# Lecture Notes for Math 104: Fall 2010 (Week 7)

Jacob Bernstein

CHAPTER 1

# Nineteenth and Twentieth Lectures

In these two lectures we look at other notions of length of vectors then the 2-norm. We also discuss notions of length for matrices.

## 1. Vector Norms

We are familiar with the two norm already.

$$||\mathbf{v}||_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^* \mathbf{v}}$$

We interpret this as the length of the vector $\mathbf{v}$. Some important properties of the two norm are that

$$\begin{aligned} ||\mathbf{v}||_2 \geq 0 \text{ and } ||\mathbf{v}||_2 = 0 \iff \mathbf{v} = 0 \\ ||\lambda \mathbf{v}||_2 = |\lambda| ||\mathbf{v}||_2 \\ ||\mathbf{v} + \mathbf{w}||_2 \leq ||\mathbf{v}||_2 + ||\mathbf{w}||_2. \end{aligned}$$

It is sometimes necessary to have other notions of length besides the 2-norm. To do this we take the three preceeding properties as a definition. We say a function $|| \cdot || : \mathbb{C}^m \to \mathbb{R}$ is a *norm* if

$$\begin{aligned} ||\mathbf{v}|| \geq 0 \text{ and } ||\mathbf{v}|| = 0 \iff \mathbf{v} = 0 \\ ||\lambda \mathbf{v}|| = |\lambda| ||\mathbf{v}|| \\ ||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||. \end{aligned}$$

There are many norms. For instance: The $p$-norms, let $\mathbf{x} = \sum_{i=1}^m x_i \mathbf{e}_i$

$$\begin{aligned} ||\mathbf{x}||_p := \left( \sum_{i=1}^m |x_i|^p) \right)^{1/p} \quad (1 \leq p < \infty) \\ ||\mathbf{x}||_\infty : - = \max_{1 \leq i \leq m} |x_i| \end{aligned}$$

Note that $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^* \mathbf{x}}$ which agrees with the usual notion of 2 norm. It is a good exercise to check that the $\infty$-norm is a actually a norm. There are lots of other norms for instance let $W \in \mathbb{C}^{m \times m}$ be a diagonal matrix with positive entries $w_{ii}$ on the diagonal We can define

$$||\mathbf{x}||_W = ||W\mathbf{x}||_2 = \sqrt{\sum_{1=1}^m |w_{ii} x_i|^2}$$

Unit ball is then some sort of ellipse.

## 2. Induced Matrix Norms

Associated to any pair of norms $||\cdot||_{(n)}$ on $\mathbb{C}^n$ and $||\cdot||_{(m)}$ on $\mathbb{C}^m$ (not necessarily $p$-norms) there is an *induced matrix norm*, $||\cdot||_{(m,n)}$ on $\mathbb{C}^{m\times n}$ This norm measures the maximum amount of "stretching" (as measured by the norms on $\mathbb{C}^m$ and $\mathbb{C}^n$) that multiplication by $A$ can achieve that is

$$||A||_{(m,n)} = \sup_{\mathbf{x}\in\mathbb{C}^n, \mathbf{x}\neq 0} \frac{||A\mathbf{x}||_{(m)}}{||\mathbf{x}||_{(n)}} = \sup_{\mathbf{x}\in\mathbb{C}^n, ||\mathbf{x}||_{(n)}=1} \frac{||A\mathbf{x}||_{(m)}}{||\mathbf{x}||_{(n)}}.$$

We leave it as an excercise to see that the two definitions are equivalent. Another way to think about this is to note that the induced norm is the smallest value $C$ so that

$$||A\mathbf{x}||_{(m)} \leq C||\mathbf{x}||_{(n)}$$

for all $\mathbf{x}\in\mathbb{C}^n$. In general this definition is hard to use computationally (as formulated it is not an algebraic property). It is very intuitive though and has good mathematical properties.

We will often consider the case when $||\cdot||_{(n)} = ||\cdot||_p$ and $||\cdot||_{(m)} = ||\cdot||_p$ (i.e. both norms are $p$-norms). We then write $||A||_p$ instead of $||A||_{(m,n)}$. A simple example. Suppose that $m = n$ and $A$ is a diagonal matrix

$$A = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ \vdots & & \ddots & \\ & & & a_m \end{bmatrix}$$

Then $||A||_p = \max 1 \leq i \leq m|a_i|$. When $p = 2$ we can see this geometrically. As $A$ maps a circle to an ellipse. And the longest vector in the image is the biggest axis.

