

Professor SVD

BY CLEVE MOLER

Stanford computer science professor Gene Golub has done more than anyone to make the singular value decomposition one of the most powerful and widely used tools in modern matrix computation.



Gene Golub's license plate, photographed by Professor P. M. Kroonenberg of Leiden University.

The SVD is a recent development. Pete Stewart, author of the 1993 paper “On the Early History of the Singular Value Decomposition”, tells me that the term *valeurs singulières* was first used by Emile Picard around 1910 in connection with integral equations. Picard used the adjective “singular” to mean something exceptional or out of the ordinary. At the time, it had

nothing to do with singular matrices.

When I was a graduate student in the early 1960s, the SVD was still regarded as a fairly obscure theoretical concept. A book that George Forsythe and I wrote in 1964 described the SVD as a nonconstructive way of characterizing the norm and condition number of a matrix. We did not yet have a practical way to actually compute

it. Gene Golub and W. Kahan published the first effective algorithm in 1965. A variant of that algorithm, published by Gene Golub and Christian Reinsch in 1970 is still the one we use today. By the time the first MATLAB appeared, around 1980, the SVD was one of its highlights.

We can generate a 2-by-2 example by working backwards, computing a matrix

from its SVD. Take $\sigma_1 = 2$, $\sigma_2 = 1/2$, $\theta = \pi/6$ and $\phi = \pi/4$. Let

$$U = \begin{pmatrix} -\cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

$$V = \begin{pmatrix} -\cos \phi & \sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

The matrices U and V are *rotations* through angles θ and ϕ , followed by *reflections* in the first dimension. The matrix Σ is a *diagonal* scaling transformation. Generate A by computing

$$A = U\Sigma V^T$$

You will find that

$$A = \begin{pmatrix} 1.4015 & -1.0480 \\ -0.4009 & 1.0133 \end{pmatrix}$$

This says that the matrix A can be generated by a rotation through 45° and a reflection, followed by independent scalings in each of the two coordinate directions by factors of 2 and $1/2$, respectively, followed by a rotation through 30° and another reflection.

The MATLAB function `eigshow` generates a figure that demonstrates the singular value decomposition of a 2-by-2 matrix. Enter the statements

```
A = [1.4015 -1.0480;
     -0.4009  1.0133];
eigshow(A)
```

The singular value decomposition (SVD), is a matrix factorization with a wide range of interesting applications.

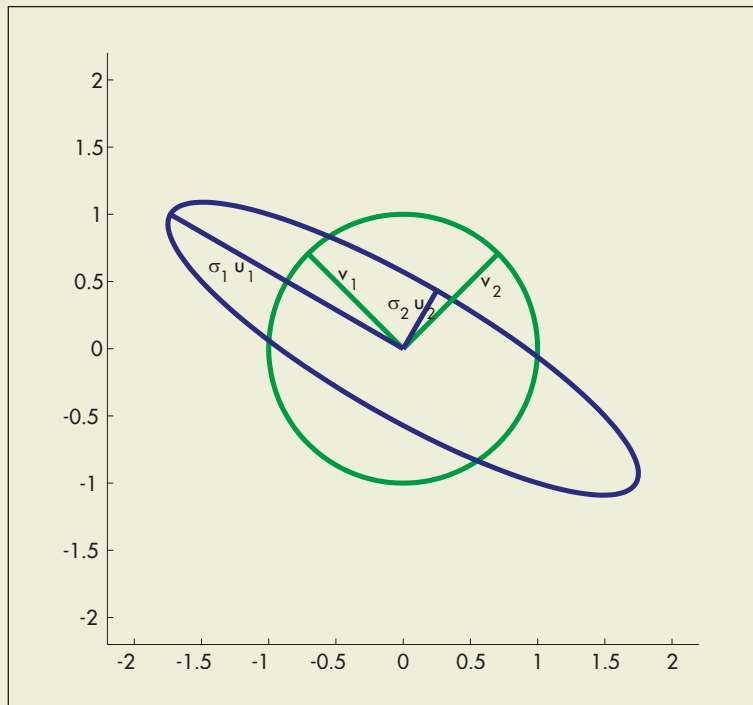


Figure 1. SVD figure produced by eigshow.

Click the SVD button and move the mouse around. You will see Figure 1, but with different labels.

The green circle is the unit circle in the plane. The blue ellipse is the image of this circle under transformation by the matrix A . The green vectors, v_1 and v_2 , which are the columns of V , and the blue vectors, u_1 and u_2 , which are the columns of U , are two different orthogonal bases for two-dimensional space. The columns of V are rotated 45° from the axes of the figure, while the columns of U , which are the major and minor axes of the ellipse, are rotated 30° . The matrix A transforms v_1 into $\sigma_1 u_1$ and v_2 into $\sigma_2 u_2$.

Let's move on to m -by- n matrices. One of the most important features of the SVD is its use of *orthogonal* matrices. A real matrix U is orthogonal, or has *orthonormal* columns, if

$$U^T U = I$$

This says that the columns of U are perpendicular to each other and have unit length.

