

Lecture Notes for Math 104: Fall 2010 (Week 8)

Jacob Bernstein

Twenty-Second and Twenty-Third Lectures

We discussed some applications of the SVD.

1. Applications of the SVD

If we know the SVD of a matrix there is lots of useful information we can deduce about the matrix A . Let $A \in \mathbb{C}^{m \times n}$ and suppose that A has the SVD

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

Let us write $p = \min(m, n)$ so p is the number of singular values of A then let $r \leq p$ denote the number of non-zero singular values. We denote by $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ the non-zero singular values of A and

$$U = [\mathbf{u}_1 \mid \dots \mid \mathbf{u}_m] \text{ and } V = [\mathbf{v}_1 \mid \dots \mid \mathbf{v}_n]$$

the left and right singular vectors.

THEOREM 1.1. *The rank of A is r the number of non-zero singular values.*

PROOF. As Σ is diagonal, it is immediate that $\mathbf{e}_1, \dots, \mathbf{e}_r$ is a basis of $R(\Sigma)$. Since U, V are intertible one then has that $U\mathbf{e}_i$ is a basis of $R(A)$. \square

A more refined result is the following:

THEOREM 1.2. $R(A) = \text{span}(\mathbf{u}_1, \dots, \mathbf{u}_r)$ and $N(A) = \text{span}(\mathbf{v}_{r+1}, \dots, \mathbf{v}_n)$

PROOF. For a diagonal matrix $R(\Sigma) = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_r)$ while $N(\Sigma) = \text{span}(\mathbf{e}_{r+1}, \dots, \mathbf{e}_n)$. The range of A has a basis $U\mathbf{e}_i = \mathbf{u}_i$ for $1 \leq i \leq r$. Similarly, by solving $V^*\mathbf{x} = \mathbf{e}_i$ for $r + 1 \leq i \leq n$ one obtains vectors in $N(A)$. We see that the solutions to this equation is $V\mathbf{e}_i = \mathbf{v}_i$. \square

REMARK 1.3. Notice that this actually gives an orthonormal basis of $R(A)$. Namely, $\mathbf{u}_1, \dots, \mathbf{u}_r$. This is *NOT* necessarily the same as the one obtained via QR factorization.

As a consequence, if one sets

$$\hat{U} = [\mathbf{u}_1 \mid \dots \mid \mathbf{u}_r]$$

then one has the matrix $P = \hat{U}\hat{U}^*$ giving orthogonal projection onto $R(A)$. In particular, we can use the SVD to solve least squares problems.

We can also use the SVD to compute 2-norms. Indeed,

THEOREM 1.4. $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$.

PROOF. To see this we note that both these norms are invariant under pre- and post- multiplication by unitary maps, so $\|A\|_2 = \|\Sigma\|_2$ and $\|A\|_F = \|\Sigma\|_F$. Since Σ is diagonal it is easy to compute the norms in this case. \square

One very important application is the SVD is that it allows one to get a good approximations of a given matrix in terms of lower rank matrices. This is important in trying to understand what the “dominant” part of the matrix is. It can also be thought of in terms of how much “compression” can be applied to the matrix.

The basic idea is that can express A is the sum of r rank one matrices

$$A = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^*$$

Which follows just by multiplying out the SVD. The point is there are lots of ways to write A as a sum of rank one matrices for instance

$$A = \sum_{j=1}^n \mathbf{a}_j \mathbf{e}_j^*$$

or

$$A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \mathbf{e}_i \mathbf{e}_j^*$$

. However, the sum given by the SVD has the property of having the k th partial sum capturing as much “energy” of A as possible, that is of being the best approximation possible in the induced 2-norm or Frobenius norm.

To make this precise

THEOREM 1.5. For any k with $0 \leq k \leq r$ define

$$A_k = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^*$$

so $A_r = A$. Then:

$$\|A - A_k\|_2 = \inf_{B \in \mathbb{C}^{m \times n} \text{rank}(B) \leq k} \|A - B\|_2 = \sigma_{k+1}$$

Here if $k = p = \min(m, n)$ we set $\sigma_{k+1} = 0$.

REMARK 1.6. That is we have that A_k is the best (in terms of the induced 2 norm approximation of A by a rank k matrix).

PROOF. Suppose one has a $B \in \mathbb{C}^{m \times n}$ with $\text{rank}(B) \leq k$ and $\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$. By the rank-nullity theorem we see that there is a $(n - k)$ -dimensional space W in \mathbb{C}^n so that for $\mathbf{w} \in W$, $B\mathbf{w} = 0$. (i.e. $W \subset N(B)$). Now for $\mathbf{w} \in W$ $A\mathbf{w} = (A - B)\mathbf{w}$ and so

$$\|A\mathbf{w}\|_2 = \|(A - B)\mathbf{w}\|_2 \leq \|A - B\|_2 \|\mathbf{w}\|_2 < \sigma_{k+1} \|\mathbf{w}\|_2$$

