CSE 521: Design and Analysis of Algorithms I **Fall 2020**

Background / Cheat Sheet

In this note I will discuss several background materials that we will discuss and exploit many times throughout this course.

1 Randomized Algorithm

Expectation: For a random variable X with domain, the discrete set S,

$$\mathbb{E}[X] = \sum_{s \in S} \mathbb{P}[X = s]\, s.$$

Linearity of Expectation: For any two Random variables X, Y,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Variance: The variance of a random variable X is defined as $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$. The following identity always holds,

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

The standard deviation of X, $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Mutual Independence A set of random variables $X_1, \ldots, X_n$ are mutually independent if for any $S \subseteq \{1, \ldots, n\}$,

$$\mathbb{E}\left[\prod_{i \in S} X_i\right] = \prod_{i \in S} \mathbb{E}[X_i].$$

k-wise Independence For an integer $k \geq 2$, a set of random variables $X_1, \ldots, X_n$ is set to be k-wise independent if for any set $S \subseteq \{1, \ldots, n\}$ of size k,

$$\mathbb{E}\left[\prod_{i \in S} X_i\right] = \prod_{i \in S} \mathbb{E}[X_i].$$

Sum of Variance: Let $X_1, \ldots, X_n$ be *pairwise independent* random variables, then

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Markov's Inequality Let X be a nonnegative random variable, then for any $k \geq 0$,

$$\mathbb{P}[X \geq k] \leq \frac{\mathbb{E}[X]}{k}.$$

Chebyshev's Inequality For any random variable X and any $\epsilon > 0$,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| > \epsilon\right] \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

So, equivalently,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| > k\sigma(X)\right] \leq \frac{1}{k^2}.$$

Hoeffding's Inequality Let $X_1, \ldots, X_n$ be independent random variables where for all i, $X_i \in [a_i, b_i]$. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i - \mathbb{E}\sum_{i=1}^n X_i\right| > \epsilon\right] \leq 2\exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n (a_i - b_i)^2}\right)$$

Multiplicative Chernoff Bound Let $X_1, \ldots, X_n$ be independent Bernoulli random variables, i.e., for all i, $X_i \in \{0, 1\}$, and let $X = X_1 + \cdots + X_n$ and $\mu = \mathbb{E}[X]$. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left[X > (1+\epsilon)\mu\right] \leq \left(\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right)^\mu \leq e^{-\frac{\epsilon^2\mu}{2+\epsilon}},$$

and

$$\mathbb{P}\left[X < (1-\epsilon)\mu\right] \leq e^{-\epsilon^2\mu/2}$$

McDiarmid's Inequality Let $X_1, \ldots, X_n \in \mathcal{X}$ be independent random variables. Let $f : \mathcal{X}^n \to \mathbb{R}$. If for all $1 \leq i \leq n$ and for all $x_1, \ldots, x_n$ and $\tilde{x}_i$,

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}, \ldots, x_n)| \leq c_i,$$

then,

$$\mathbb{P}\left[|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| > \epsilon\right] \leq 2\exp\left(-\frac{-2\epsilon^2}{\sum_i c_i^2}\right).$$

Concentration of Gaussians Let $X_1, \ldots, X_n$ be independent standard normal random variables i.e., for all i, $X_i \sim \mathcal{N}(0, 1)$. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i^2 - n\right| > \epsilon\right] \leq 2\exp\left(\frac{\epsilon^2}{8}\right)$$

Gaussian Density Function The density function of a 1-dimensional normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is as follows:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

More generally, we say $X_1, \ldots, X_n$ form a multivariate normal random variable when they have following density function:

$$\det(2\pi\Sigma)^{-1/2} e^{-(x-\mu)^\intercal \Sigma^{-1}(x-\mu)/2}$$

where Σ is the covariance matrix of $X_1, \ldots, X_n$. In particular, for all i, j,

$$\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}\left[X_i - \mathbb{E}[X_i]\right]\mathbb{E}\left[X_j - \mathbb{E}[X_j]\right] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j].$$

As a special case, if $X_1, \ldots, X_n$ are standard normals chosen independently then Σ is just the identity matrix.