Geometrically, transformations by orthogonal matrices are generalizations of rotations and reflections; they preserve lengths and angles. Computationally, orthogonal matrices are very desirable because they do not magnify roundoff or other kinds of errors.

Any real matrix A , even a nonsquare one, can be written as the product of three matrices.

$$A = U \Sigma V^T$$

The matrix U is orthogonal and has as many rows as A . The matrix V is orthogonal and has as many columns as A . The matrix Σ is the same size as A , but its only nonzero elements are on the main diagonal. The diagonal elements of Σ are the *singular values*, and the columns of U and V are the left and right *singular vectors*.

In abstract linear algebra terms, a matrix represents a linear transformation from one vector space, the *domain*, to another, the *range*. The SVD says that for any linear transformation it is possible to choose an orthonormal basis for the domain and

a possibly different orthonormal basis for the range. The transformation becomes independent of scalings or dilatations in each coordinate direction.

The *rank* of a matrix is the number of linearly independent rows, which is the same as the number of linearly independent columns. The rank of a diagonal matrix is clearly the number of nonzero diagonal elements. Orthogonal transforms preserve linear independence. Thus, the rank of any matrix is the number of nonzero singular values. In MATLAB, enter the statement

`type rank`

to see how we choose a tolerance and count nonnegligible singular values.

Traditional courses in linear algebra make considerable use of the reduced row echelon form (RREF), but the RREF is an unreliable tool for computation in the face of inexact data and arithmetic. The SVD can be regarded as a modern, computationally powerful replacement for the RREF.

A square diagonal matrix is *nonsingular* if, and only if, its diagonal elements are nonzero. The SVD implies that any square matrix is nonsingular if, and only if, its singular values are nonzero. The most numerically reliable way to determine whether matrices are singular is to test their singular values. This is far better than trying to compute determinants, which have atrocious scaling properties.

With the singular value decomposition, the system of linear equations

$$Ax = b$$

becomes

$$U \Sigma V^T x = b$$

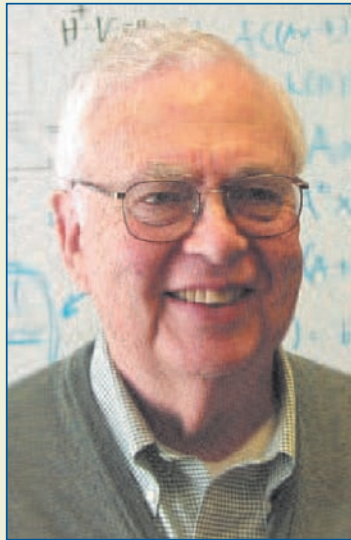
The solution is

$$x = V \Sigma^{-1} U^T b$$

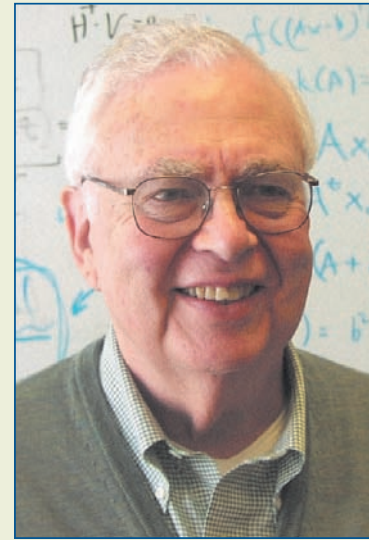
Multiply by an orthogonal matrix, divide by the singular values, then multiply by another orthogonal matrix. This is much more



12



50



120

Figure 2. Rank 12, 50, and 120 approximations to a rank 598 color photo of Gene Golub.

computational work than Gaussian elimination, but it has impeccable numerical properties. You can judge whether the singular values are small enough to be regarded as negligible, and if they are, analyze the relevant singular system.

Let E_k denote the outer product of the k -th left and right singular vectors, that is

$$E_k = u_k v_k^T$$

Then A can be expressed as a sum of rank-1 matrices,

$$A = \sum_{k=1}^n \sigma_k E_k$$

If you order the singular values in decreasing order, $\sigma_1 > \sigma_2 > \dots > \sigma_n$, and truncate the sum after r terms, the result is a rank- r approximation to the original matrix. The error in the approximation depends upon the magnitude of the neglected singular values. When you do this with a matrix of data that has been centered, by subtracting the mean of each column from the entire column, the process is known as *principal component analysis* (PCA). The right singular vectors, v_k , are

the *components*, and the scaled left singular vectors, $\sigma_k u_k$, are the *scores*. PCAs are usually described in terms of the eigenvalues and eigenvectors of the covariance matrix, AA^T , but the SVD approach sometimes has better numerical properties.

SVD and matrix approximation are often illustrated by approximating images. Our example starts with the photo on Gene Golub's Web page (Figure 2). The image is 897-by-598 pixels. We stack the red, green, and blue JPEG components vertically to produce a 2691-by-598 matrix. We then do just one SVD computation. After computing a low-rank approximation, we repartition the matrix into RGB components. With just rank 12, the colors are accurately reproduced and Gene is recognizable, especially if you squint at the picture to allow your eyes to reconstruct the original image. With rank 50, you can begin to read the mathematics on the white board behind Gene. With rank 120, the image is almost indistinguishable from the full rank 598. (This is not a particularly effective image compression technique. In fact, my friends in image processing call it "image degradation.")

