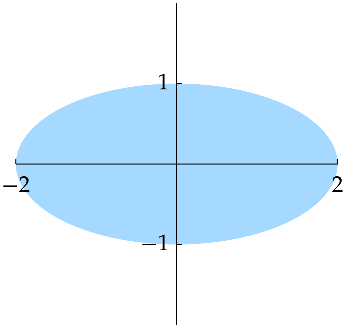
7.96 definition: *ellipsoid, $E(s_1 f_1, \dots, s_n f_n)$, principal axes*

Suppose that $f_1, \dots, f_n$ is an orthonormal basis of V and $s_1, \dots, s_n$ are positive numbers. The *ellipsoid* $E(s_1 f_1, \dots, s_n f_n)$ with *principal axes* $s_1 f_1, \dots, s_n f_n$ is defined by

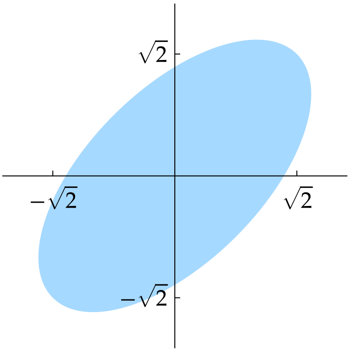
$$E(s_1 f_1, \dots, s_n f_n) = \left\{ v \in V : \frac{|\langle v, f_1 \rangle|^2}{s_1^2} + \cdots + \frac{|\langle v, f_n \rangle|^2}{s_n^2} < 1 \right\}.$$

The ellipsoid notation $E(s_1 f_1, \dots, s_n f_n)$ does not explicitly include the inner product space V, even though the definition above depends on V. However, the inner product space V should be clear from the context and also from the requirement that $f_1, \dots, f_n$ be an orthonormal basis of V.

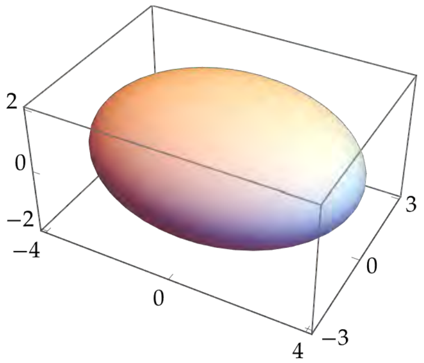
7.97 example: *ellipsoids*



The ellipsoid $E(2f_1, f_2)$ in $\mathbf{R}^2$, where f_1, f_2 is the standard basis of $\mathbf{R}^2$.



The ellipsoid $E(2f_1, f_2)$ in $\mathbf{R}^2$, where $f_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $f_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.



The ellipsoid $E(4f_1, 3f_2, 2f_3)$ in $\mathbf{R}^3$, where f_1, f_2, f_3 is the standard basis of $\mathbf{R}^3$.

The ellipsoid $E(f_1, \dots, f_n)$ equals the ball B in V for every orthonormal basis $f_1, \dots, f_n$ of V [by Parseval's identity 6.30(b)].

7.98 notation: $T(\Omega)$

For T a function defined on V and $\Omega \subseteq V$, define $T(\Omega)$ by

$$T(\Omega) = \{Tv : v \in \Omega\}.$$

Thus if T is a function defined on V, then $T(V) = \operatorname{range} T$.

The next result states that every invertible operator $T \in \mathcal{L}(V)$ maps the ball B in V onto an ellipsoid in V. The proof shows that the principal axes of this ellipsoid come from the singular value decomposition of T.

7.99 *invertible operator takes ball to ellipsoid*

Suppose $T \in \mathcal{L}(V)$ is invertible. Then T maps the ball B in V onto an ellipsoid in V.

Proof Suppose T has singular value decomposition

7.100
$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$; here $s_1, \dots, s_n$ are the singular values of T and $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are both orthonormal bases of V. We will show that $T(B) = E(s_1 f_1, \dots, s_n f_n)$.

First suppose $v \in B$. Because T is invertible, none of the singular values $s_1, \dots, s_n$ equals 0 (see 7.68). Thus 7.100 implies that

$$\frac{|\langle Tv, f_1\rangle|^2}{s_1^2} + \cdots + \frac{|\langle Tv, f_n\rangle|^2}{s_n^2} = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 < 1.$$

Thus $Tv \in E(s_1 f_1, \dots, s_n f_n)$. Hence $T(B) \subseteq E(s_1 f_1, \dots, s_n f_n)$.

