Lecture 5. More on the SVD

We continue our discussion of the singular value decomposition, emphasizing its connection with low-rank approximation of matrices in the 2-norm and the Frobenius norm.

A Change of Bases

The SVD makes it possible for us to say that every matrix is diagonal—if only one uses the proper bases for the domain and range spaces.

Here is how the change of bases works. Any $b \in \mathbb{C}^m$ can be expanded in the basis of left singular vectors of A (columns of U), and any $x \in \mathbb{C}^n$ can be expanded in the basis of right singular vectors of A (columns of V). The coordinate vectors for these expansions are

$$b' = U^*b, \qquad x' = V^*x.$$

By (4.3), the relation $b = Ax$ can be expressed in terms of b' and x':

$$b = Ax \iff U^*b = U^*Ax = U^*U\Sigma V^*x \iff b' = \Sigma x'.$$

Whenever $b = Ax$, we have $b' = \Sigma x'$. Thus A reduces to the diagonal matrix Σ when the range is expressed in the basis of columns of U and the domain is expressed in the basis of columns of V.

SVD vs. Eigenvalue Decomposition

The theme of diagonalizing a matrix by expressing it in terms of a new basis also underlies the study of eigenvalues. A nondefective square matrix A can be expressed as a diagonal matrix of eigenvalues Λ, if the range and domain are represented in a basis of eigenvectors.

If the columns of a matrix $X \in \mathbb{C}^{m\times m}$ contain linearly independent eigenvectors of $A \in \mathbb{C}^{m\times m}$, the *eigenvalue decomposition* of A is

$$A = X\Lambda X^{-1}, \tag{5.1}$$

where Λ is an $m \times m$ diagonal matrix whose entries are the eigenvalues of A. This implies that if we define, for $b, x \in \mathbb{C}^m$ satisfying $b = Ax$,

$$b' = X^{-1}b, \qquad x' = X^{-1}x,$$

then the newly expanded vectors b' and x' satisfy $b' = \Lambda x'$. Eigenvalues are treated systematically in Lecture 24.

There are fundamental differences between the SVD and the eigenvalue decomposition. One is that the SVD uses two different bases (the sets of left and right singular vectors), whereas the eigenvalue decomposition uses just one (the eigenvectors). Another is that the SVD uses orthonormal bases, whereas the eigenvalue decomposition uses a basis that generally is not orthogonal. A third is that not all matrices (even square ones) have an eigenvalue decomposition, but all matrices (even rectangular ones) have a singular value decomposition, as we established in Theorem 4.1. In applications, eigenvalues tend to be relevant to problems involving the behavior of iterated forms of A, such as matrix powers A^k or exponentials e^{tA}, whereas singular vectors tend to be relevant to problems involving the behavior of A itself, or its inverse.

Matrix Properties via the SVD

The power of the SVD becomes apparent as we begin to catalogue its connections with other fundamental topics of linear algebra. For the following theorems, assume that A has dimensions $m \times n$. Let p be the minimum of m and n, let $r \le p$ denote the number of nonzero singular values of A, and let $\langle x, y, \ldots, z\rangle$ denote the space spanned by the vectors $x, y, \ldots, z$.

Theorem 5.1. *The rank of A is r, the number of nonzero singular values.*

Proof. The rank of a diagonal matrix is equal to the number of its nonzero entries, and in the decomposition $A = U\Sigma V^*$, U and V are of full rank. Therefore $\operatorname{rank}(A) = \operatorname{rank}(\Sigma) = r$. □

Theorem 5.2. $\operatorname{range}(A) = \langle u_1, \ldots, u_r\rangle$ *and* $\operatorname{null}(A) = \langle v_{r+1}, \ldots, v_n\rangle$.

Proof. This is a consequence of the fact that $\operatorname{range}(\Sigma) = \langle e_1, \ldots, e_r\rangle \subseteq \mathbb{C}^m$ and $\operatorname{null}(\Sigma) = \langle e_{r+1}, \ldots, e_n\rangle \subseteq \mathbb{C}^n$. □

Theorem 5.3. $\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.

