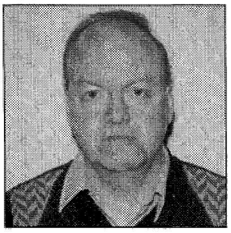
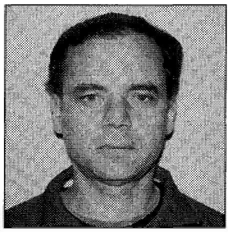
A Fresh Approach to the Singular Value Decomposition

Colm Mulcahy and John Rossi



Colm Mulcahy (colm@spelman.edu) earned his B.Sc. and M.Sc. at University College Dublin, Ireland. Alex F. T. W. Rosenberg directed his Ph.D work (Cornell, 1985), concerning higher level forms on ordered fields. Recent interests include CAGD, image processing, graphics, wavelets, geometry, and statistics. He and John Rossi met in 1985 in Hawaii (disputing over a bed; it's a long story), and since 1988 Colm has been at Spelman College.



John Rossi (rossi@math.vt.edu) got his Ph.D. (1980) from Jack Williamson at the University of Hawaii and joined Virginia Tech in 1983. His research area is complex analysis, especially value distribution theory, and he recently helped create an on-line first year linear algebra class. Over the objections of his still-furious spouse (see above bio), John undertook this project with Colm at an ATLAST developers workshop in Seattle in 1996.

The singular value decomposition (SVD) concerns the factorization of an arbitrary matrix A into a product UDV' of orthogonal matrices U and V and a "diagonal" matrix D. The SVD is applied frequently in numerical linear algebra; for example, it provides an excellent method for estimating the rank of a matrix, the standard method for computing the matrix norm $\| A \|_2$, and a robust method for solving ill-conditioned least squares problems [**2, 3, 14**]. The SVD and its close cousin the polar decomposition also come up in continuum mechanics, materials science [**4**], and light polarization [**11**].

Few instructors seem to find the enthusiasm or time to teach the SVD in the first (and too frequently *only*) linear algebra course. Even Dan Kalman, in his delightful article [**7**], concluded that "it is probably not feasible to include the SVD in the first linear algebra course." We aim to convince the reader that the singular value decomposition is in fact a very natural and approachable topic. It is equivalent, in the case of invertible square matrices, to the easy to motivate (and establish) polar decomposition. Using the SVD we can deduce a useful theorem that students—and indeed many instructors—will surely find surprising:

Theorem. *Every matrix is invertible, and hence every linear system can be solved.*

What we have of course stumbled on here is the concept of pseudo-inverses and the method of least squares, but without the baggage normally associated with those topics. Students with access to software such as *Matlab*, *Maple*, *Mathematica*, *Derive*, or *Macsyma* can then "solve" any inconsistent (as well as any under-determined) system in a meaningful way, thus supplementing their knowledge of the unique solution case. Since this returns to the topic with which most linear algebra courses start—namely, solving linear systems—it seems a fitting note on which to end an introductory course.

We also mention in passing some geometric interpretations of the SVD. While our treatment is far from complete, we still feel that this whirlwind tour of the singular value decomposition, omitting proofs and technicalities when it suits us, has pedagogical value. For simplicity we only work with very small matrices, and ignore the interesting Hermitian considerations that arise when complex-valued matrices are allowed. The point is that the concepts and methods introduced all apply to arbitrary matrices, where computer software such as *Matlab* can be used to gain access to useful solutions to real-world problems. For a more thorough, in-depth, and traditional introduction to the SVD, see [7].

Polar Decomposition

In what follows, we assume only this standard basic result on eigenvalues and eigenvectors: *If A is a real symmetric matrix, then $A = UDU'$, where D is diagonal and U is orthogonal (i.e., $UU' = I$); moreover, the diagonal entries of D are the eigenvalues of A, and the columns of U are corresponding eigenvectors.* (We adopt *Matlab* notation, denoting the transpose of a matrix U by U'.) Following Kalman's lead [7], we refer to this as an EVD (for Eigenvalue Decomposition) of A. It is trivial to check that nonsymmetric matrices have no such decomposition. Throughout, all matrices considered are real.