To prove inclusion in the other direction, now suppose $w \in E(s_1 f_1, \dots, s_n f_n)$. Let

$$v = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_n\rangle}{s_n} e_n.$$

Then $\|v\| < 1$ and 7.100 implies that $Tv = \langle w, f_1\rangle f_1 + \cdots + \langle w, f_n\rangle f_n = w$. Thus $T(B) \supseteq E(s_1 f_1, \dots, s_n f_n)$. ■

We now use the previous result to show that invertible operators take all ellipsoids, not just the ball of radius 1, to ellipsoids.

7.101 *invertible operator takes ellipsoids to ellipsoids*

Suppose $T \in \mathcal{L}(V)$ is invertible and E is an ellipsoid in V. Then $T(E)$ is an ellipsoid in V.

Proof There exist an orthonormal basis $f_1, \dots, f_n$ of V and positive numbers $s_1, \dots, s_n$ such that $E = E(s_1 f_1, \dots, s_n f_n)$. Define $S \in \mathcal{L}(V)$ by

$$S(a_1 f_1 + \cdots + a_n f_n) = a_1 s_1 f_1 + \cdots + a_n s_n f_n.$$

Then S maps the ball B of V onto E, as you can verify. Thus

$$T(E) = T\big(S(B)\big) = (TS)(B).$$

The equation above and 7.99, applied to TS, show that $T(E)$ is an ellipsoid in V. ■

Recall (see 3.95) that if $u \in V$ and $\Omega \subseteq V$ then $u + \Omega$ is defined by

$$u + \Omega = \{u + w : w \in \Omega\}.$$

Geometrically, the sets Ω and $u + \Omega$ look the same, but they are in different locations.

In the following definition, if $\dim V = 2$ then the word *parallelogram* is often used instead of *parallelepiped*.

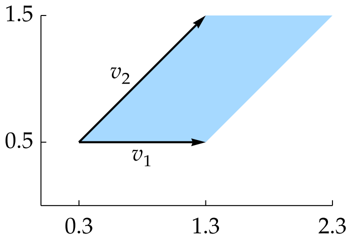
7.102 definition: *$P(v_1, \dots, v_n)$, parallelepiped*

Suppose $v_1, \dots, v_n$ is a basis of V. Let

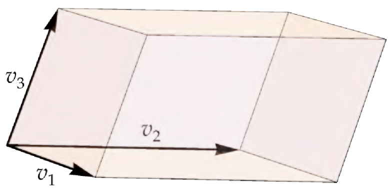
$$P(v_1, \dots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \dots, a_n \in (0, 1)\}.$$

A *parallelepiped* is a set of the form $u + P(v_1, \dots, v_n)$ for some $u \in V$. The vectors $v_1, \dots, v_n$ are called the *edges* of this parallelepiped.

7.103 example: *parallelepipeds*



The parallelepiped $(0.3, 0.5) + P\big((1,0), (1,1)\big)$ in $\mathbf{R}^2$.



A parallelepiped in $\mathbf{R}^3$.

7.104 *invertible operator takes parallelepipeds to parallelepipeds*

Suppose $u \in V$, $v_1, \ldots, v_n$ is a basis of V, and $T \in \mathcal{L}(V)$ is invertible. Then

$$T\big(u + P(v_1, \ldots, v_n)\big) = Tu + P(Tv_1, \ldots, Tv_n).$$

Proof Because T is invertible, the list $Tv_1, \ldots, Tv_n$ is a basis of V. The linearity of T implies that

$$T(u + a_1 v_1 + \cdots + a_n v_n) = Tu + a_1 Tv_1 + \cdots + a_n Tv_n$$

for all $a_1, \ldots, a_n \in (0, 1)$. Thus $T\big(u + P(v_1, \ldots, v_n)\big) = Tu + P(Tv_1, \ldots, Tv_n)$. ■

Just as the rectangles are distinguished among the parallelograms in $\mathbf{R}^2$, we give a special name to the parallelepipeds in V whose defining edges are orthogonal to each other.

7.105 definition: *box*

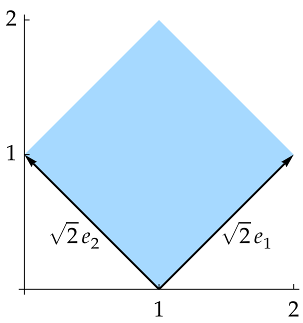
A *box* in V is a set of the form

$$u + P(r_1 e_1, \ldots, r_n e_n),$$

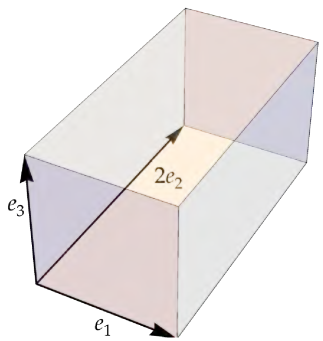
where $u \in V$ and $r_1, \ldots, r_n$ are positive numbers and $e_1, \ldots, e_n$ is an orthonormal basis of V.