Another example. Lets compute the 1-norm of a matrix. This turns out to be easy to determine in terms of the lengths of columns. We calim that with

$$A = \begin{bmatrix}\mathbf{a}_1 & |\cdots & |\mathbf{a}_n\end{bmatrix}$$

one has

$$||A||_1 = \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

To see this we calculate for $\mathbf{x} = \sum_{j=1}^n x_j\mathbf{e}_j$ with $||\mathbf{x}||_1 = 1$. In this case we see that $\sum_{j=1} |x_j| = 1$. Then

$$||A\mathbf{x}||_1 = ||\sum_{j=1}^n x_j\mathbf{a}_j||_1 \leq \sum_{j=1}^n ||x_j\mathbf{a}_j||_1 = \sum_{j=1}^n |x_j|||\mathbf{a}_j||_1$$

But then

$$\leq \left(\max_{1\leq j\leq n} ||\mathbf{a}_j||_1\right)\sum_{j=1}^n |x_j| = \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

This implies

$$||A||_1 \leq \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

To get the equality we we suppose the maximum is achieved at the $j_0$ column i.e.

$$\max_{1\leq j\leq n} ||\mathbf{a}_j||_1 = ||\mathbf{a}_{j_0}||_1$$

then with $\mathbf{x}_0 = \mathbf{e}_{j_0}$ one has $||\mathbf{x}_0||_1 = 1$ and $||A\mathbf{x}_0||_1 = ||\mathbf{a}_{j_0}||_1$. In a similar fashion one can show that

$$||A||_\infty = \max_{1 \leq i \leq m} ||\mathbf{a}_i^*||_1$$

I.e. is the maximum length of the rows. We leave this as an exercise.

Computing matrix $p$-norms for $1 < p < \infty$ is much harder. We will see a method to do this for 2 norms (which is the most important). One fact that can be useful in at least getting a bound on induced norms is a generalization of the Cauchy-Schwarz inequality called H olders Inequality:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q$$

provided $1/p + 1/q = 1$. When $p = q = 2$ this is the Cauchy-Schwarz inequality.

## 3. General Matrix Norms

There are many more norms on matrices then just the induced norms. In general we say a map $||\cdot|| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a matrix norm if it is just a norm on the vector space $\mathbb{C}^{mn}$. That is

$$||A|| \leq 0 \text{ and } ||A|| = 0 \iff A = 0$$
$$||A + B|| \leq ||A|| + ||B||$$
$$||\lambda A|| = |\lambda| ||A||$$

It is easy to see any induced norm satisfies these conditions.

One important norm that is not an induced norm is the so called *Frobenius norm*. This is given by

$$||A||_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \left( \sum_{j=1}^n ||\mathbf{a}_j||_2^2 \right)^{1/2}$$

Which is just the 2-norm on $\mathbb{C}^{mn}$.

One other way to compute this (which is useful from a theoretical point of view) is

$$||A||_F = \sqrt{tr(A^*A)} = \sqrt{tr(AA^*)}$$

Here $tr(A) = \sum_{i=1}^{\min(m,n)} a_{ii}$. It is a simple exercise to check this.

General matrix norms do not interact with matrix multiplication. However, for induced norms and the Frobenius norm the norm of the product is controlled by the product of the norms. Indeed,

$$||AB||_{(l,n)} \leq ||A||_{(l,m)} ||B||_{(m,n)}$$

here $A \in \mathbb{C}^{l\times m}$ and $B \in \mathbb{C}^{m\times n}$. To see this just consider

$$||A\mathbf{x}||_{(l)} \leq ||A||_{(l,m)} ||B\mathbf{x}||_{(m)} \leq ||A||_{(l,m)} ||B||_{(m,n)} ||\mathbf{x}||_{(n)}$$

and

$$||AB||_F \leq ||A||_F ||B||_F$$

Final useful property the induced 2-norm and the Frobenius norm is that they are invariant under pre- or post-multiplication by a unitary matrix. That is let $Q \in \mathbb{C}^{m\times m}$ and $Q' \in \mathbb{C}^{n\times n}$ both be unitary. Then for $A \in \mathbb{C}^{m\times n}$ one has

$$||QA||_2 = ||A||_2 \text{ and } ||QA||_F = ||A||_F$$

and

$$||AQ'||_2 = ||A||_2 \text{ and } ||AQ'||_F = ||A||_F.$$

CHAPTER 2

# Twenty-First Lecture

We introduce the singular value decomposition (SVD).

