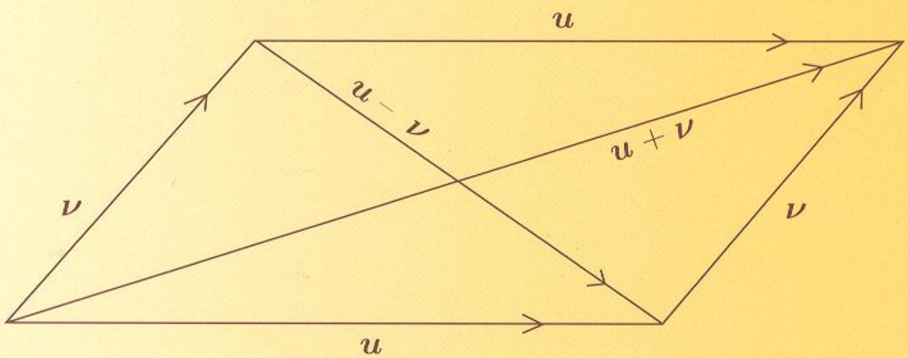
Sheldon Axler

LINEAR ALGEBRA DONE RIGHT

Second Edition



Springer

Sheldon Axler
Mathematics Department
San Francisco State University
San Francisco, CA 94132
USA

Mathematics Subject Classification (1991): 15-01

Library of Congress Cataloging-in-Publication Data
Axler, Sheldon Jay
Linear algebra done right / Sheldon Axler. - 2nd ed.
p. cm. — (Undergraduate texts in mathematics)
Includes index.
ISBN 0-387-98259-0 (alk. paper). - ISBN 0-387-98258-2 (pbk. : alk. paper)
1. Algebra, Linear. I. Title. II. Series.
QA184.A96 1997
512'.5—dc20 97-16664

Printed on acid-free paper.

Production managed by Victoria Evaretta; manufacturing supervised by Jeffrey Taub.
Photocomposed copy prepared from the author's TEX files.
Printed and bound by Hamilton Printing Co., Rensselaer, NY.
Printed in the United States of America.

9 8 7 6 5 4 3 (Corrected third printing, 1999)

ISBN 0-387-98259-0 Springer-Verlag New York Berlin Heidelberg SPIN 10629393 (hardcover)
ISBN 0-387-98258-2 Springer-Verlag New York Berlin Heidelberg SPIN 10719130 (softcover)

matrix 7.40 has the required form 7.39, completing the proof in this direction.

Conversely, now suppose that there is an orthonormal basis of V with respect to which the matrix of S has the form required by the theorem. Thus there is a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a subspace of V of dimension 1 or 2. Furthermore, any two vectors belonging to distinct U's are orthogonal, and each $S|_{U_j}$ is an isometry mapping U_j into U_j. If $v \in V$, we can write

$$v = u_1 + \cdots + u_m,$$

where each $u_j \in U_j$. Applying S to the equation above and then taking norms gives

$$\begin{aligned}\|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\ &= \|Su_1\|^2 + \cdots + \|Su_m\|^2 \\ &= \|u_1\|^2 + \cdots + \|u_m\|^2 \\ &= \|v\|^2.\end{aligned}$$

Thus S is an isometry, as desired. ■

Polar and Singular-Value Decompositions

Recall our analogy between $\mathbf{C}$ and $\mathcal{L}(V)$. Under this analogy, a complex number z corresponds to an operator T, and $\bar{z}$ corresponds to T^*. The real numbers correspond to the self-adjoint operators, and the nonnegative numbers correspond to the (badly named) positive operators. Another distinguished subset of $\mathbf{C}$ is the unit circle, which consists of the complex numbers z such that $|z| = 1$. The condition $|z| = 1$ is equivalent to the condition $\bar{z}z = 1$. Under our analogy, this would correspond to the condition $T^*T = I$, which is equivalent to T being an isometry (see 7.36). In other words, the unit circle in $\mathbf{C}$ corresponds to the isometries.

Continuing with our analogy, note that each complex number z except 0 can be written in the form

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

where the first factor, namely, $z/|z|$, is an element of the unit circle. Our analogy leads us to guess that any operator $T \in \mathcal{L}(V)$ can be written as an isometry times $\sqrt{T^*T}$. That guess is indeed correct, as we now prove.

If you know a bit of complex analysis, you will recognize the analogy to polar coordinates for complex numbers: every complex number can be written in the form $e^{\theta i}r$, where $\theta \in [0, 2\pi)$ and $r \geq 0$. Note that $e^{\theta i}$ is in the unit circle, corresponding to S being an isometry, and r is nonnegative, corresponding to $\sqrt{T^*T}$ being a positive operator.