Thus W is an $(n - k)$ dimensional subspace with $\|A\mathbf{w}\|_2 < \sigma_{k+1} \|\mathbf{w}\|_2$. However, by considering $W' = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$ one obtains a $k + 1$ dimensional space (all vectors are orthogonal hence linearly independent) with $\|A\mathbf{w}'\|_2 \geq \sigma_{k+1} \|\mathbf{w}'\|_2$ for all $\mathbf{w}' \in W'$. Now W' and W must have a non-zero vector in common (otherwise one would get $n + 1$ linearly independent vectors in \mathbb{C}^n . But this is a contradiction. \square

Notice when σ_{k+1} is small this means that $\|A\mathbf{x} - A_k\mathbf{x}\|_2$ is small (at least relative to $\|\mathbf{x}\|_2$). I.e. multiplication by A is well approximated by multiplication by A_k . A similar result also holds for the Frobenius norm i.e.

THEOREM 1.7. For any k with $0 \leq k \leq r$ one has

$$\|A - A_k\|_F = \inf_{B \in \mathbb{C}^{m \times n} \text{rank}(B) \leq k} \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$$

Notice that when $\sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$ is small all of the entries of A are close to the entries of A_k . That is the array of numbers making up A are all well approximated by the array of numbers making up A_k . Notice that A takes mn numbers to determine (i.e. each entry) while A_k takes $(m+n+1)k$ to represent (i.e. the left and right singular vector and the singular value). If k is small relative to $p = \min(m, n)$ this is a significant savings.

2. Least squares via SVD: NIC

As we've seen the SVD of a matrix A gives a orthonormal basis of $R(A)$. More than that it gives an approach to solving least squares problems.

Assume that $A \in \mathbb{C}^{m \times n}$ with $m > n$. We assume also that $N(A) = \{0\}$ though this isn't necessary. We want to solve the overdetermined problem

$$A\mathbf{x} = \mathbf{b}$$

in a least squares sense using the SVD. To that end, let A have reduced SVD

$$A = \hat{U}\hat{\Sigma}V^*$$

with

$$\hat{U} = [\mathbf{u}_1 | \dots | \mathbf{u}_n] \in \mathbb{C}^{m \times n}$$

Now orthogonal projection onto $R(A)$ is given by $P = \hat{U}\hat{U}^*$. Hence to solve the equation in the least squares sense it is enough to solve

$$A\mathbf{x} = P\mathbf{b}$$

but this leads to

$$\hat{U}\hat{\Sigma}V^*\mathbf{x} = \hat{U}\hat{U}^*\mathbf{b}$$

since the columns of \hat{U} are linearly independent this is equivalent to solving

$$\hat{\Sigma}V^*\mathbf{x} = \hat{U}^*\mathbf{b}$$

But this consists just of solving a diagonal system and multiplying by a unitary matrix.

CHAPTER 2

Twenty-Fourth Lecture

In this lecture we recall some definitions related to the study of eigenvectors and eigenvalues. This will allow us to compute the SVD of a matrix by solving a related eigenvalue problem (which is slightly more algebraically tractable).

1. Eigenvalues and Eigenvectors

We review some of the very important linear concept of eigenvectors and eigenvalues. It is helpful to compare and contrast these with singular vectors and singular values.

Recall that for $A \in \mathbb{C}^{m \times m}$ we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* and $0 \neq \mathbf{v} \in \mathbb{C}^m$ is an *eigenvector* if

$$A\mathbf{v} = \lambda\mathbf{v}$$

that is multiplication of \mathbf{v} by A scales \mathbf{v} by λ . We call the set of all eigenvalues of A , $\Lambda(A)$ the *spectrum* of A . We point out that if A is singular then $0 \in \Lambda(A)$ and any non-zero vector in $N(A)$ is then an eigenvector (with eigenvalue 0).

When everything is real – i.e. both the matrix A and the eigenvalue λ and eigenvectors \mathbf{x} – then one can geometrically understand this as saying A scales \mathbf{x} by $|\lambda|$ (and possibly reverses its direction if $\lambda < 0$). However, it is possible for A to be real and for λ and \mathbf{v} to be complex. This contrasts with much of what we have seen previously and should be kept in mind. For instance the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

which geometrically rotates by 90° has eigenvalues $\pm\sqrt{-1} = \pm i$. Roughly speaking, for real matrices, complex eigenvalues correspond geometrically to such a “rotation” (possible also with a scaling) while real eigenvalues correspond to pure scaling.

How do we find eigenvalues and eigenvectors? We can recast the question slightly and see that we are trying to find non-trivial solutions to

$$(A - \lambda I)\mathbf{x} = 0$$

That is we try to find λ so that $N(A - \lambda I) \neq \{0\}$ and then in this case try and find elements in the null space. The latter problem is easy (as it is just solving a linear system) and the difficulty arises mostly in the former. Indeed, determining the spectrum $\Lambda(A)$ is an essentially non-linear problem.