So far in this column I have hardly mentioned eigenvalues. I wanted to show that it is possible to discuss singular values without discussing eigenvalues—but, of course, the two are closely related. In fact, if A is square, symmetric, and positive definite, its singular values and eigenvalues are equal, and its left and right singular vectors are equal to each other and to its eigenvectors. More generally, the singular values of A are the square roots of the eigenvalues of $A^T A$ or AA^T .

Singular values are relevant when the matrix is regarded as a transformation from one space to a different space with possibly different dimensions. Eigenvalues are relevant when the matrix is regarded as a transformation from one space into itself—as, for example, in linear ordinary differential equations.

Google finds over 3,000,000 Web pages that mention "singular value decomposition" and almost 200,000 pages that mention "SVD MATLAB." I knew about a few of these pages before I started to write this column. I came across some other interesting ones as I surfed around.

Professor SVD made all of this, and much more, possible. Thanks, Gene. ◀

A Few Search Results for “Singular Value Decomposition”

- The Wikipedia pages on SVD and PCA are quite good and contain a number of useful links, although not to each other.
en.wikipedia.org/wiki/Singular_value_decomposition
en.wikipedia.org/wiki/Principal_component_analysis
- Rasmus Bro, a professor at the Royal Veterinary and Agricultural University in Denmark, and Barry Wise, head of Eigenvector Research in Wenatchee, Washington, both do chemometrics using SVD and PCA. One example involves the analysis of the absorption spectrum of water samples from a lake to identify upstream sources of pollution.
www.models.kvl.dk/users/rasmus
www.eigenvector.com
- Tammy Kolda and Brett Bader, at Sandia National Labs in Livermore, ca, developed the Tensor Toolbox for MATLAB, which provides generalizations of PCA to multidimensional data sets.
csmr.ca.sandia.gov/~tgkolda/TensorToolbox
- In 2003, Lawrence Sirovich of the Mount Sinai School of Medicine published “A pattern analysis of the second Rehnquist U.S. Supreme Court” in the *Proceedings of the US National Academy of Sciences*. His paper led to articles in the *New York Times* and the *Washington Post* because it provides a nonpolitical, phenomenological model of court decisions. Between 1994 and 2002, the court heard 468 cases. Since there are nine justices, each of whom takes a majority or minority position on each case, the data is a 468-by-9 matrix of +1s and -1s. If the judges had made their decisions by flipping coins, this matrix would almost certainly have rank 9. But Sirovich found that the third singular value is an order of magnitude smaller than the first one, so the matrix is well approximated by a matrix of rank 2. In other words, most of the court’s decisions are close to being in a two-dimensional subspace of all possible decisions.
www.pnas.org/cgi/reprint/100/13/7432
- Latent Semantic Indexing involves the use of SVD with term-document matrices to perform document retrieval. For example, should a search for “singular value” also look for “eigenvalue”? See a 1999 *SIAM Review* paper by Michael Berry, Zlato Drmac, and Liz Jessup, “Matrices, Vector Spaces, and Information Retrieval.”
epubs.siam.org/SIREV/volume41/art_34703.html
- The first Google hit on “protein svd” is “Protein Substate Modeling and Identification Using the SVD,” by Tod Romo at Rice University. The site provides an electronic exposition of the use of SVD in the analysis of the structure and motion of proteins, and includes some gorgeous graphics.
bioc.rice.edu/~tromo/Sprez/toc.html
- Los Alamos biophysicists Michael Wall, Andreas Rechsteiner, and Luis Rocha provide a good online reference about SVD and PCA, phrased in terms of applications to gene expression analysis.
public.lanl.gov/mewall/kluwer2002.html
- “Representing cyclic human motion using functional analysis” (2005), by Dirk Ormoneit, Michael Black, Trevor Hastie, and Hedvig Kjellstrom, describes techniques involving Fourier analysis and principal component analysis for analyzing and modeling motion-capture data from activities such as walking.
www.csc.kth.se/~hedvig/publications/ivc_05.pdf
- A related paper is “Decomposing biological motion: a framework for analysis and synthesis of human gait patterns,” (2002), by Nicholaus Troje. Troje’s work is the basis for an “eigenwalker” demo.
www.journalofvision.org/2/5/2
www.mathworks.com/moler/ncm/walker.m
- A search at the US Patent and Trademark Office Web page lists 1,197 U.S. patents that mention “singular value decomposition.” The oldest, issued in 1987, is for “A fiber optic inspection system for use in the inspection of sandwiched solder bonds in integrated circuit packages”. Other titles include “Compression of surface light fields”, “Method of seismic surveying”, “Semantic querying of a peer-to-peer network”, “Biochemical markers of brain function”, and “Diabetes management.”
www.uspto.gov/patft

RESOURCES

- ▶ **On the Early History of the Singular Value Decomposition**
locus.siam.org/SIREV/volume35/art_1035134.html
- ▶ **Cleve’s Corner Collection**
www.mathworks.com/res/cleve