Real square matrices enjoy some properties with respect to the transpose operation that are reminiscent of complex numbers with respect to complex conjugation. First, recall that each $z \in \mathbb{C}$ can be written uniquely as $z = x + iy$, where x is real and iy is purely imaginary, and that if z is nonzero it can also be written uniquely as $z = rw$, where r is nonnegative real and w is "unitary." The first (rectangular representation) facilitates the addition of complex numbers, and the second (polar decomposition) facilitates multiplication. In terms of the conjugation map $z \to \overline{z}$, we have z is real when $\overline{z} = z$, purely imaginary when $\overline{z} = -z$, and "unitary" when $\overline{z}z = 1$—that is, when $z^{-1} = 1/z = \overline{z}i$ (so that z lies on the unit circle and has modulus 1).

Now consider real square matrices of some fixed size. It is well known that any such matrix A can be written (uniquely) as $A = S + K$, the sum of a symmetric matrix S and a skew-symmetric matrix K (by definition, B is *symmetric* when $B' = B$ and *skew-symmetric* when $B' = -B$). It is easy to prove that $S = (A + A')/2$ and $K = (A - A')/2$. Thus the representation $A = S + K$ parallels the rectangular representation $z = x + iy$ for complex numbers, where of course $x = (z + \overline{z})/2$, and $iy = (z - \overline{z})/2$.

This raises the question: Does the polar form of complex numbers have an analogue for invertible square matrices? It appears that we would like to prove:

Theorem 1. *Any invertible square matrix A can be written as $A = RW$, where R is "positive" symmetric and W is orthogonal.*

Proof. This is modeled on the proof in the complex number case. How *do* we show that each nonzero $z \in \mathbb{C}$ can be written as $z = rw$, where $r > 0$ and $w\overline{w} = 1$? One way is to pretend that we have such a representation, see what r and w have to be, and then go back and check that these values really do work. If $z = rw$, with r, w as above, then $\overline{z} = \overline{w}r$ (note the optional switch in order) and $z\overline{z} = rw\overline{w}r = r^2$. Thus r must be the positive square root of $z\overline{z}$, which is a positive real number since $\overline{(z\overline{z})} = z\overline{z} = |z|^2$. Conversely, given some nonzero $z \in \mathbb{C}$, if we define r to be the positive square root of $z\overline{z}$, and then take w to be $r^{-1}z$, it can easily be verified that $w\overline{w} = 1$.

Let's mimic this argument for an invertible square matrix A. If $A = RW$, with R symmetric and W orthogonal, then $A' = W'R$ (the switch in order here is forced upon us) and $AA' = RWW'R = R^2$. If we could claim that R must be "the positive square root of AA'," then matrix division on the left of $A = RW$ would yield $W = R^{-1}A$, which can easily be shown to satisfy $WW' = 1$. So it all boils down to our ability to take a "positive square root" of AA'. It turns out that there are several ways to do this. Certainly $(AA')' = AA'$, i.e., AA' is symmetric, and it is routine to check that the eigenvalues of AA' must be positive, bearing in mind that AA' is invertible since A is. So there exist matrices B and P such that $AA' = PBP'$, where B is diagonal—with positive diagonal entries, which are the eigenvalues of AA'—and P is orthogonal. The columns of P are eigenvectors of AA', and we can order them so that the entries of B are in non-decreasing order down the diagonal. Now if we let D be the diagonal matrix whose entries are the square roots of the corresponding entries of B, and define $R = PDP'$, it is clear that $R^2 = PDP'PDP' = PD^2P = PBP' = AA'$. This R will do nicely as a "positive square root of AA'." We note that P is not unique, and neither is R, but once an R is chosen it determines W. $\square$

Example 1. If $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, then $AA' = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. Proceeding as above, we have

$AA' = PDP'$, where $B = \begin{bmatrix} \frac{3+\sqrt{5}}{2} & 0 \\ 0 & \frac{3-\sqrt{5}}{2} \end{bmatrix}$, and $P = \begin{bmatrix} \sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} & \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} \\ \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} & -\sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} \end{bmatrix}$.

Thus $D = \begin{bmatrix} \sqrt{\frac{3+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\frac{3-\sqrt{5}}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$, and $R = PDP' = \begin{bmatrix} \frac{3}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$.