## 1. What is the SVD: A Geometric point of view

The SVD is a factorization of an arbitrary matrix that follows from geometric properties of linear maps. In particular, one tries to understand what the image of the unit sphere is under multiplication by $A$. To work geometrically we first focus on the reals. To that end, let $A \in \mathbb{R}^{m\times n}$ and consider the unit sphere in $\mathbb{R}^n$ i.e. the vectors $\mathbf{x} \in \mathbb{R}^n$ with $||\mathbf{x}||_2 = 1$. We denote this set by $S$ and the consider it is $AS$. Formally, $AS = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = A\mathbf{x}, \mathbf{x} \in S\}$. We claim that $AS$ is in general a hyperellipse (i.e. a higher dimenisonal analog of an ellipse).

For $n = m = 2$ this means that $S$ is the unit circle and $AS$ should be a rotation and stretching of some ellipse. Note that we are allowed to stretch so much that $AS$ is actually a line segment.

To be more precise we suppose that $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ and suppose also that $A$ has full rank (i.e. the columns linearly independent). We define the *singular values* $\sigma_1, \sigma_2, \ldots, \sigma_n$ to be the length of the principal semiaxes of $AS$. We usually order these so $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$. Note that $\sigma_n > 0$ as $N(A) = \{0\}$.

We define the *left singular vectors* of $A$ to be the set of orthogonal unit vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ in $\mathbb{C}^m$ so that $\sigma_i\mathbf{u}_i$ is a principal semiaxis of $AS$. In particular $\sigma_1\mathbf{u}_1$ is the largest semiaxis of $AS$. The *right singular vectors* are the set of orthgonal unit vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{C}^n$ so that $A\mathbf{v}_i = \mathbf{u}_i$. As $N(A) = \{0\}$ the $\mathbf{u}_i$ are unique. We mention that it is not a priori clear that the $\mathbf{u}_i$ need to be orthgonal, this is however true and is something we will show.

In terms of matrices:

$$AV = \hat{U}\Sigma$$

Here

$$V = \begin{bmatrix}\mathbf{v}_1| & \cdots & |\mathbf{v}_n\end{bmatrix} \in \mathbb{C}^{n\times n}$$

While

$$\hat{U} = \begin{bmatrix}\mathbf{u}_1| & \cdots & |\mathbf{u}_n\end{bmatrix} \in \mathbb{C}^{m\times n}$$

and

$$\Sigma = \begin{bmatrix}\sigma_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & & \sigma_n\end{bmatrix}$$

This means one has a *reduced SVD* factorization:

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

As with the $QR$ factorization we can form a full SVD by adding additionally columns to $\hat{U}$ to make a unitary square matrix $U$. This requires adding additional zeros to $\Sigma$. This gives

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

## 2. What is the SVD: an Algebraic Point of View

While the geometric point of view discussed above is imporatant to understanding the SVD it is hard to make rirgourous (and not easy to compute with). We will now discuss a more algebraic point of view.

We let $m, n$ now be arbitary integers and let $A \in \mathbb{C}^{m\times n}$ also be arbitary. A (full) *Singular Value Decomposition* of $A$ is a factorization

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

Where $U \in \mathbb{C}^{m\times m}$ is unitary. $V \in \mathbb{C}^{n\times n}$ is unitary and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal. We assume in addition that the diagonal elements of $\Sigma$ are non-negative and ordered so $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ where $p = \min(m, n)$. That is we can write (here we have $m = n$):

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \\ 0 & \cdots & 0 & \sigma_p \end{bmatrix}$$

Notice that it is then clear that the image of the unit sphere under $A$ is a hyperellipse.

