Introduction to the Singular Value Decomposition

The Singular Value Decomposition (SVD) is a topic rarely reached in undergraduate linear algebra courses and often skipped over in graduate courses. Consequently relatively few mathematicians are familiar with what M.I.T. Professor Gilbert Strang calls "absolutely a high point of linear algebra."

The singular value decomposition is a powerful technique in many matrix computations and analyses. Using the SVD of a matrix in computations, rather than the original matrix, has the advantage of being more robust to numerical error. Additionally, the SVD exposes the geometric structure of a matrix, an important aspect of many matrix calculations. A matrix can be described as a transformation from one vector space to another. The components of the SVD quantify the resulting change between the underlying geometry of those vector spaces.

The SVD is employed in a variety of applications, from least-squares problems to solving systems of linear equations. Each of these applications exploit key properties of the SVD -- its relation to the rank of a matrix and its ability to approximate matrices of a given rank. Many fundamental aspects of linear algebra rely on determining the rank of a matrix, making the SVD an important and widely-used technique.

Some Basics

The null space of a matrix A is the set of x for which Ax = 0, and the range of A is the set of b for which Ax = b has a solution for x.

We say two vectors x and y are orthogonal if $x^T y = 0$: In two or three dimensional space, this simply means that the vectors are perpendicular. Let A be a square matrix such that its columns are mutually orthogonal vectors of length 1, i.e. $x^T x = 1$. Then A is an orthogonal matrix and $A^T A = I$, the identity matrix. To simplify the notation, assume that a matrix A has at least as many rows as columns (m >= n).

A singular value decomposition of an mxn matrix A is any factorization of the form

$$A = U D V^T$$

where U is an mxm orthogonal matrix, V is an nxn orthogonal matrix, andD is an mxn diagonal matrix with $\sigma_1 \geq \sigma_1 \geq \ldots\ \sigma_n \geq 0$. The quantities σ_i are called the singular values of A, and the columns of U and V are called the left and right singular vectors, respectively.

SVD and Matrix Rank

Fundamental to linear algebra is the notion of rank. Numerous theorems begin with the condition ``If matrix A is of full rank, then the following property holds". However, if the matrix is rank deficient (or nearly so), then small perturbations of the matrix values (from round-off errors or fuzzy data) will yield a matrix which is of full rank. Hence, determining the rank of a matrix is non-trivial. The SVD lends us a practical definition of rank, as well as allows us to quantify the notion of near rank deficiency.

The familiar definition of rank is the number of linearly independent columns of a matrix. Let the matrix A have the $U D V^T$. Since multiplication by orthogonal matrices preserves linear independence, the rank of A is precisely the rank of the diagonal matrix D, or equivalently, the number of non-zero singular values. If A is nearly rank deficient (singular), then the singular values will be small. Moreover, suppose that Rank(A) = m and we wish to approximate A by a matrix B of lower rank k. Then we can use the following methodo compute a matr ix with the best approximation: B= $U D_k V^T$ where D_k is obtained by setting the singular values of D smaller than σ_k to 0.

SVD and Linear Independence

Another use of the SVD provides a measure, called a condition number, which is related to the measure of linear independence between the column vectors of the matrix.

The condition number (with respect to the Euclidean norm) of a matrix A is

$$\text{cond}(A) = \sigma_{max} / \sigma_{min}$$

where σ_{max} and σ_{min} are the largest and smallest singular values of A. If A is rank deficient, then $\sigma_i = 0$ and we consider cond(A) = ∞. Using the condition number, we can quantify the independence of the columns of A. Note that cond(A) >= 1. If cond(A) is close to 1, then the columns of A are very independent. When the condition number is large, the columns of A are nearly dependent.

As we will see in the next section, the notion of a condition number becomes important in solving linear systems, where cond(A) in some sense measures the sensitivity of the system to noise in the data.

Solutions to Linear Equations

Numerous practical problems can be expressed in the language of linear algebra. A linear system involves a set of equations in N variables. For example, consider the following linear system.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 8 \\ 10x_1 + 18x_2 + 12x_3 &= 78 \\ 20x_1 + 22x_2 + 40x_3 &= 144 \end{aligned}$$

This problem can be expressed in terms of a coefficient matrix A, a vector x of variables, and a vector b, such that a solution to the linear system Ax = b is an assignment to the values of the vector x.

As an extension of solving linear systems where we have more equations than unknown (m>n), we wish to find a solution where Ax is approximately equal to b. By this we mean the least-squares solution x to minimize

$$||Ax-b||^2$$

An advantage of using the SVD for this problem is that it can reliably handle the rank deficient case as well as the full rank case. Here when we say the LS system is full rank, it means A has full column rank, i.e., rank n or in other words, there is no dependency between the unknown variables. When we say the LS system is deficient rank, it means A has rank less than n.

The least square solution is $\boldsymbol{x} = A^+ \boldsymbol{b}$, where $A^+ = V D^+ U^{\mathrm{T}}$, and D^+ is a diagonal matrix, with diagonal entries given by $1/\sigma_i$ for nonzero singular values σ_i and zero otherwise.

In the case of deficient-rank system, $\boldsymbol{x} = A^+ \boldsymbol{b}$ is the solution that minimizes $\| \boldsymbol{x} \|$.

In the case where $\boldsymbol{b}$=0, $\boldsymbol{x} = A^+ \boldsymbol{b} = 0$ gives you the trivial solution of course. The nontrivial solution $\mathbf{x}$ is given by the eigenvector associated with the smallest eigenvalue of A^TA, i.e., the last column of V.

Reference

Gene H. Golub and Charles F. Van Loan. Matrix Computations, pages 16--21, 293. Johns Hopkins University Press, Baltimore, Maryland, 1983.