Hence, $W = R^{-1}A = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{3}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$, so a polar decomposition of A is given by:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \simeq \begin{bmatrix} 1.3416 & 0.4472 \\ 0.4472 & 0.8944 \end{bmatrix} \begin{bmatrix} 0.8944 & 0.4472 \\ -0.4472 & 0.8944 \end{bmatrix}.$$

In most of the examples here we follow the convention of giving precise (but radical-intensive) expressions for everything. Decimal approximations, such as those which *Matlab* yields, make the calculations look less fearsome but often obscure the patterns. It is easy to write a short program to automate finding such polar decompositions using one of the standard software packages; we offer a short *Matlab* routine later on.

Remark. We could just as easily have attempted to factor a given A as WR with W orthogonal and R symmetric. An interesting exercise is to repeat our previous analysis in this case and apply it to the matrices A in the last two examples. What changes? What stays the same?

For 2×2 or 3×3 matrices A, interpreted as geometric transformations acting on 2- or 3-dimensional space via left-multiplication, the polar decomposition $A = RW$ shows that the effect of A on an object can be interpreted as a rigid rotation (or a rotation and reflection) induced by W, followed by stretching and/or compression corresponding to R. In materials science one studies deformations of elastic bodies, and when such a transformation is decomposed into a rotation followed by a stretching and/or compression, much of the interest is in the energy required to elongate or compress the body [**4**]. The following example illustrates this geometrical view of the polar decomposition.

Example 2. (*The golden pumpkin.*) The shearing linear transformation induced by the matrix of example 1, $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, is a self-mapping of the plane. Applying this mapping to the 2-dimensional pumpkin face in Figure 1a, we obtain Figure 1c.

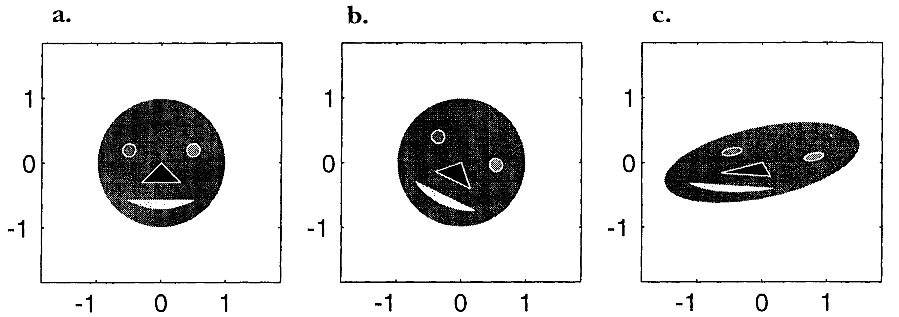


Figure 1. Squashing the pumpkin without losing face.

Now consider the polar decomposition $A = RW$ in example 1. Here W is a rotation matrix, corresponding to a rotation of θ radians counterclockwise about the origin, where $\cos(\theta) = 2/\sqrt{5}$ and $\sin(\theta) = -1/\sqrt{5}$. Hence $\theta = \arctan(-1/2) = -0.4637$ radians, or -26.4585 degrees. Thus W produces a clockwise rotation of about 26.5 degrees. The effect of this on the pumpkin face in Figure 1a is shown in Figure 1b. It remains to describe what R does to this rotated pumpkin face. As we saw,

$$R = \begin{bmatrix} \frac{3}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = PDP', \quad \text{where} \quad D = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \simeq \begin{bmatrix} 1.6180 & 0 \\ 0 & 0.6180 \end{bmatrix},$$

and

$$P = \begin{bmatrix} \sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} & \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} \\ \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} & -\sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} \end{bmatrix} \simeq \begin{bmatrix} 0.8507 & 0.5257 \\ 0.5257 & -0.8507 \end{bmatrix}.$$

Since the eigenvalues of R are $\left(1 + \sqrt{5}\right)/2$ (the golden ratio) and its reciprocal, R has the effect of stretching and compressing (by this factor) along the perpendicular axes determined by the columns of P, yielding the golden pumpkin in Figure 1c. Note that D, R, and A all have determinant 1, so that the area of each displayed pumpkin face is the same—at no stage did we lose face!