Phrase things in a manner that is amenable to algebraic investigation we must recall the determinant. This is a function:

$$\det : \mathbb{C}^{m \times m} \rightarrow \mathbb{C}$$

defined by

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

and inductively by

$$\det A = \sum_{i=1}^m (-1)^i a_{1i} \det A_{1i}.$$

Here A_{1i} is the matrix in $\mathbb{C}^{(m-1) \times (m-1)}$ obtained by omitting the first row and i th column. In other words, we have defined the determinant by expanded along the first row. The determinant has many properties that allow one to compute it in other ways. We refer to Strang for instance for more detailed discussion. You should be able to compute the determinant of small matrices (i.e. 2×2 and 3×3).

There is a big theory of determinants. The main property we will need is the fact that A is non-singular when and only when $\det(A) \neq 0$. Using this fact and an expansion of the determinant we see that $\lambda \in \Lambda(A)$ when and only when λ is a root of the polynomial

$$p_A(z) = \det(zI - A) = z^m + c_{m-1}z^{m-1} + \dots + c_0$$

is a degree m polynomial. We call this the *characteristic* polynomial of A . The coefficients c_i are determined by the entries of A in an explicit (but non-linear) way.

This is one place that working over \mathbb{C} greatly simplifies things. Indeed, the fundamental theorem of algebra tells us that over \mathbb{C} . $p_A(z)$ has exactly m roots (counting multiplicity). That is we can factor

$$p_A(z) = (z - \lambda_1)^{m_1} \dots (z - \lambda_k)^{m_k}$$

where $k \leq m$, $m_i \geq 1$ and $\sum_i m_i = m$. We call the value $m_i = m_{\lambda_i}$ the *algebraic multiplicity* of the eigenvalue λ_i . Notice one cannot always produce such a factorization over \mathbb{R} .

For a $\lambda \in \Lambda(A)$ we say the *eigenspace* associated λ is the vector space

$$E_\lambda = N(\lambda I - A)$$

this is always a non-empty vector space all the non-zero vectors of E_λ are eigenvectors with eigenvalue λ . We let $g_\lambda = \dim E_\lambda$ and call this number the *geometric multiplicity* of λ . One always has $1 \leq g_\lambda \leq m_\lambda$ (for a proof we refer to Trefethen-Bau Lecture 24). We say A is *non-defective* if $g_\lambda = m_\lambda$ for all $\lambda \in \Lambda(A)$.

The point above is that for a non-defective $A \in \mathbb{C}^{m \times m}$ one has the dimensions of the eigenspaces summing up to m (since the algebraic multiplicites have this property). In this case there is a set $\mathbf{x}_1, \dots, \mathbf{x}_m$ a basis of \mathbb{C}^m where each \mathbf{x}_i is an eigenvector of A . In particular, there is a non-singular matrix

$$X = [\mathbf{x}_1 | \dots | \mathbf{x}_m]$$

so that

$$A = X \Lambda X^{-1}$$

where Λ is diagonal. Notice that unlike the SVD we have only one set of vectors, they are not necessarily orthogonal, we must start with a square matrix, the diagonal matrix may have complex or negative entries and we aren't guaranteed of such a decomposition existing.

One important result we will need is the following:

THEOREM 1.1. *Let $A \in \mathbb{C}^{m \times m}$ be hermitian, i.e. $A^* = A$. Then A is non-defective, all the eigenvalues of A are real and one may choose a orthonormal basis of eigenvectors.*

COROLLARY 1.2. *There is a $Q \in \mathbb{C}^{m \times m}$ that is unitary so that*

$$A = Q\Lambda Q^*$$

where Λ is diagonal with real entries.

We will prove this later. The point is that hermitian matrices are rather nice from an eigenvalue point of view.

2. Eigenvalues and the SVD

Despite the differences noted above, there is a clear important relationship between eigenvalues and singular values. Indeed, for hermitian matrices they are (practically) the same. One thing that is useful about this is that eigenvalues and eigenvectors can be found algebraically (though this is not an easy problem for large matrices). This allows one to find singular values in an algebraic manner:

THEOREM 2.1. *Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ be the non-zero singular values of $A \in \mathbb{C}^{m \times n}$. Then σ_i^2 are precisely the non-zero eigenvalues of A^*A and of AA^* (i.e. these matrices have the same non-zero eigenvalues).*

PROOF. : Let A have SVD $A = U\Sigma V^*$ then

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^*U^*U\Sigma V^* = V\Sigma^*\Sigma V^* = V\Sigma^2V^*$$

This says exactly that A^*A has eigenvalues $\sigma_1^2, \dots, \sigma_r^2$ with associated eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_r$. Of course there may be more eigenvalues but these must all be zero. Similarly,

$$AA^* = (U\Sigma V^*)(U\Sigma V^*)^* = U\Sigma V^*V\Sigma^*U^* = U\Sigma\Sigma^*U^* = U\Sigma^2U^*$$

This says that AA^* has eigenvalues $\sigma_1^2, \dots, \sigma_r^2$ with associated eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_r$. As above there may be more eigenvalues but they are zero. \square

It is important to note that this given the eigenvalues of A^*A one gets the singular values by taking the (positive) square root. More over, by taking the associated eigenvectors of A^*A one gets the right singular vectors of A .