2 Spectral Algorithms

Determinant Let $A \in \mathbb{R}^{n\times n}$, the determinant of A can be written as follows:

$$\det(A) = \sum_{\sigma} \prod_{i=1}^{n} A_{i,\sigma(i)} \operatorname{sgn}(\sigma).$$

where the sum is over all permutations σ of the numbers $1, \dots, n$, and $\operatorname{sgn}(\sigma) \in \{+1, -1\}$. For a permutation σ, $\operatorname{sgn}(\sigma)$ is the parity of the number of swaps one needs to transform σ into the identity permutations. For example, for $n = 4$, $\operatorname{sgn}(1, 2, 3, 4) = +1$ because we need no swaps, $\operatorname{sgn}(2, 1, 3, 4) = -1$ because we can transform it to the identity just by swapping $1, 2$ and $\operatorname{sgn}(3, 1, 2, 4) = +1$.

Properties of Determinant

- For a matrix $A \in \mathbb{R}^{n\times n}$, $\det(A) \neq 0$ if and only if the columns of A are linearly independent. Recall that for a set of vectors $v_1, \dots, v_n \in \mathbb{R}^n$, we say they are linearly independent if for any set of coefficients $c_1, \dots, c_n$

  $$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0$$

  only when $c_1 = c_2 = \dots = c_n = 0$. In other words, $v_1, \dots, v_n$ are linearly independent if no v_i can be written as a linear combination of the rest of the vectors.

- For any matrix $A \in \mathbb{R}^{n\times n}$, with eigenvalues $\lambda_1, \dots, \lambda_n$,

  $$\det(A) = \prod_{i=1}^{n} \lambda_i$$

  So, $\det(A) = 0$ iff A has at least one zero eigenvalue. So, it follows from the previous fact that A has a zero eigenvalue iff columns of A are linearly independent.

- For any two square matrices $A, B \in \mathbb{R}^{n\times n}$,

  $$\det(AB) = \det(A)\det(B).$$

Characteristic Polynomial For a matrix $A \in \mathbb{R}^{n\times n}$ we write $\det(xI - A)$ for an indeterminant (variable) x is called the characteristic polynomial of A. The roots of this polynomial are the eigenvalues of A. In particular,

$$\det(xI - A) = (x - \lambda_1)(x - \lambda_2) \dots (x - \lambda_n),$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. It follows from the above identity that for $x = 0$, $\det(-A) = \prod_{i=1}^{-} \lambda_i$ or equivalently, $\det(A) = \prod_{i=1}^{n} \lambda_i$.

Rank The rank of a matrix $A \in \mathbb{R}^{n\times n}$ is the number of nonzero eigenvalues of A. More generally, the rank of a matrix $A \in \mathbb{R}^{m\times n}$ is the number of nonzero singular values of A. Or in other words, it is the number of nonzero eigenvalues of $AA^\intercal$.

PSD matrices We discuss several equivalent defnitions of PSD matrices. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefnite (PSD) iff

- All eigenvalues of A are nonnegative
- A can be written as $BB^\intercal$ for some matrix $B \in \mathbb{R}^{n\times m}$.
- $x^\intercal Ax \geq 0$ for all vectors $x \in \mathbb{R}^n$.
- $\det(A_{S,S}) \geq 0$ for all $S \subseteq \{1, \ldots, n\}$ where $A_{S,S}$ denotes the square submatrix of A with rows and columns indexed by S.

The following fact about PSD matrices is immediate. If $A \succeq 0$ is an $n \times n$ matrix, then for any matrix $C \in \mathbb{R}^{k\times n}$,

$$CAC^T \succeq 0.$$

This is because for any vector $x \in \mathbb{R}^k$,

$$x^T CAC^T x = (C^T x)^T A(C^T x) = y^T Ay \geq 0,$$

where $y = C^T x$.

For two symmetric $A, B \in \mathbb{R}^n$ we write $A \preceq B$ if and only if $B - A \succeq 0$. In other words, $A \preceq B$ if and only if for any vector $x \in \mathbb{R}^n$,

$$x^T Ax \leq x^T Bx.$$

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ be the eigenvalues of B. If $A \preceq B$, then for all i, $\lambda_i \leq \tilde{\lambda}_i$.