Singular Value Decomposition

In one sense, the singular value decomposition does for all matrices what orthogonal diagonalization (EVD) can only hope to do for symmetric square matrices. In the case of invertible square matrices, we can obtain SVD factorizations directly from polar decompositions. What makes an SVD special is that one can be found for any matrix, square or not, invertible or singular.

Theorem 2. *Given any $m \times n$ matrix A, there exists a "diagonal" $m \times n$ matrix D, together with orthogonal matrices U ($n \times n$) and V ($m \times m$) such that $A = UDV'$. More specifically, if r is the rank of A, then we can arrange it so that the only nonzero entries of D are the positive square roots $d_1, d_2, \ldots, d_r$, of the nonzero eigenvalues of AA'* (these are known as the "singular values of A"), *listed in non-increasing order down the principal diagonal of D. The first r columns of U (respectively V) are eigenvectors of AA' (respectively $A'A$), corresponding to $d_1, d_2, \ldots, d_r$.*

Proof. (For the invertible square case.) Assume $m = n$ and A is invertible, and fix a polar decomposition $A = RW$. As we saw before, $R = PDP'$, where P is orthogonal, and D has positive entries on its diagonal and zeros elsewhere. Thus $A = (PDP')W = (P)(D)(P'W)$. Setting $U = P$, and $V = P'W$ (which is easily shown to be orthogonal), we then have $A = UDV'$ as desired. $\square$

The first question we should ask is how does such an SVD compare with an EVD (orthogonal diagonalization) in the case of a symmetric matrix? The answer is simple: the SVD takes the *absolute values* of the eigenvalues, arranging them in non-increasing order.

Example 3. The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has eigenvalues 3 and -1, and singular values 3 and 1. This time we find that $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}'$ is an SVD. This should be compared to the EVD:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}'.$$

We now move on to nonsymmetric matrices—which cannot be orthogonally diagonalized.

Example 4. Using the matrices P, D, and W worked out in example 1 for $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, we see that the singular values of A are $\left(\sqrt{5}+1\right)/2$ and $\left(\sqrt{5}-1\right)/2$, and $P'W = \begin{bmatrix} \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} & -\sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} \\ \sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} & \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} \end{bmatrix}$. We thus get the following SVD of A:

$$\begin{bmatrix} \sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} & \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} \\ \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} & -\sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{\sqrt{5}-1}{2} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} & -\sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} \\ \sqrt{\frac{1}{2} + \frac{1}{10}\sqrt{5}} & \sqrt{\frac{1}{2} - \frac{1}{10}\sqrt{5}} \end{bmatrix}'$$

$$\simeq \begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix} \begin{bmatrix} 1.6180 & 0 \\ 0 & 0.6180 \end{bmatrix} \begin{bmatrix} 0.5257 & -0.8507 \\ 0.8507 & 0.5257 \end{bmatrix}'.$$

What do we do if A is not invertible, or not even square? In such cases we may not use the polar decomposition presented earlier, as our derivation required A to be invertible. According to the statement of Theorem 2, we must ensure that if $r = \text{rank}(A)$, then the first r columns of U (respectively V) are eigenvectors of AA' (respectively $A'A$).

Example 5. The rank one matrix $A = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}$ has eigenvalues 0 and -5. Here $AA' = A'A = \begin{bmatrix} 5 & -10 \\ -10 & 20 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 25 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}'$ is an EVD of AA'. Consequently, we have $A = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}'$ is an SVD. (*Note:* $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}'$ is an EVD of A.)

Example 6. If $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then $AA' = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}'$ and $A'A = [\,2\,]$. We get the SVD $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} [1]'$.

Pseudo-inverses

One popular application of orthogonal diagonalizations of symmetric matrices is to the computation of matrix powers. This is based on the observations that if $A = PDP'$, with P orthogonal and D diagonal, then $A^k = PD^kP'$, and if k is an integer, then D^k is trivial to find. In discussing polar decomposition, we effectively looked at the case $k = 1/2$. When $k = -1$, this approach yields a quick way to find the inverse of A (when all of the eigenvalues of A are nonzero). A similar application of the SVD can also be given: If $A = UDV'$ is an SVD of an invertible square matrix, then $A^{-1} = (UDV')^{-1} = VD^{-1}U'$, where D^{-1} can be computed by simply reciprocating each diagonal entry.

