18.409 The Behavior of Algorithms in Practice 2/12/2

Lecture 2

Lecturer: Dan Spielman *Scribe: Steve Weis*

Linear Algebra Review

A n x n matrix has n **singular values**. For a matrix A, the largest singular value is denoted as $\sigma_n(A)$. Similarly, the smallest is denoted as $\sigma_1(A)$. They are defined as follows:

$$\sigma_n(A) = ||A|| = \max_x \frac{||Ax||}{||x||}$$

$$\sigma_1(A) = ||A^{-1}||^{-1} = \min_x \frac{||Ax||}{||x||}$$

There are several other equivalent definitions:

$$\{\sigma_n(A), \ldots, \sigma_1(A)\} = \{\sqrt{\lambda_n(A^TA)}, \ldots, \sqrt{\lambda_1(A^TA)}\}$$

$$\sigma_i(A) = \min_{subspaces S, dim(S)=i} \max_{x \in S} \frac{||Ax||}{||x||} = \max_{subspaces S, dim(S)=(n-i+1)} \min_{x \in S} \frac{||Ax||}{||x||}$$

Another classic definition is to take a unit sphere and apply A to it, resulting in some hyper-ellipse. σ_n will be the length of the largest axis, σ_{n-1} will be the length of the next largest orthogonal axis, etc..

Exercise: Prove that every real matrix A has a *singular-value decompsition* as $A = USV$, where U and V are orthogonal matrices and S is non-negative diagonal, and all entries in U, S, and V are real.

Condition Numbers

The singular values define a **condition number** of a matrix as follows:

$\kappa(A) := \frac{\sigma_n(A)}{\sigma_1(A)} = \frac{||A||}{||A^{-1}||^{-1}}$

Lemma 1. *If* $Ax = b$ *and* $A(x + \delta x) = b + \delta b$ *then*

$$\frac{||\delta x||}{||x||} \leq \kappa(A) \frac{||\delta b||}{||b||}$$

Proof of Lemma 1:

$$A\delta x = \delta b \Rightarrow \delta x = A^{-1}\delta b \Rightarrow ||b|| \leq ||\delta b|| \cdot ||A^{-1}|| = \frac{||\delta b||}{\sigma_1(A)}$$

$$Ax = b \Rightarrow ||b|| \leq ||A|| \cdot ||x|| = \sigma_n(A) \cdot ||x|| \Rightarrow \frac{1}{||x||} \leq \frac{\sigma_n(A)}{||b||}$$

Lemma 1 follows from these two inequalities.

Lemma 2. *If $Ax = b$ and $(A + \delta A)(x + \delta x) = b$ then*

$$\frac{||\delta x||}{||x + \delta x||} \leq \kappa(A)\frac{||\delta A||}{||A||}$$

Exercise: Prove Lemma 2.

In regards to the condition number, sometimes people state things like:

For any function f, the condition number of f at x is defined as:

$$\lim_{\delta \to 0} \sup_{||\delta x|| < \delta} \frac{||f(x) - f(x + \delta x)||}{||\delta x||}$$

If f is differentiable, this is equivalent to the Jacobian of f: $||J(f)||$. A result of Demmel's is that condition numbers are related to a problem being "ill-posed". A problem $Ax = b$ is ill-posed if the condition number $\kappa(A) = \infty$, which occurs iff $\sigma_1(A) = 0$. Letting $V := \{A : \sigma_1(A) = 0\}$, we state the following Lemma:

Lemma 3. *$\sigma_1(A) = dist(A, V)$, i.e. the Euclidian distance from A to the set V.*

Proof to Lemma 3: Consider the singular value decomposition (SVD), $A = USV^T$, U, V orthogonal. S is defined as the diagonal matrix composed of singular values, $\sigma_1, \ldots, \sigma_n$.

Construct a matrix B to be the singular matrix closest to A. Then $A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$ and $B = \sum_{i=2}^{n} \sigma_i u_i v_i^T$. Now consider the Frobenius norm, denoted $||M||_F$ of A and B: $||A - B||_F = ||\sigma_1 u_1 v_1^T||_F = \sigma_1$. Since $\sigma_1(B) = 0$ and B, $dist(A, V) \leq \sigma_1(A)$.