Nonsymmetric Matrices Any matrix $A \in \mathbb{R}^{m\times n}$ (for $m \leq n$) can be written as

$$A = \sum_{i=1}^{m} \sigma_i u_i v_i^\intercal$$

where

- $u_1, \ldots, u_m \in \mathbb{R}^m$ form an orthonormal set of vectors. These are called left singular vectors of A and they have the property, $u_i A = \sigma_i v_i$. These vectors are the eigenvectors of the matrix $AA^\intercal$.
- $v_1, \ldots, v_m \in \mathbb{R}^n$ form an orthonormal set of vectors. Note that these vectors do not necessarily span the space. These vectors are eigenvectors of the matrix $A^\intercal A$.
- $\sigma_1, \ldots, \sigma_m$ are called the singular values of A. They are always real an nonnegative. In fact they are eigenvalues of the PSD matrix $AA^\intercal$.

Rotation Matrix A matrix $R^{n\times n}$ is a rotation matrix iff $\|Rx\|_2 = \|x\|_2$ for all vectors $x \in \mathbb{R}^n$. In other words, R as an *operator* preserves the norm of all vectors. Next, we discuss equivalent definitions of R being a rotation matrix. R is a rotation matrix iff

- $RR^\intercal = I$.
- All singular values of R are 1.
- Columns of R form an orthonormal set of vectors in $\mathbb{R}^n$.

Projection Matrix A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is a projection matrix iff

- It can be written as $P = \sum_{i=1}^k v_i v_i^\intercal$ for some $1 \leq k \leq n$.
- All eigenvalues of P are 0 or 1.
- $PP = P$.

It follows from the spectral theorem that there is a unique projection matrix of rank n and that is the identity matrix. In general a projection matrix *projects* any given vector x to the linear subspace corresponding to span of the vectors $v_1, \ldots, v_k$.

Trace For a square matrix $A \in \mathbb{R}^{n\times n}$ we write

$$\mathrm{Tr}(A) = \sum_{i=1}^n A_{i,i}$$

to denote the sum of entries on the diagonal of A. Next, we discuss several properties of the trace.

- Trace of A is equal to the sum of all eigenvalues of A.
- Trace is a linear operator, for any two square matrices $A, B \in \mathbb{R}^{n\times n}$

$$\begin{aligned} \mathrm{Tr}(A+B) &= \mathrm{Tr}(A) + \mathrm{Tr}(B) \\ \mathrm{Tr}(tA) &= t\,\mathrm{Tr}(A), \forall t \in \mathbb{R}. \end{aligned}$$

- It follows by the previous fact that for a random matrix X, $\mathbb{E}[\mathrm{Tr}(X)] = \mathrm{Tr}(\mathbb{E}[X])$.
- For any pair of matrices $A \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{k\times n}$ such that AB is a square matrix we have

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

  So, in particular, for any vector $v \in \mathbb{R}^n$,

$$\mathrm{Tr}(vv^\intercal) = \mathrm{Tr}(v^\intercal v) = \|v\|^2.$$

- For any matrix $A \in \mathbb{R}^{m\times n}$

$$\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 = \mathrm{Tr}(AA^\intercal).$$

Matrix Chernoff Bound Let X be a random $n \times n$ PSD matrix. Suppose that $X \preceq \alpha\mathbb{E}[X]$with probability 1 for some $\alpha \geq 0$. Let $X_1, \ldots, X_k$ be independent copies of X. Then, for any $0 < \epsilon < 1$,

$$\mathbb{P}\left[(1-\epsilon)\mathbb{E}[X] \preceq \frac{1}{k}(X_1 + \cdots + X_k) \preceq (1+\epsilon)\mathbb{E}[X]\right] \geq 1 - 2ne^{-\epsilon^2 k/4\alpha}.$$