Example 7. Given the SVD $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}'$, we get $A^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}' = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$.

We can try this in general: first if D is any diagonal $m \times n$ matrix (meaning any nonzero entries occur on the principal diagonal) then we define the *pseudo-inverse* D^+ of D to be the matrix obtained by transposing D and reciprocating each nonzero (diagonal) entry. See [**1, 6, 8, 9, 12, 13**] for further discussion. This is a dream come true for linear algebra novices—invert what you can and ignore the rest!

(Don't forget to transpose too when D is not square.) For example,

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad [-\tfrac{1}{3} \;\; 0]^+ = \begin{bmatrix} -3 \\ 0 \end{bmatrix}.$$

Theorem 3. *Every matrix is "invertible."*

Proof. If $A = UDV'$ is an SVD, then A has *pseudo-inverse* $A^+ = VD^+U'$. □

Remark. All we are really doing here is providing a *definition* of the pseudo-inverse of an arbitrary matrix. It can be shown that this is well defined, in other words, while SVDs are not unique, pseudo-inverses are. One way to see this **[6]** entails proving that the pseudo-inverse (also known as the Moore-Penrose inverse) we have defined is the unique matrix A^+ such that

i. AA^+ and A^+A are symmetric.
ii. $AA^+A = A$.
iii. $A^+AA^+ = A^+$.

From this one can show various algebraic identities, for example $(A^+)^+ = A$. We offer another proof of uniqueness in the next section.

If A is square and invertible, so D has only nonzero entries on its diagonal, then as we saw in example 7 the pseudo-inverse $A^+ = VD^+U'$ is the ordinary inverse A^{-1}.

Example 8. From example 5,

$$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}^+ = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{25} & \frac{2}{25} \\ \frac{2}{25} & -\frac{4}{25} \end{bmatrix}.$$

Example 9. From example 6,

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}^+ = [1] \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}' \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = [\tfrac{1}{2} \;\; \tfrac{1}{2}].$$

Solving All Systems

Now that we have a way to invert any matrix, what could be more natural than to try to apply this to solving linear systems?

Theorem 4. *Every linear system is "solvable."*

Proof. If $Ax = b$ is a linear system, where A is $m \times n$, x is a column vector of n unknowns, and b is a known vector of length m, then $w = A^+b$ is the "solution." □

In the case where A is square and invertible, $w = A^+b = A^{-1}b$ is clearly the usual unique solution to the system; that is, $Ax = b$ if and only if $x = w$. The point is that we now have a way to "solve" all systems, even the traditionally discredited underdetermined and overdetermined (inconsistent) ones!

Example 10. The underdetermined system: $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ yields the solution $w = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}^+ \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{25} & \frac{2}{25} \\ \frac{2}{25} & -\frac{4}{25} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{25} \\ -\frac{4}{25} \end{bmatrix}$, using the pseudo-inverse found in example 8. One can easily check that out of the infinitude of solutions, this solution is the closest to the origin, i.e., the one of smallest norm.

Example 11. The simplest inconsistent system is surely $\begin{bmatrix} 1 \\ 1 \end{bmatrix} (x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. What sort of answer seems reasonable here? Using the pseudo-inverse found in example 9, we find $w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}^+ \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{2}$, which is a pretty good compromise given that we started with $x = 0$ and $x = 1$!

Remark. What we have here is the *method of least squares*. It turns out that $w = A^+b$ gives the least squares solution to $Ax = b$ with minimum norm. In other words the technique of the last two examples works for any underdetermined or any inconsistent linear system, provided we can compute the pseudo-inverse of the coefficient matrix [**5, 6, 13, 14**].

The proof that $w = A^+b$ gives the least squares solution to $Ax = b$ with minimum norm in no way depends on the particular matrices U and V in the SVD of A. So if A_1^+ and A_2^+ are pseudo-inverses obtained from distinct SVDs of A, they must give the same result when multiplied on the right by b, for all vectors b; hence A_1^+ and A_2^+ must be identical. This is another way to see that pseudo-inverses are unique.