Note that in the special case of $\mathbf{R}^2$ each box is a rectangle, but the terminology *box* can be used in all dimensions.

7.106 example: *boxes*



The box $(1, 0) + P(\sqrt{2}\,e_1, \sqrt{2}\,e_2)$, *where* $e_1 = \big(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\big)$ *and* $e_2 = \big(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\big)$.



The box $P(e_1, 2e_2, e_3)$, *where* e_1, e_2, e_3 *is the standard basis of* $\mathbf{R}^3$.

Suppose $T \in \mathcal{L}(V)$ is invertible. Then T maps every parallelepiped in V to a parallelepiped in V (by 7.104). In particular, T maps every box in V to a parallelepiped in V. This raises the question of whether T maps some boxes in V to boxes in V. The following result answers this question, with the help of the singular value decomposition.

7.107 *every invertible operator takes some boxes to boxes*

Suppose $T \in \mathcal{L}(V)$ is invertible. Suppose T has singular value decomposition

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n,$$

where $s_1, \dots, s_n$ are the singular values of T and $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthonormal bases of V and the equation above holds for all $v \in V$. Then T maps the box $u + P(r_1 e_1, \dots, r_n e_n)$ onto the box $Tu + P(r_1 s_1 f_1, \dots, r_n s_n f_n)$ for all positive numbers $r_1, \dots, r_n$ and all $u \in V$.

Proof If $a_1, \dots, a_n \in (0, 1)$ and $r_1, \dots, r_n$ are positive numbers and $u \in V$, then

$$T(u + a_1 r_1 e_1 + \cdots + a_n r_n e_n) = Tu + a_1 r_1 s_1 f_1 + \cdots + a_n r_n s_n f_n.$$

Thus $T\big(u + P(r_1 e_1, \dots, r_n e_n)\big) = Tu + P(r_1 s_1 f_1, \dots, r_n s_n f_n)$. ■

Volume via Singular Values

Our goal in this subsection is to understand how an operator changes the volume of subsets of its domain. Because notions of volume belong to analysis rather than to linear algebra, we will work only with an intuitive notion of volume. Our intuitive approach to volume can be converted into appropriate correct definitions, correct statements, and correct proofs using the machinery of analysis.

Our intuition about volume works best in real inner product spaces. Thus the assumption that $\mathbf{F} = \mathbf{R}$ will appear frequently in the rest of this subsection.

If $\dim V = n$, then by *volume* we will mean n-dimensional volume. You should be familiar with this concept in $\mathbf{R}^3$. When $n = 2$, this is usually called area instead of volume, but for consistency we use the word volume in all dimensions. The most fundamental intuition about volume is that the volume of a box (whose defining edges are by definition orthogonal to each other) is the product of the lengths of the defining edges. Thus we make the following definition.

7.108 definition: *volume of a box*

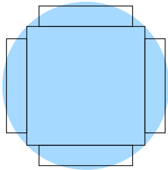
Suppose $\mathbf{F} = \mathbf{R}$. If $u \in V$ and $r_1, \dots, r_n$ are positive numbers and $e_1, \dots, e_n$ is an orthonormal basis of V, then

$$\text{volume}\big(u + P(r_1 e_1, \dots, r_n e_n)\big) = r_1 \times \cdots \times r_n.$$

The definition above agrees with the familiar formulas for the area (which we are calling the volume) of a rectangle in $\mathbf{R}^2$ and for the volume of a box in $\mathbf{R}^3$. For example, the first box in Example 7.106 has two-dimensional volume (or area) 2 because the defining edges of that box have length $\sqrt{2}$ and $\sqrt{2}$. The second box in Example 7.106 has three-dimensional volume 2 because the defining edges of that box have length 1, 2, and 1.

To define the volume of a subset of V, approximate the subset by a finite collection of disjoint boxes, and then add up the volumes of the approximating collection of boxes. As we approximate a subset of V more accurately by disjoint unions of more boxes, we get a better approximation to the volume.