Proof. The first result was already established in the proof of Theorem 4.1: since $A = U\Sigma V^*$ with unitary U and V, $\|A\|_2 = \|\Sigma\|_2 = \max\{|\sigma_j|\} = \sigma_1$, by Theorem 3.1. For the second, note that by Theorem 3.1 and the remark following, the Frobenius norm is invariant under unitary multiplication, so $\|A\|_F = \|\Sigma\|_F$, and by (3.16), this is given by the stated formula. □

Theorem 5.4. *The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*. (These matrices have the same nonzero eigenvalues.)*

Proof. From the calculation

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^*U^*U\Sigma V^* = V(\Sigma^*\Sigma)V^*,$$

we see that A^*A is similar to $\Sigma^*\Sigma$ and hence has the same n eigenvalues (see Lecture 24). The eigenvalues of the diagonal matrix $\Sigma^*\Sigma$ are $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$, with $n - p$ additional zero eigenvalues if $n > p$. A similar calculation applies to the m eigenvalues of AA^*. □

Theorem 5.5. *If $A = A^*$, then the singular values of A are the absolute values of the eigenvalues of A.*

Proof. As is well known (see Exercise 2.3), a hermitian matrix has a complete set of orthogonal eigenvectors, and all of the eigenvalues are real. An equivalent statement is that (5.1) holds with X equal to some unitary matrix Q and Λ a real diagonal matrix. But then we can write

$$A = Q\Lambda Q^* = Q|\Lambda|\operatorname{sign}(\Lambda)Q^*, \tag{5.2}$$

where $|\Lambda|$ and $\operatorname{sign}(\Lambda)$ denote the diagonal matrices whose entries are the numbers $|\lambda_j|$ and $\operatorname{sign}(\lambda_j)$, respectively. (We could equally well have put the factor $\operatorname{sign}(\Lambda)$ on the left of $|\Lambda|$ instead of the right.) Since $\operatorname{sign}(\Lambda)Q^*$ is unitary whenever Q is unitary, (5.2) is an SVD of A, with the singular values equal to the diagonal entries of $|\Lambda|$, $|\lambda_j|$. If desired, these numbers can be put into nonincreasing order by inserting suitable permutation matrices as factors in the left-hand unitary matrix of (5.2), Q, and the right-hand unitary matrix, $\operatorname{sign}(\Lambda)Q^*$. □

Theorem 5.6. *For $A \in \mathbb{C}^{m\times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.*

Proof. The determinant of a product of square matrices is the product of the determinants of the factors. Furthermore, the determinant of a unitary matrix is always 1 in absolute value; this follows from the formula $U^*U = I$ and the property $\det(U^*) = (\det(U))^*$. Therefore,

$$|\det(A)| = |\det(U\Sigma V^*)| = |\det(U)|\,|\det(\Sigma)|\,|\det(V^*)| = |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i.$$

□

Low-Rank Approximations

But what *is* the SVD? Another approach to an explanation is to consider how a matrix A might be represented as a sum of rank-one matrices.

Theorem 5.7. *A is the sum of r rank-one matrices:*

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*. \tag{5.3}$$

Proof. If we write Σ as a sum of r matrices Σ_j, where $\Sigma_j = \mathrm{diag}(0, \ldots, 0, \sigma_j, 0, \ldots, 0)$, then (5.3) follows from (4.3). □

There are many ways to express an $m \times n$ matrix A as a sum of rank-one matrices. For example, A could be written as the sum of its m rows, or its n columns, or its mn entries. For another example, Gaussian elimination reduces A to the sum of a full rank-one matrix, a rank-one matrix whose first row and column are zero, a rank-one matrix whose first two rows and columns are zero, and so on.

Formula (5.3), however, represents a decomposition into rank-one matrices with a deeper property: *the νth partial sum captures as much of the energy of A as possible.* This statement holds with "energy" defined by either the 2-norm or the Frobenius norm. We can make it precise by formulating a problem of best approximation of a matrix A by matrices of lower rank.

Theorem 5.8. *For any ν with $0 \le \nu \le r$, define*

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^*; \tag{5.4}$$

if $\nu = p = \min\{m, n\}$, define $\sigma_{\nu+1} = 0$. Then

$$\|A - A_\nu\|_2 \;=\; \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ \mathrm{rank}(B) \le \nu}} \|A - B\|_2 \;=\; \sigma_{\nu+1}.$$

Proof. Suppose there is some B with $\operatorname{rank}(B) \le \nu$ such that $\|A - B\|_2 < \|A - A_\nu\|_2 = \sigma_{\nu+1}$. Then there is an $(n-\nu)$-dimensional subspace $W \subseteq \mathbb{C}^n$ such that $w \in W \Rightarrow Bw = 0$. Accordingly, for any $w \in W$, we have $Aw = (A-B)w$ and

$$\|Aw\|_2 = \|(A - B)w\|_2 \le \|A - B\|_2 \, \|w\|_2 < \sigma_{\nu+1}\|w\|_2.$$