3 Optimization

Convex Functions A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex on a set $S \subseteq \mathbb{R}^n$ if for any two points $x, y \in S$, we have

$$f\left(\frac{x+y}{2}\right) \le \frac{1}{2}(f(x) + f(y)).$$

We say f is concave if for any such $x, y \in S$, we have

$$f\left(f(\alpha x + (1-\alpha)y\right) \ge \alpha f(x) + (1-\alpha) f(y),$$

for any $0 \le \alpha \le 1$. There is an equivalent definition of convexity: For a function $f : \mathbb{R}^n \to \mathbb{R}$, the Hessian of f, $\nabla^2 f$ is a $n \times n$ matrix defined as follows:

$$(\nabla^2 f)_{i,j} = \partial_{x_i} \partial_{x_j} f$$

for all $1 \le i, j \le n$. We can show that f is convex over S if and only if for all $a \in S$,

$$\nabla^2 f\Big|_{x=a} \succeq 0.$$

For example, consider the function $f(x) = x^T A x$ for $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$. Then, $\nabla^2 f = A$. So, f is convex (over $\mathbb{R}^n$) if and only if $A \succeq 0$.

For another example, let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^k$ for some integer $k \ge 2$. Then, $f''(x) = k(k-1)x^{k-2}$. If k is an even integer, $f''(x) \ge 0$ over all $x \in \mathbb{R}$, so f is convex over all real numbers. On the other hand, if k is an odd integer then $f''(x) \ge 0$ if and only if $x \ge 0$. So, in this f is convex only over non-negative reals.

Similarly, f is concave over S, if $\nabla^2 f\Big|_{x=a} \preceq 0$ for all $a \in S$. For example, $x \mapsto \log x$ is concave over all positive reals.

Convex set We say a set $S \subseteq \mathbb{R}^n$ is convex if for any pair of points $x, y \in S$, the line segment connecting x to y is in S.

For example, let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function over a set $S \subseteq \mathbb{R}^n$. Let $t \in \mathbb{R}$, and define

$$T = \{x \in \mathbb{R}^n : f(x) \le t\}.$$

Then, T is convex. This is because if $x, y \in T$, then for any $0 \le \alpha \le 1$,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y) \le \alpha t + (1-\alpha)t = t$$

where the first inequality follows by convexity of f. So, $\alpha x + (1-\alpha) \in T$ and T is convex.

Norms are Convex functions A norm $\|\cdot\|$ is defined as a function that maps $\mathbb{R}^n$ to $\mathbb{R}$ and satisfies the following three properties,

i) $\|x\| \ge 0$ for all $x \in \mathbb{R}^n$,

ii) $\|\alpha x\| = \alpha \|x\|$ for all $\alpha \ge 0$ and $x \in \mathbb{R}^n$,

iii) Triangle inequality: $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

It is easy to see that any norm function is a convex function: This is because for any $x, y \in R^n$, and $0 \le \alpha \le 1$,

$$\|\alpha x + (1-\alpha)y\| \le \|\alpha x\| + \|(1-\alpha)y\| = \alpha\|x\| + (1-\alpha)\|y\|.$$

4 Useful Inequalities

- For real numbers, $a_1, \ldots, a_n$ and nonnegative reals $b_1, \ldots, b_n$,

$$\min_i \frac{a_i}{b_i} \le \frac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n} \le \max_i \frac{a_i}{b_i}$$

- **Cauchy-Schwartz inequality:** For real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$,

$$\sum_{i=1}^{n} a_i \cdot b_i \le \sqrt{\sum_i a_i^2} \cdot \sqrt{\sum_i b_i^2}$$

  There is an equivalent vector-version of the above inequality. For any two vectors $u, v \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} u_i \cdot v_i = \langle u, v \rangle \le \|u\| \cdot \|v\|$$

  The equality in the above holds only when u, v are parallel.

- **AM-GM inequality:** For any n nonnegative real numbers $a_1, \ldots, \ldots, a_n$,

$$\frac{a_1 + \cdots + a_n}{n} \ge (a_1 \cdot a_2 \cdot \ldots a_n)^{1/n}.$$

- Relation between norms: For any vector $a \in \mathbb{R}^n$,

$$\|a\|_2 \le \|a\|_1 \le \sqrt{n} \cdot \|a\|_2$$

  The right inequality is just Cauchy-Schwartz inequality.

- For any real numbers $a_1, \ldots, a_n$,

$$(|a_1| + \cdots + |a_n|)^2 \le n(a_1^2 + \cdots + a_n^2).$$

  This is indeed a special case of Cauchy-Schwartz inequality.

- For any real number x, $1 - x \le e^{-x}$. In this course we use $1 - x \approx e^{-x}$ to simplify calculations.