These ideas should remind you of how the Riemann integral is defined by approximating the area under a curve by a disjoint collection of rectangles. This discussion leads to the following nonrigorous but intuitive definition.



Volume of this ball ≈ sum of the volumes of the five boxes.

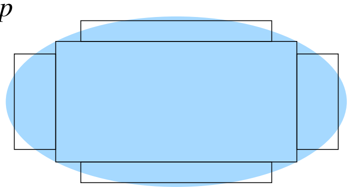
7.109 definition: *volume*

Suppose $\mathbf{F} = \mathbf{R}$ and $\Omega \subseteq V$. Then the *volume* of Ω, denoted by volume Ω, is approximately the sum of the volumes of a collection of disjoint boxes that approximate Ω.

We are ignoring many reasonable questions by taking an intuitive approach to volume. For example, if we approximate Ω by boxes with respect to one basis, do we get the same volume if we approximate Ω by boxes with respect to a different basis? If Ω_1 and Ω_2 are disjoint subsets of V, is $\operatorname{volume}(\Omega_1 \cup \Omega_2) = \operatorname{volume}\Omega_1 + \operatorname{volume}\Omega_2$? Provided that we consider only reasonably nice subsets of V, techniques of analysis show that both these questions have affirmative answers that agree with our intuition about volume.

7.110 example: *volume change by a linear map*

Suppose that $T \in \mathcal{L}(\mathbf{R}^2)$ is defined by $Tv = 2\langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2$, where e_1, e_2 is the standard basis of $\mathbf{R}^2$. This linear map stretches vectors along the e_1-axis by a factor of 2 and leaves vectors along the e_2-axis unchanged. The ball approximated by five boxes above gets mapped by T to the ellipsoid shown here. Each of the five boxes in the original figure gets mapped to a box of twice the width and the same height as in the original figure. Hence each box gets mapped to a box of twice the volume (area) as in the original figure. The sum of the volumes of the five new boxes approximates the volume of the ellipsoid. Thus T changes the volume of the ball by a factor of 2.



Each box here has twice the width and the same height as the boxes in the previous figure.

In the example above, T maps boxes with respect to the basis e_1, e_2 to boxes with respect to the same basis; thus we can see how T changes volume. In general, an operator maps boxes to parallelepipeds that are not boxes. However, if we choose the right basis (coming from the singular value decomposition!), then boxes with respect to that basis get mapped to boxes with respect to a possibly different basis, as shown in 7.107. This observation leads to a natural proof of the following result.

7.111 *volume changes by a factor of the product of the singular values*

Suppose $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$ is invertible, and $\Omega \subseteq V$. Then

$$\text{volume}\, T(\Omega) = (\text{product of singular values of } T)(\text{volume}\, \Omega).$$

Proof Suppose T has singular value decomposition

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$, where $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthonormal bases of V.

Approximate Ω by boxes of the form $u + P(r_1 e_1, \dots, r_n e_n)$, which have volume $r_1 \times \cdots \times r_n$. The operator T maps each box $u + P(r_1 e_1, \dots, r_n e_n)$ onto the box $Tu + P(r_1 s_1 f_1, \dots, r_n s_n f_n)$, which has volume $(s_1 \times \cdots \times s_n)(r_1 \times \cdots \times r_n)$.

The operator T maps a collection of boxes that approximate Ω onto a collection of boxes that approximate $T(\Omega)$. Because T changes the volume of each box in a collection that approximates Ω by a factor of $s_1 \times \cdots \times s_n$, the linear map T changes the volume of Ω by the same factor. ■

Suppose $T \in \mathcal{L}(V)$. As we will see when we get to determinants, the product of the singular values of T equals $|\det T|$; see 9.60 and 9.61.

Properties of an Operator as Determined by Its Eigenvalues

We conclude this chapter by presenting the table below. The context of this table is a finite-dimensional complex inner product space. The first column of the table shows a property that a normal operator on such a space might have. The second column of the table shows a subset of $\mathbf{C}$ such that the operator has the corresponding property if and only if all eigenvalues of the operator lie in the specified subset. For example, the first row of the table states that a normal operator is invertible if and only if all its eigenvalues are nonzero (this first row is the only one in the table that does not need the hypothesis that the operator is normal).