The following claim will help us prove that $dist(A, V) \geq \sigma_1(A)$. For a singular matrix B and let $\delta A = A - B$. The following claim implies that $||A - B|| \geq \sigma_1$, and Lemma 5 implies that $||A - B||_F \geq ||A - B||$,

Claim 4. *If $(A + \delta A)$ is singular, then $||\delta A||_F \geq \sigma_1$*

Proof: $\exists v, ||v|| = 1, s.t. (A + \delta A)v = 0,$

$||Av|| \geq \sigma_1(A) \Rightarrow ||\delta Av|| \geq \sigma_1 \Rightarrow \sigma_n(\delta A) \geq \sigma_1.$

by the next lemma, we

Lemma 5. $||A||_F \geq \sigma_n(A)$

Proof: The Froebinus norm, which is the root of the sum of the squares of the entries in a matrix, does not change under a change of basis. That is, if V is an orthonormal matrix, then: $||AV||_F = ||A||_F$. In particular, if $A = USV$ is the singular-value decomposition of A, then

$$||A||_F = ||USV||_F = ||S||_F = \sqrt{\sum \sigma_i^2}.$$

We now state the main theorem that will be proved in this and the next lectures.

Theorem 6. *Let A be a d-by-d matrix such that $\forall i, j, |a_{ij}| \leq 1$. Let G be a d-by-d with Gaussian random variance $\sigma^2 \leq 1$. We will start to prove the following claims:*

a. $Pr[\sigma_1(A+G) \leq \epsilon] \leq \sqrt{\frac{2}{\pi}} \frac{d^{3/2}\epsilon}{\sigma}$

b. $Pr[\kappa(A) > d^2(1 + \sigma\frac{\sqrt{\log 1/\epsilon}}{\epsilon\sigma})] \leq 2\epsilon$

To give a geometric characterization of what it means for σ_1 to be small. Let $a_1, \ldots, a_d$ be the columns of A. Each a_i is a $d-$element vector. We now define $height(a_1, \ldots, a_d)$ as the shortest distance from some a_i to the span of the remaining vectors:

$$height(a_1, \ldots, a_d) = \min_i dist(a_i, span(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_d))$$

Lemma 7. $height(a_1, \ldots, a_d) \leq \sqrt{d}\sigma_1(A)$

Proof: Let v be a vector such that $||v|| = 1$, $||Av|| = \sigma_1(A) = ||\sum_{i=1}^d a_i v_i||$ Since v is a unit vector, some coordinate $|v_i| \geq \frac{1}{\sqrt{d}}$. Assume it is v_1. Then:

$$||\sum_{i=2}^d a_i \frac{v_i}{v_1} + a_1|| = \frac{\sigma_1}{v_1} \leq \sqrt{d}\sigma_1(A) \Rightarrow dist(a_1, span(a_2, \ldots, a_d)) \leq \sigma_1(A)\sqrt{d}$$

Lemma 8. $Pr[height(a_1 + g_1, \ldots, a_n + g_n) \leq \epsilon] \leq \frac{d\epsilon}{\sigma}$

This lemma follows from the union bound applied to the following Lemma:

Lemma 9. *$Pr[dist(a_1 + g_1, span(a_2 + g_2, \ldots, a_d + g_d)) \leq \epsilon] \leq \frac{\epsilon}{\sigma}$*

Proof of Lemma 9: This proof will take advantage of the following lemmas regarding gaussian distributions:

Lemma 10. *A a Gaussian distribution g has density:*

$$(\frac{1}{\sqrt{2\pi}\sigma})^d \cdot e^{\frac{||g||^2}{2\sigma^2}}$$

Lemma 11. *A univariate Gaussian x with mean x_0 and standard deviation σ has density:*

$$\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-x_0)^2}{2\sigma^2}}$$

Lemma 12. *The Gaussian distribution is spherically symmetric. That is, it is invariant under orthogonal changes of basis.*

Exercise: Prove Lemma 12.

Returning to the proof, fix $a_2, \ldots, a_d$ and $g_2, \ldots, g_d$. Let $S = span(a_2 + g_2, \ldots, a_d + g_d)$. We want to upperbound the distance of the vector $a_1 + g_1$ to the multi-dimensional plane S, which has $dim(S) = d - 1$. Since the vector is of higher dimension, the distance to the span will be bounded by one element. We can then just select x to be a univariate Gaussian random variable such that $x = g_{11}$ and $x_0 = a_{11}$. Using Lemma 11 and the fact that $e^{\frac{-g^2}{2\sigma^2}} \leq 1$, we can prove lemma 12:

$$Pr[|g_{11} - a_{11}| < \epsilon] = \int_{a_{11}-\epsilon}^{a_{11}+\epsilon} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-g_{11}^2}{2\sigma^2}} \leq \frac{2\epsilon}{\sqrt{2\pi}\sigma} = \sqrt{\frac{2}{\pi}}\frac{\epsilon}{\sigma} \leq \frac{\epsilon}{\sigma}$$

Part (a) of Theorem 6 follows from these lemmas and claims.