DS-GA 1014: Extended Syllabus Lecture 8

Optimization and Computational Linear Algebra for Data Science (Fall 2018)

Brett Bernstein
brettb@cims.nyu.edu
http://www.cims.nyu.edu/~brettb

October 30, 2018

These are not meant to be Lecture Notes. They are simply extended syllabi with the most important definitions and results from the lecture. As such, they lack the intuition and motivation and so they are not a good place to learn the material the first time, just to briefly review it. These extended syllabi will also have references.

There are many amazing books about linear algebra and virtually all of them will contain the material for this particular lecture, examples include the book suggested for the course [2]. Another place you can read about some of these is the Lecture Notes for DSGA1002 [1].

Please let me know if you find any typos

- Given a matrix $A \in \mathbb{R}^{m\times n}$ its *Frobenius norm* $\|A\|_F$ is given by

$$\|A\|_F = \sqrt{\operatorname{Tr}(A^TA)} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2},$$

  where $\sigma_1, \ldots, \sigma_{\min(m,n)}$ are the singular values of A.

- Given a matrix $A \in \mathbb{R}^{m\times n}$ its *spectral norm* $\|A\|$ (also written $\|A\|_s$ or $\|A\|_2$) is given by

$$\|A\| = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\| = \sigma_1,$$

  where σ_1 is the largest singular value of A.

- The following properties hold. [The last two look like Cauchy-Schwarz, but they are different.]

  1. $\|A\| \le \|A\|_F$.
  2. $\|Ax\| \le \|A\|\|x\|$ for all $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$.
  3. $\|AB\| \le \|A\|\|B\|$ for all $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$.

- Let $A \in \mathbb{R}^{n\times n}$ be invertible, and suppose that $Ax = b$ for some $x, b \in \mathbb{R}^n$. Then the solution $\tilde{x}$ to the perturbed linear system

$$A\tilde{x} = b + \epsilon,$$

for $\epsilon \in \mathbb{R}^n$ satisfies

$$\|\tilde{x} - x\| \leq \|A^{-1}\|\|\epsilon\| = \frac{\|\epsilon\|}{\sigma_n},$$

where σ_n is the smallest singular value of A. If $x, b \neq 0$ we also have

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \kappa(A)\frac{\|\epsilon\|}{\|b\|}$$

where the *condition number* $\kappa(A)$ is defined by

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

with σ_1 denoting the largest singular value of A.

- Numerical linear algebra libraries, like numpy, use the SVD to compute the rank of a matrix by counting all singular values that are larger than some threshold.
- If $A \in \mathbb{R}^{n\times n}$ has spectral decomposition $V\Lambda V^T$ then we have

$$A^k = V\Lambda^k V^T$$

for all $k \geq 0$. More generally, if $p(x) = \sum_{k=0}^n a_k x^k$ is a polynomial, then we have

$$p(A) = Vp(\Lambda)V^T$$

where

$$p(\Lambda) = \begin{bmatrix} p(\lambda_1) & & & \\ & p(\lambda_2) & & \\ & & \ddots & \\ & & & p(\lambda_n) \end{bmatrix}.$$

- If $A \in \mathbb{R}^{n\times n}$ can be written as $A = V\Lambda V^T = UDU^T$ where U, V are orthogonal, and D, Λ are diagonal, then $D = \Lambda$ up to reordering of the columns. In other words, there are exactly n eigenvalues (including repeats) that are uniquely determined by the matrix A.
- Similarly, if $B \in \mathbb{R}^{m\times n}$ can be written as $B = U\Sigma V^T = PDQ^T$ where U, V, P, Q are orthogonal and $\Sigma, D \in \mathbb{R}^{m\times n}$ are diagonal with non-negative entries, then $\Sigma = D$ up to reordering.
- We can compute other functions of a symmetric matrix $A = V\Lambda V^T$ as well:
  1. $A^{1/2} = V\Lambda^{1/2}V^T$ if $A \succeq 0$ where $\Lambda^{1/2}$ is obtained by taking the square roots of the entries of Λ.
  2. $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k = V\left(\sum_{k=0}^{\infty}\Lambda^k\right)V^T$ if $\max_i |\lambda_i| < 1$.
  3. $e^A = \sum_{k=0}^{\infty}\frac{1}{k!}A^k$ (used to solve systems of differential equations).
  4. $A^{-1} = \sum_{k=0}^{\infty}(I - A)^k$ if the maximum absolute eigenvalue of $I - A$ is strictly less than 1.

- If $A \in \mathbb{R}^{n\times n}$ is invertible we can solve $Ax = b$ by writing

$$x = A^{-1}b = \sum_{k=0}^{\infty}(I-A)^k b$$

  if the largest absolute eigenvalue of $I - A$ is strictly less than 1. This gives us the following iterative method for solving the linear system:

  1. Let $x^{(0)} = b$.
  2. For $k = 1, 2, \ldots$
     (a) Set $x^{(k)} = (I - A)x^{(k-1)} + b$.
     (b) Stop if $\|Ax^{(k)} - b\|$ is sufficiently small.

  Then we have

$$x^{(k)} = \sum_{j=0}^{k}(I-A)^k b \to \sum_{j=0}^{\infty}(I-A)^k b = A^{-1}b.$$

- Given an invertible matrix $C \in \mathbb{R}^{n\times n}$ we can instead solve the equivalent system $CAx = Cb$. Our iterative method converges for the new system if the largest absolute eigenvalue of $I - CA$ is strictly smaller than 1. The goal is to find an easy to compute C that makes the eigenvalues of $I - CA$ small. Such a C is called a preconditioner.

References

[1] Carlos Fernandez-Granda, *Lecture Notes of DSGA1002*, available at `http://www.cims.nyu.edu/~cfgranda/pages/DSGA1002_fall15/notes.html`, 2015

[2] Gilbert Strang, *Introduction to Linear Algebra*, Fifth Edition, 2016