The issue now is to see whether every matrix admits a singular value decomposition. This turns out to always be the case:

THEOREM 2.1. *Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition. Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined and if $A$ is square and the $\sigma_j$ are distinct the ne the left and right singular vectors $\{u_j\}$ and $\{v_j\}$ are uniquely determined up to (complex) sign.*

PROOF. The method of proof is an induction on the dimension of $A$ where what me really mean is an induction on $l = \min(m, n)$. Set $\sigma_1 = ||A||_2$. Because the unit sphere is a compact and the map $\mathbf{x} \to ||A\mathbf{x}||_2$ is continuous there must be vectors $\mathbf{v}_1' \in \mathbb{C}^n$ and $\mathbf{u}_1' \in \mathbb{C}^m$ with $||\mathbf{v}_1'||_2 = ||\mathbf{u}_1'||_2 = 1$ and so that $A\mathbf{v}_1' = \sigma_1\mathbf{u}_1'$. You should take this for granted as it is beyond the scope of this class to discuss it further. Consider a basis extension of $\mathbf{v}_1'$ to $\left\{\mathbf{v}_j'\right\}$ an orthonormal basis of $\mathbb{C}^n$ and a basis extension of $\mathbf{u}_1'$ to $\left\{\mathbf{u}_j'\right\}$ an orthonormal basis of $\mathbb{C}^m$. Let $U_1$ and $V_1$ denote the matrices with columns $\left\{\mathbf{v}_j'\right\}$ and $\left\{\mathbf{u}_j'\right\}$. Then one has

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & \mathbf{w}^* \\ 0 & B \end{bmatrix}$$

Here $B \in \mathbb{C}^{(m-1)\times(n-1)}$ and $\mathbf{w} \in \mathbb{C}^{n-1}$. One estimates

$$\left|\left|\begin{bmatrix} \sigma_1 & \mathbf{w}^* \\ 0 & B \end{bmatrix}\begin{bmatrix} \sigma_1 \\ w \end{bmatrix}\right|\right|_2 \geq \sigma_1^2 + \mathbf{w}^*\mathbf{w} = (\sigma_1^2 + \mathbf{w}^*\mathbf{w})^{1/2}\left|\left|\begin{bmatrix} \sigma_1 \\ w \end{bmatrix}\right|\right|_2, .$$

This means $\sigma_1 = ||A||_2 = ||S||_2 \geq (\sigma_1^2 + \mathbf{w}^*\mathbf{w})^{1/2}$ which can only occur if $\mathbf{w}^*\mathbf{w} = ||\mathbf{w}||_2^2 = 0$. In particular,

$$U_1^* A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}$$

If $n = 1$ or $m = 1$ we are done – this is the base case $l = 1$. Otherwise $B$ describes an action on $span(\mathbf{v}_1)^\perp$. By the induction hypothesis one has an SVD of $B$

$$B = U_2 \Sigma_2 V_2^*$$

One verifies that

$$A = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^*$$

is an SVD of $A$. The point is that the first two matrices are unitary so their product is also unitary, same true for last two and the middle one is diagonal. Notice that $||B||_2 \leq ||A||_2$ so the singular values are ordered as desired.

To verify uniqueness we note that $\sigma_1$ is uniquely determined by being equal to $||A||_2$. Now suppose that in addition to $\mathbf{v}_1$ there is another (linearly independent) vector $\mathbf{w}$ with $||\mathbf{w}||_2 = 1$ and $||A\mathbf{w}||_2 = \sigma_1$. Let

$$\mathbf{v}_2 = \frac{P_{\mathbf{v}_1^\perp}\mathbf{w}}{||P_{\mathbf{v}_1^\perp}\mathbf{w}||_2}$$

As $||A||_2 = \sigma_1$ one has $||A\mathbf{v}_2||_2 \leq \sigma_1$. We claim this is an equality.

To see this we note that as $\mathbf{v}_1$ is a left singular vector and $\mathbf{v}_2$ is perpendicular to $\mathbf{v}_1$ one has

$$\langle A\mathbf{v}_1, A\mathbf{v}_2 \rangle = 0$$

To see this we note that $A$ has the SVD $A = U\Sigma V^*$ where $\mathbf{v}_1$ is the first column of $V$. Thus $\langle A\mathbf{v}_1, A\mathbf{v}_2 \rangle = \langle A^*A\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \sigma_1^2 \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. Now we can write $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ and since $\mathbf{v}_1$ and $\mathbf{v}_2$ are an orthonormal set $|c_1|^2 + |c_2|^2 = 1$ with both non-zero. Withouth equality the Pythagorean theorem would imply $||A\mathbf{w}||_2 < \sigma_1$ a contradiction. Thus $\mathbf{v}_2$ is a second right singular vector of $A$ corresponding to $\sigma_1$. This implies that the singular values would not be distinct and so cannot occur. The result follows by induction. $\square$