Make sure you can explain why all results in the table hold. For example, the last row of the table holds because the norm of an operator equals its largest singular value (by 7.85) and the singular values of a normal operator, assuming $\mathbf{F} = \mathbf{C}$, equal the absolute values of the eigenvalues (by Exercise 7 in Section 7E).

properties of a normal operator	**eigenvalues are contained in**
invertible	$\mathbf{C}\backslash\{0\}$
self-adjoint	$\mathbf{R}$
skew	$\{\lambda \in \mathbf{C} : \operatorname{Re}\lambda = 0\}$
orthogonal projection	$\{0, 1\}$
positive	$[0, \infty)$
unitary	$\{\lambda \in \mathbf{C} : \lvert\lambda\rvert = 1\}$
norm is less than 1	$\{\lambda \in \mathbf{C} : \lvert\lambda\rvert < 1\}$

Exercises 7F

1 Prove that if $S, T \in \mathcal{L}(V, W)$, then $\big| \|S\| - \|T\| \big| \le \|S - T\|$.

*The inequality above is called the **reverse triangle inequality**.*

2 Suppose that $T \in \mathcal{L}(V)$ is self-adjoint or that $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that

$$\|T\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

3 Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Prove that

$$\|Tv\| = \|T\|\,\|v\| \iff T^*Tv = \|T\|^2 v.$$

4 Suppose $T \in \mathcal{L}(V, W)$, $v \in V$, and $\|Tv\| = \|T\|\,\|v\|$. Prove that if $u \in V$ and $\langle u, v \rangle = 0$, then $\langle Tu, Tv \rangle = 0$.

5 Suppose U is a finite-dimensional inner product space, $T \in \mathcal{L}(V, U)$, and $S \in \mathcal{L}(U, W)$. Prove that

$$\|ST\| \le \|S\|\,\|T\|.$$

6 Prove or give a counterexample: If $S, T \in \mathcal{L}(V)$, then $\|ST\| = \|TS\|$.

7 Show that defining $d(S, T) = \|S - T\|$ for $S, T \in \mathcal{L}(V, W)$ makes d a metric on $\mathcal{L}(V, W)$.

This exercise is intended for readers who are familiar with metric spaces.

8 (a) Prove that if $T \in \mathcal{L}(V)$ and $\|I - T\| < 1$, then T is invertible.

(b) Suppose that $S \in \mathcal{L}(V)$ is invertible. Prove that if $T \in \mathcal{L}(V)$ and $\|S - T\| < 1/\|S^{-1}\|$, then T is invertible.

This exercise shows that the set of invertible operators in $\mathcal{L}(V)$ is an open subset of $\mathcal{L}(V)$, using the metric defined in Exercise 7.

9 Suppose $T \in \mathcal{L}(V)$. Prove that for every $\epsilon > 0$, there exists an invertible operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

10 Suppose $\dim V > 1$ and $T \in \mathcal{L}(V)$ is not invertible. Prove that for every $\epsilon > 0$, there exists $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$ and S is not invertible.

11 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that for every $\epsilon > 0$ there exists a diagonalizable operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

12 Suppose $T \in \mathcal{L}(V)$ is a positive operator. Show that $\big\|\sqrt{T}\big\| = \sqrt{\|T\|}$.

13 Suppose $S, T \in \mathcal{L}(V)$ are positive operators. Show that

$$\|S - T\| \le \max\{\|S\|, \|T\|\} \le \|S + T\|.$$

14 Suppose U and W are subspaces of V such that $\|P_U - P_W\| < 1$. Prove that $\dim U = \dim W$.

15 Define $T \in \mathcal{L}(\mathbf{F}^3)$ by

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

Find (explicitly) a unitary operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T = S\sqrt{T^*T}$.

16 Suppose $S \in \mathcal{L}(V)$ is a positive invertible operator. Prove that there exists $\delta > 0$ such that T is a positive operator for every self-adjoint operator $T \in \mathcal{L}(V)$ with $\|S - T\| < \delta$.

17 Prove that if $u \in V$ and φ_u is the linear functional on V defined by the equation $\varphi_u(v) = \langle v, u \rangle$, then $\|\varphi_u\| = \|u\|$.

Here we are thinking of the scalar field $\mathbf{F}$ *as an inner product space with* $\langle \alpha, \beta \rangle = \alpha\overline{\beta}$ *for all* $\alpha, \beta \in \mathbf{F}$. *Thus* $\|\varphi_u\|$ *means the norm of* φ_u *as a linear map from* V *to* $\mathbf{F}$.

18 Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $T \in \mathcal{L}(V, W)$.

(a) Prove that $\max\{\|Te_1\|, \ldots, \|Te_n\|\} \leq \|T\| \leq \left(\|Te_1\|^2 + \cdots + \|Te_n\|^2\right)^{1/2}$.