7.41 Polar Decomposition: *If $T \in \mathcal{L}(V)$, then there exists an isometry $S \in \mathcal{L}(V)$ such that*

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

PROOF: Suppose $T \in \mathcal{L}(V)$. If $v \in V$, then

$$\begin{aligned}
\|Tv\|^2 &= \langle Tv, Tv\rangle \\
&= \langle T^*Tv, v\rangle \\
&= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\
&= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \\
&= \|\sqrt{T^*T}v\|^2.
\end{aligned}$$

Thus

7.42 $$\|Tv\| = \|\sqrt{T^*T}v\|$$

for all $v \in V$.

Define a linear map $S_1\colon \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ by

7.43 $$S_1(\sqrt{T^*T}v) = Tv.$$

The idea of the proof is to extend S_1 to an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Now for the details.

First we must check that S_1 is well defined. To do this, suppose $v_1, v_2 \in V$ are such that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$. For the definition given by 7.43 to make sense, we must show that $Tv_1 = Tv_2$. However,

$$\begin{aligned}
\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| \\
&= \|\sqrt{T^*T}(v_1 - v_2)\| \\
&= \|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\| \\
&= 0,
\end{aligned}$$

where the second equality holds by 7.42. The equation above shows that $Tv_1 = Tv_2$, so S_1 is indeed well defined. You should verify that S_1 is a linear map.

In the rest of the proof all we are doing is extending S_1 to an isometry S on all of V.

We see from 7.43 that S_1 maps range $\sqrt{T^*T}$ onto range T. Clearly 7.42 and 7.43 imply that $\|S_1 u\| = \|u\|$ for all $u \in$ range $\sqrt{T^*T}$. In particular, S_1 is injective. Thus from 3.4, applied to S_1, we have

$$\dim \operatorname{range} \sqrt{T^*T} = \dim \operatorname{range} T.$$

This implies that $\dim(\operatorname{range} \sqrt{T^*T})^\perp = \dim(\operatorname{range} T)^\perp$ (see Exercise 15 in Chapter 6). Thus orthonormal bases $(e_1, \dots, e_m)$ of $(\operatorname{range} \sqrt{T^*T})^\perp$ and $(f_1, \dots, f_m)$ of $(\operatorname{range} T)^\perp$ can be chosen; the key point here is that these two orthonormal bases have the same length. Define a linear map $S_2\colon (\operatorname{range} \sqrt{T^*T})^\perp \to (\operatorname{range} T)^\perp$ by

$$S_2(a_1 e_1 + \dots + a_m e_m) = a_1 f_1 + \dots + a_m f_m.$$

Obviously $\|S_2 w\| = \|w\|$ for all $w \in (\operatorname{range} \sqrt{T^*T})^\perp$.

Now let S be the operator on V that equals S_1 on range $\sqrt{T^*T}$ and equals S_2 on $(\operatorname{range} \sqrt{T^*T})^\perp$. More precisely, recall that each $v \in V$ can be written uniquely in the form

7.44 $$v = u + w,$$

where $u \in$ range $\sqrt{T^*T}$ and $w \in (\operatorname{range} \sqrt{T^*T})^\perp$ (see 6.29). For $v \in V$ with decomposition as above, define Sv by

$$Sv = S_1 u + S_2 w.$$

For each $v \in V$ we have

$$S(\sqrt{T^*T} v) = S_1(\sqrt{T^*T} v) = Tv,$$

so $T = S\sqrt{T^*T}$, as desired. All that remains is to show that S is an isometry. However, this follows easily from the two uses of the Pythagorean theorem: if $v \in V$ has decomposition as in 7.44, then

$$\begin{aligned} \|Sv\|^2 &= \|S_1 u + S_2 w\|^2 \\ &= \|S_1 u\|^2 + \|S_2 w\|^2 \\ &= \|u\|^2 + \|w\|^2 \\ &= \|v\|^2, \end{aligned}$$

where the second equality above holds because $S_1 u \in$ range T and $S_2 u \in (\operatorname{range} T)^\perp$. ■

The polar decomposition (7.41) states that each operator on V is the product of an isometry and a positive operator. Thus we can write each operator on V as the product of two operators, each of which comes from a class that we have completely described and that we understand reasonably well. The isometries are described by 7.37 and 7.38; the positive operators (which are all self-adjoint) are described by the spectral theorem (7.9 and 7.13).