Thus W is an $(n - \nu)$-dimensional subspace where $\|Aw\| < \sigma_{\nu+1}\|w\|$. But there is a $(\nu + 1)$-dimensional subspace where $\|Aw\| \ge \sigma_{\nu+1}\|w\|$, namely the space spanned by the first $\nu + 1$ right singular vectors of A. Since the sum of the dimensions of these spaces exceeds n, there must be a nonzero vector lying in both, and this is a contradiction. □

Theorem 5.8 has a geometric interpretation. What is the best approximation of a hyperellipsoid by a line segment? Take the line segment to be the longest axis. What is the best approximation by a two-dimensional ellipsoid? Take the ellipsoid spanned by the longest and the second-longest axis. Continuing in this fashion, at each step we improve the approximation by adding into our approximation the largest axis of the hyperellipsoid not yet included. After r steps, we have captured all of A. This idea has ramifications in areas as disparate as image compression (see Exercise 9.3) and functional analysis.

We state the analogous result for the Frobenius norm without proof.

Theorem 5.9. *For any ν with $0 \le \nu \le r$, the matrix A_ν of* (5.4) *also satisfies*

$$\|A - A_\nu\|_F \;=\; \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ \operatorname{rank}(B) \le \nu}} \|A - B\|_F \;=\; \sqrt{\sigma_{\nu+1}^2 + \cdots + \sigma_r^2}.$$

Computation of the SVD

In this and the previous lecture, we have examined the properties of the SVD but not considered how it can be computed. As it happens, the computation of the SVD is a fascinating subject. The best methods are variants of algorithms used for computing eigenvalues, and we shall discuss them in Lecture 31.

Once one can compute it, the SVD can be used as a tool for all kinds of problems. In fact, most of the theorems of this lecture have computational consequences. The best method for determining the rank of a matrix is to count the number of singular values greater than a judiciously chosen tolerance (Theorem 5.1). The most accurate method for finding an orthonormal basis of a range or a nullspace is via Theorem 5.2. (For both of these examples, QR factorization provides alternative algorithms that are faster but not always as accurate.) Theorem 5.3 represents the standard method for computing $\|A\|_2$, and Theorems 5.8 and 5.9, the standards for computing low-rank approximations with respect to $\|\cdot\|_2$ and $\|\cdot\|_F$. Besides these examples, the SVD is also an ingredient in robust algorithms for least squares fitting, intersection of subspaces, regularization, and numerous other problems.

Exercises

5.1. In Example 3.1 we considered the matrix (3.7) and asserted, among other things, that its 2-norm is approximately 2.9208. Using the SVD, work out (on paper) the exact values of $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ for this matrix.

5.2. Using the SVD, prove that any matrix in $\mathbb{C}^{m\times n}$ is the limit of a sequence of matrices of full rank. In other words, prove that the set of full-rank matrices is a dense subset of $\mathbb{C}^{m\times n}$. Use the 2-norm for your proof. (The norm doesn't matter, since all norms on a finite-dimensional space are equivalent.)

5.3. Consider the matrix

$$A = \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix}.$$

(a) Determine, on paper, a real SVD of A in the form $A = U\Sigma V^T$. The SVD is not unique, so find the one that has the minimal number of minus signs in U and V.

(b) List the singular values, left singular vectors, and right singular vectors of A. Draw a careful, labeled picture of the unit ball in $\mathbb{R}^2$ and its image under A, together with the singular vectors, with the coordinates of their vertices marked.

(c) What are the 1-, 2-, ∞-, and Frobenius norms of A?

(d) Find A^{-1} not directly, but via the SVD.

(e) Find the eigenvalues λ_1, λ_2 of A.

(f) Verify that $\det A = \lambda_1\lambda_2$ and $|\det A| = \sigma_1\sigma_2$.

(g) What is the area of the ellipsoid onto which A maps the unit ball of $\mathbb{R}^2$?

5.4. Suppose $A \in \mathbb{C}^{m\times m}$ has an SVD $A = U\Sigma V^*$. Find an eigenvalue decomposition (5.1) of the $2m \times 2m$ hermitian matrix

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}.$$