

Mathematical background and pre work

Mathematical background

We will not assume a lot of mathematical background in this course but will use some basic notions from **linear algebra**, such as *vector spaces* (finite dimensional and almost always over the real numbers), matrices, and associated notions such as rank, eigenvalues and eigenvectors. We will use the notion of *convexity* (of functions and sets) and some of its basic properties. We will also use basic notions from probability such as random variables, expectation, variance, tail bounds as well as properties of the normal (a.k.a. Gaussian) distribution. Though this will not be our main focus, we will assume some comfort with algorithms and notions such as order of growth ($O(n)$, $2^{\Omega(n)}$, etc..) and some notions from computational complexity such as the notion of a reduction and the classes **P** and **NP**.

Probably the most important mathematical background for this course is that ever elusive notion of “mathematical maturity” which basically means the ability to pick up on the needed notions as we go along. At any point, please do not hesitate to ask questions when you need clarifications or pointers to some references, either in the class or on the Piazza forum.

Some references for some of this material (that include much more than what we need are):

- All these topics are covered to some extent in Ryan O’Donnell’s CMU class [15-859T: A Theorist’s Toolkit](#) see in particular Lectures 6-8 (spectral graph theory) and Lectures 13-14 (linear programming). See also the lecture notes for Jonathan Kelner’s MIT course [18.409 Topics in Theoretical Comp Sci](#). While not strictly necessary, you may find Luca Trevisan [series of blog posts on expanders](#) (from 2006, 2008, and 2011) illuminating.
- We will sometimes touch upon Fourier analysis of Boolean functions which is covered by [O’Donnell’s excellent book and lecture notes](#)
- For basic linear algebra and probability, see the [lecture notes by Papadimitriou and Vazirani](#), the [lecture notes of Lehman, Leighton and Meyer](#) from MIT Course 6.042 “Mathematics For Computer Science” (Chapters 1-2 and 14 to 19 are particularly relevant). The “Probabilistic Method” book by Alon and Spencer is a great resource for discrete probability. Also, the books of Mitzenmacher

and Upfal and Prabhakar and Raghavan cover probability from a more algorithmic perspective.

- Convexity, linear programming duality: see [Boyd and Parrilo's lecture notes](#) and in particular Lectures 1-5. The book [Convex Optimization by Boyd and Vandenberghe](#), which is available online, is an excellent resource for this area, which includes much more than what we will use here.

Pre work ("homework o")

Please do the following reading and exercises before the first lecture.

Reading:

Please read the lecture notes for [the introduction to this course](#) and for [definitions of sum of squares over the hypercube](#). You don't have to do the exercises in the lecture notes, but you may find attempting them useful. (See [here](#) for all notation used in these lecture notes.)

Exercises:

You do not need to submit these exercises, or even to write them down properly, and feel free to collaborate with others while working on them.

All matrices and vectors are over the reals. In all the exercises below you can use the fact that any $n \times n$ matrix A has a singular value decomposition (SVD) $A = \sum_{i=1}^r \sigma_i u_i \otimes v_i$ with $\sigma_i \in \mathbb{R}$ and $u_i, v_i \in \mathbb{R}^n$, and for every i, j $\|u_i\| = 1$, $\|v_j\| = 1$ (where $\|v\| = \sqrt{\sum v_i^2}$), and for all $i \neq j$, $\langle u_i, u_j \rangle = 0$ and $\langle v_i, v_j \rangle = 0$. (For vectors u, v , their tensor product is defined as $u \otimes v$ is the matrix $T = uv^\top$ where $T_{i,j} = u_i v_j$.) Equivalently $A = U \Sigma V^\top$ where Σ is a diagonal matrix and U and V are orthogonal matrices (satisfying $U^\top U = V^\top V = I$). If A is symmetric then there is such a decomposition with $u_i = v_i$ for all i (i.e., $U = V$). In this case the values $\sigma_1, \dots, \sigma_r$ are known as *eigenvalues* of A and the vectors v_1, \dots, v_r are known as *eigenvectors*. (This decomposition is unique if $r = n$ and all the σ_i 's are distinct.) Moreover the SVD of A can be found in polynomial time. (You can ignore issues of numerical accuracy in all exercises.)

1. Exercise. For an $n \times n$ matrix A , the *spectral norm* of A is defined as the maximum of $\|Av\|$ over all vectors $v \in \mathbb{R}^n$ with $\|v\| = 1$. * Prove that if A is symmetric (i.e., $A = A^\top$), then $\|A\| \leq \max_i \sum_j |A_{i,j}|$. See footnote for hint¹ * Show that if A is the adjacency matrix of a d -regular graph then $\|A\| = d$.

¹ You can do this via the following stronger inequality: for any (not necessarily symmetric) matrix A , $\|A\| \leq \sqrt{\alpha\beta}$ where $\alpha = \max_i \sum_j |A_{i,j}|$ and $\beta = \max_j \sum_i |A_{i,j}|$.

2. Exercise. Let A be a symmetric $n \times n$ matrix. The *Frobenius norm* of A , denoted by $\|A\|_F$, is defined as $\sqrt{\sum_{i,j} A_{i,j}^2}$.

3. Exercise. Prove that $\|A\| \leq \|A\|_F \leq \sqrt{n}\|A\|$. Give examples where each of those inequalities is tight.

4. Exercise. Let $\text{Tr}(A) = \sum A_{i,i}$. Prove that for every even k , $\|A\| \leq \text{Tr}(A^k)^{1/k} \leq n^{1/k}\|A\|$.

5. Exercise (Spectral norm of a random matrix). Let A be a symmetric matrix such that $A_{i,i} = 0$ for all i and $A_{i,j}$ is chosen to be a random value in $\{\pm 1\}$ independently of all others.

- Prove that (for n sufficiently large) with probability at least 0.99, $\|A\| \leq n^{0.9}$.
- (harder) Prove that with probability at least 0.99, $\|A\| \leq n^{0.51}$.

While $\|A\|$ can be computed in polynomial time, both $\max_i \sum_j |A_{i,j}|$ and $\|A\|_F$ give even simpler to compute upper bounds for $\|A\|$. However the examples in the previous exercise show that they are not always tight. It is often easier to compute $\text{Tr}(A^k)^{1/k}$ than trying to compute $\|A\|$ directly, and as k grows this yields a better and better estimate.

6. Exercise. Let A be an $n \times n$ symmetric matrix. Prove that the following are equivalent:

1. A is positive semi-definite. That is, for every vector $v \in \mathbb{R}^n$, $v^\top Av \geq 0$ (where we think of vectors as column vectors and so $v^\top Av = \sum_{i,j} A_{i,j} v_i v_j$).
2. All eigenvalues of A are non-negative. That is, if $Av = \lambda v$ then $\lambda \geq 0$.
3. The quadratic polynomial P_A defined as $P_A(x) = \sum A_{i,j} x_i x_j$ is a *sum of squares*. That is, there are linear functions L_1, \dots, L_m such that $P_A = \sum_i (L_i)^2$.
4. $A = B^\top B$ for some $n \times r$ matrix B
5. There exist a set of correlated random variables (X_1, \dots, X_m) such that for every i, j , $\mathbb{E} X_i X_j = A_{i,j}$ and moreover, for every i , the random variable X_i is distributed like a Normal variable with mean 0 and variance $A_{i,i}$.

References