Specifically, suppose $T = S\sqrt{T^*T}$ is the polar decomposition of $T \in \mathcal{L}(V)$, where S is an isometry. Then there is an orthonormal basis of V with respect to which S has a diagonal matrix (if $\mathbf{F} = \mathbf{C}$) or a block diagonal matrix with blocks of size at most 2-by-2 (if $\mathbf{F} = \mathbf{R}$), 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 forms (diagonal or block diagonal with small blocks). In other words, S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

Suppose $T \in \mathcal{L}(V)$. The ***singular values*** of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim \operatorname{null}(\sqrt{T^*T} - \lambda I)$ times. The singular values of T are all nonnegative because they are the eigenvalues of the positive operator $\sqrt{T^*T}$.

For example, if $T \in \mathcal{L}(\mathbf{F}^4)$ is defined by

7.45 $$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4),$$

then $T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4)$, as you should verify. Thus

$$\sqrt{T^*T}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4),$$

and we see that the eigenvalues of $\sqrt{T^*T}$ are $3, 2, 0$. Clearly

$$\dim \operatorname{null}(\sqrt{T^*T} - 3I) = 2,\ \dim \operatorname{null}(\sqrt{T^*T} - 2I) = 1,\ \dim \operatorname{null} \sqrt{T^*T} = 1.$$

Hence the singular values of T are $3, 3, 2, 0$. In this example -3 and 0 are the only eigenvalues of T, as you should verify.

Each $T \in \mathcal{L}(V)$ has $\dim V$ singular values, as can be seen by applying the spectral theorem and 5.21 (see especially part (e)) to the positive (hence self-adjoint) operator $\sqrt{T^*T}$. For example, the operator T defined by 7.45 on the four-dimensional vector space $\mathbf{F}^4$ has four singular values (they are $3, 3, 2, 0$), as we saw in the previous paragraph.

The next result shows that every operator on V has a nice description in terms of its singular values and two orthonormal bases of V.

7.46 Singular-Value Decomposition: *Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \dots, s_n$. Then there exist orthonormal bases $(e_1, \dots, e_n)$ and $(f_1, \dots, f_n)$ of V such that*

7.47
$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$.

PROOF: By the spectral theorem (also see 7.14) applied to $\sqrt{T^*T}$, there is an orthonormal basis $(e_1, \dots, e_n)$ of V such that $\sqrt{T^*T}e_j = s_j e_j$ for $j = 1, \dots, n$. We have

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

for every $v \in V$ (see 6.17). Apply $\sqrt{T^*T}$ to both sides of this equation, getting

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

This proof illustrates the usefulness of the polar decomposition.

for every $v \in V$. By the polar decomposition (see 7.41), there is an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Apply S to both sides of the equation above, getting

$$Tv = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n$$

for every $v \in V$. For each j, let $f_j = Se_j$. Because S is an isometry, $(f_1, \dots, f_n)$ is an orthonormal basis of V (see 7.36). The equation above now becomes

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

for every $v \in V$, completing the proof. ■

When we worked with linear maps from one vector space to a second vector space, we considered the matrix of a linear map with respect to a basis for the first vector space and a basis for the second vector space. When dealing with operators, which are linear maps from a vector space to itself, we almost always use only one basis, making it play both roles.

The singular-value decomposition allows us a rare opportunity to use two different bases for the matrix of an operator. To do this, suppose $T \in \mathcal{L}(V)$. Let $s_1, \dots, s_n$ denote the singular values of T, and let $(e_1, \dots, e_n)$ and $(f_1, \dots, f_n)$ be orthonormal bases of V such that the singular-value decomposition 7.47 holds. Then clearly

$$\mathcal{M}(T,(e_1,\dots,e_n),(f_1,\dots,f_n)) = \begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{bmatrix}.$$

In other words, every operator on V has a diagonal matrix with respect to some orthonormal bases of V, provided that we are permitted to use two different bases rather than a single basis as customary when working with operators.

Singular values and the singular-value decomposition have many applications (some are given in the exercises), including applications in computational linear algebra. To compute numeric approximations to the singular values of an operator T, first compute T^*T and then compute approximations to the eigenvalues of T^*T (good techniques exist for approximating eigenvalues of positive operators). The nonnegative square roots of these (approximate) eigenvalues of T^*T will be the (approximate) singular values of T (as can be seen from the proof of 7.28). In other words, the singular values of T can be approximated without computing the square root of T^*T.