(b) Prove that $\|T\| = \left(\|Te_1\|^2 + \cdots + \|Te_n\|^2\right)^{1/2}$ if and only if $\dim \operatorname{range} T \leq 1$.

Here $e_1, \ldots, e_n$ *is an arbitrary orthonormal basis of* V, *not necessarily connected with a singular value decomposition of* T. *If* $s_1, \ldots, s_n$ *is the list of singular values of* T, *then the right side of the inequality above equals* $\left(s_1^2 + \cdots + s_n^2\right)^{1/2}$, *as was shown in Exercise 11(a) in Section 7E.*

19 Prove that if $T \in \mathcal{L}(V, W)$, then $\|T^*T\| = \|T\|^2$.

This formula for $\|T^*T\|$ *leads to the important subject of* C^**-algebras.*

20 Suppose $T \in \mathcal{L}(V)$ is normal. Prove that $\|T^k\| = \|T\|^k$ for every positive integer k.

21 Suppose $\dim V > 1$ and $\dim W > 1$. Prove that the norm on $\mathcal{L}(V, W)$ does not come from an inner product. In other words, prove that there does not exist an inner product on $\mathcal{L}(V, W)$ such that

$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} = \sqrt{\langle T, T \rangle}$$

for all $T \in \mathcal{L}(V, W)$.

22 Suppose $T \in \mathcal{L}(V, W)$. Let $n = \dim V$ and let $s_1 \geq \cdots \geq s_n$ denote the singular values of T. Prove that if $1 \leq k \leq n$, then

$$\min\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\} = s_{n-k+1}.$$

23 Suppose $T \in \mathcal{L}(V, W)$. Show that T is uniformly continuous with respect to the metrics on V and W that arise from the norms on those spaces (see Exercise 23 in Section 6B).

24 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that

$$\|T^{-1}\| = \|T\|^{-1} \iff \frac{T}{\|T\|} \text{ is a unitary operator.}$$

25 Fix $u, x \in V$ with $u \neq 0$. Define $T \in \mathcal{L}(V)$ by $Tv = \langle v, u \rangle x$ for every $v \in V$. Prove that

$$\sqrt{T^*T}v = \frac{\|x\|}{\|u\|}\langle v, u \rangle u$$

for every $v \in V$.

26 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if there exists a unique unitary operator $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$.

27 Suppose $T \in \mathcal{L}(V)$ and $s_1, \dots, s_n$ are the singular values of T. Let $e_1, \dots, e_n$ and $f_1, \dots, f_n$ be orthonormal bases of V such that

$$Tv = s_1\langle v, e_1 \rangle f_1 + \cdots + s_n\langle v, e_n \rangle f_n$$

for all $v \in V$. Define $S \in \mathcal{L}(V)$ by

$$Sv = \langle v, e_1 \rangle f_1 + \cdots + \langle v, e_n \rangle f_n.$$

(a) Show that S is unitary and $\|T - S\| = \max\{|s_1 - 1|, \dots, |s_n - 1|\}$.
(b) Show that if $E \in \mathcal{L}(V)$ is unitary, then $\|T - E\| \geq \|T - S\|$.

This exercise finds a unitary operator S that is as close as possible (among the unitary operators) to a given operator T.

28 Suppose $T \in \mathcal{L}(V)$. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $T = \sqrt{TT^*}\,S$.

29 Suppose $T \in \mathcal{L}(V)$.

(a) Use the polar decomposition to show that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $TT^* = ST^*TS^*$.
(b) Show how (a) implies that T and T^* have the same singular values.

30 Suppose $T \in \mathcal{L}(V)$, $S \in \mathcal{L}(V)$ is a unitary operator, and $R \in \mathcal{L}(V)$ is a positive operator such that $T = SR$. Prove that $R = \sqrt{T^*T}$.

This exercise shows that if we write T as the product of a unitary operator and a positive operator (as in the polar decomposition 7.93), then the positive operator equals $\sqrt{T^*T}$.

31 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$ and such that S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis of V.

32 Suppose that $T \in \mathcal{L}(V, W)$ and $T \neq 0$. Let $s_1, \dots, s_m$ denote the positive singular values of T. Show that there exists an orthonormal basis $e_1, \dots, e_m$ of $(\operatorname{null} T)^{\perp}$ such that

$$T\Bigl(E\Bigl(\frac{e_1}{s_1}, \dots, \frac{e_m}{s_m}\Bigr)\Bigr)$$

equals the ball in range T of radius 1 centered at 0.