Many software packages will provide an SVD and the pseudo-inverse of a matrix. For example, in *Matlab*, given A, to obtain matrices U, D, V with $A = UDV'$, just enter the command `[U,D,V]= SVD(A)`. To obtain the pseudo-inverse of A enter `pinv(A)`. The easiest way to find a polar decomposition of a matrix A using software is perhaps to work backwards from an SVD of A. In *Matlab* this simple m-file routine does the job:

```
function [R,W]=polard(A)
[U,D,V]=svd(A);
R=U*D*U'
W=inv(R)*A
```

This approach gives an answer for non-invertible matrices (and even rectangular ones), which raises some interesting questions worth pondering.

Closing Remarks

The SVD is one of several matrix factorization techniques (also including LU and QR decompositions) that seem to be making a long overdue appearance in some basic undergraduate linear algebra courses. These techniques are discussed in the emerging generation of basic undergraduate textbooks such as [**1, 5, 8, 13, 15**].

The SVD has a reputation for being relatively difficult, and indeed its usual presentation is rather different in flavor from the bulk of the material in a standard introductory course. Also, it seems to require more time to develop than we may have available. As a result, even if instructors include the SVD in their syllabi, fully intending to cover it in some fashion, when push comes to shove late in the term, it often gets dropped unceremoniously.

Perhaps there is another reason why the SVD gets left out in cold: *We teach what we know*. How many of us are as familiar with the SVD as we might be? The simple truth is that the SVD did not feature in the pure mathematics training that many of us received, not even in graduate school. Under these circumstances, is it any wonder that we fail to "pass it on" to so many generations of our own students?

While our treatment of the SVD is far from mathematically complete, we believe that it is nevertheless worthwhile because it gives students *some* appreciation of an important topic in a short amount of time. What we present here can be discussed in class in two hours. We believe that the advantages of knowing a little about the SVD (and how to implement it with software) outweigh the disadvantages of not having seen every claimed result proved in detail, and certainly our approach is preferable to learning absolutely nothing about it.

Figure 1 was generated using the ATLAST *Matlab* routine `transfor.m`. (ATLAST is an NSF-funded project to Augment the Teaching of Linear Algebra through the use of Software Tools [**10**].) This routine, and other m-files, some of which also explore the SVD, are available from < http://www2.gasou.edu/atlast >.

Acknowledgment. We thank ATLAST Project Director Steven Leon, and David Hill, who led the ATLAST Developers Workshop at the University of Washington, Seattle, in August 1996. This article grew out of our experiences at that workshop. Thanks also to Rick Elderkin for guidance in the early stages.

References

1. Otto Bretscher, *Linear Algebra with Applications*, Prentice Hall, Englewood Cliffs, NJ, 1997.
2. Charles G. Cullen, *An Introduction to Numerical Linear Algebra*, PWS, Boston, 1994.
3. G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.
4. Morton Gurtin, *An Introduction to Continuum Mechanics*, Academic Press, New York, 1981.
5. Richard Hill, *Elementary Linear Algebra with Applications*, 3rd ed., Saunders, Philadelphia, 1996.
6. Roger Horn and Charles Johnson, *Matrix Analysis*, Cambridge Press, New York, 1985.
7. Dan Kalman, A singularly valuable decomposition: The SVD of a matrix, *College Mathematics Journal* 27 (1996) 2–23.
8. David Lay, *Linear Algebra and Its Applications*, 2nd ed., Addison-Wesley, Reading, MA, 1997.
9. Stephen J. Leon, *Linear Algebra with Applications*, 5th ed., Prentice Hall, Englewood Cliffs, NJ, 1997.
10. Stephen J. Leon, Eugene Herman, and Richard Falkenberry, eds., *ATLAST Computer Exercises for Linear Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1996.
11. Michael Renardy, Singular value decomposition in Minkowski space, *Linear Algebra and Its Applications* 253 (1996) 53–58.
12. Gilbert Strang, *Linear Algebra and Its Applications*, 3rd ed., Saunders, Philadelphia, 1986.
13. ———, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, Wellesley, MA, 1993.
14. Lloyd N.Trefethen and David Bau III *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
15. David Winter, *Matrix Algebra*, Macmillan, 1992.