The Singular Value Decomposition

Higham, Nicholas J.

2015

MIMS EPrint: **2015.104**

ISSN 1749-9097

The Singular Value Decomposition†

Nicholas J. Higham

One of the most useful matrix factorizations is the *singular value decomposition* (SVD), which is defined for an arbitrary rectangular matrix $A \in \mathbb{C}^{m\times n}$. It takes the form

$$A = U\Sigma V^*, \quad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}, \quad (1)$$

where $p = \min(m, n)$, Σ is a diagonal matrix with diagonal elements $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$, and $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary. The σ_i are the *singular values* of A, and they are the nonnegative square roots of the p largest eigenvalues of A^*A. The columns of U and V are the left and right *singular vectors* of A, respectively.

Postmultiplying (1) by V gives $AV = U\Sigma$ since $V^*V = I$, which shows that the ith columns of U and V are related by $Av_i = \sigma_i u_i$ for $i = 1\colon p$. Similarly, $A^*u_i = \sigma_i v_i$ for $i = 1\colon p$. A geometrical interpretation of the former equation is that the singular values of A are the lengths of the semiaxes of the hyperellipsoid $\{ Ax : \|x\|_2 = 1 \}$.

Assuming that $m \geqslant n$ for notational simplicity, from (1) we have

$$A^*A = V(\Sigma^*\Sigma)V^*, \quad (2)$$

with $\Sigma^*\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$, which shows that the columns of V are eigenvectors of the matrix A^*A with corresponding eigenvalues the squares of the singular values of A. Likewise, the columns of U are eigenvectors of the matrix AA^*.

The SVD reveals a great deal about the matrix A and the key subspaces associated with it. The rank, r, of A is equal to the number of nonzero singular values and the range and the null space of A are spanned by the first r columns of U and the last $n - r$ columns of V, respectively.

The SVD reveals not only the rank but also how close A is to a matrix of a given rank, as shown by a classic 1936 theorem of Eckart and Young.

Theorem 1 (Eckart–Young). *Let $A \in \mathbb{C}^{m\times n}$ have the SVD (1). If $k < r = \mathrm{rank}(A)$, then for the 2-norm and the Frobenius norm,*

$$\min_{\mathrm{rank}(B)=k} \|A-B\| = \|A-A_k\| = \begin{cases} \sigma_{k+1}, & \text{2-norm,} \\ \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}, & F\text{-norm,} \end{cases}$$

where

$$A_k = UD_kV^*, \quad D_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0).$$

In many situations the matrices that arise are necessarily of low rank but errors in the underlying data make the matrices actually obtained of full rank. The Eckart–Young result tells us that in order to obtain a lower rank matrix we are justified in discarding (i.e., setting to zero) singular values that are of the same order of magnitude as the errors in the data.

The SVD (1) can be written as an outer product expansion

$$A = \sum_{i=1}^{p} \sigma_i u_i v_i^*,$$

and A_k in the Eckart–Young theorem is given by the same expression with p replaced by k. If $k \ll p$ then A_k requires much less storage than A and so the SVD can provide *data compression* (or *data reduction*). As an example, consider the monochrome image at the top

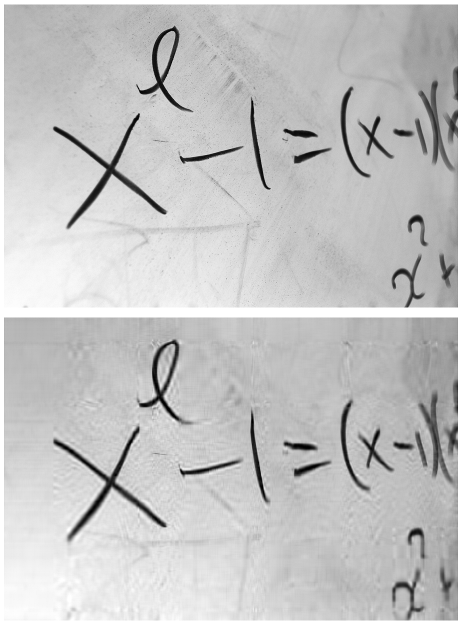

Figure 1 Photo of blackboard, inverted so that white and black are interchanged in order to show more clearly the texture of the board. Top: original 1067×1600 image. Bottom: image compressed using rank 40 approximation A_{40} computed from SVD.

†. Author's final version, before copy editing and cross-referencing, of: N. J. Higham. The singular value decomposition. In N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa, and J. Tanner, editors, *The Princeton Companion to Applied Mathematics*, pages 126–127. Princeton University Press, Princeton, NJ, USA, 2015.

of figure 1 represented by a 1067×1600 array of RGB values ($R = G = B$ since the image is monochrome). Let $A \in \mathbb{R}^{1067\times1600}$ contain the values from any one of the three channels. The singular values of A range from 8.4×10^4 down to 1.3×10^1. If we retain only the singular values down to the 40th, $\sigma_{40} = 2.1 \times 10^3$ (a somewhat arbitrary cutoff since there is no pronounced gap in the singular values), we obtain the image at the bottom of figure 1. The reduced SVD requires only 6 percent of the storage of the original matrix. Some degradation is visible in the compressed image (and more can be seen when it is viewed at 100 percent size on-screen), but it retains all the key features of the original image. While this example illustrates the power of the SVD, image compression is in general done much more effectively by the JPEG scheme.

A pleasing feature of the SVD is that the singular values are not unduly affected by perturbations. Indeed, if A is perturbed to $A + E$ then no singular value of A changes by more than $\|E\|_2$.

The SVD is a valuable tool in applications where two-sided orthogonal transformations can be carried out without "changing the problem", as it allows the matrix of interest to be diagonalized. Foremost among such problems is the linear least squares problem $\min_{x\in\mathbb{C}^n} \|b - Ax\|_2$.

The SVD was first derived by Beltrami in 1873. The first reliable method for computing it was published by Golub and Kahan in 1965; this method applies two-sided unitary transformations to A and does not form and solve the equation (2), or its analog for AA^*. Once software for computing the SVD became readily available, in the 1970s, the use of the SVD proliferated. Among the wide variety of uses of the SVD are for text mining, deciphering encrypted messages, and image deblurring.

Further Reading

[1] Lars Eldén. *Matrix Methods in Data Mining and Pattern Recognition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007. x+224 pp. ISBN 978-0-898716-26-9.

[2] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Fourth edition, Johns Hopkins University Press, Baltimore, MD, USA, 2013. xxi+756 pp. ISBN 978-1-4214-0794-4.

[3] Cleve B. Moler and Donald Morrison. Singular value analysis of cryptograms. *Amer. Math. Monthly*, 90: 78–87, 1983.