Exercises

1. **Lots of Flows.** Suppose you wanted to find an approximate solution to the following “multi-commodity” flow problem: given a digraph \( G = (V, E) \) with unit arc capacities, send \( F_i \) flow from node \( s_i \) to node \( t_i \) in the graph, for all \( i \in [k] \). You should imagine that the flow from \( s_i \) to \( t_i \) is of commodity \( i \) (e.g., oil, water, sand...) which are all distinct.

   (a) Suppose \( \mathcal{P}_i \) is the set of all paths from \( s_i \) to \( t_i \): show that the following LP captures the problem we are trying to solve. The variables are \( f_P \), one for each path in \( \cup_i \mathcal{P}_i \).

   \[
   \sum_{P \in \mathcal{P}_i} f_P = F_i \quad \forall i \in [k]
   \]

   \[
   \sum_i \sum_{P \in \mathcal{P}_i; e \in P} f_P \leq 1 \quad \forall e \in E
   \]

   \[
   f \geq 0
   \]

   (b) Define an appropriate “easy” polytope \( K \) for this problem.

   (c) Given weights \( q \in \Delta_m \), how would you solve the oracle for this problem? Show you can find a flow that satisfies the demands, but uses at most \((1 + \varepsilon)\) capacity on each edge, in time \( O(k(m + n \log n) / \text{poly}(\varepsilon)) \).

2. **Strength in Convexity.** A function \( f : \mathbb{R}^n \to \mathbb{R} \) is called \( \ell \)-strongly-convex if

   \[
   f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\ell}{2} \|y - x\|^2.
   \]

   I.e., if the function is not just convex, but “locally it grows at least as fast as a quadratic”. Modify the basic gradient descent analysis to show that using the same update rule \( x_{t+1} \leftarrow x_t - \eta_t \nabla f(x_t) \) with suitably chosen \( \eta_t \), then we can find \( \hat{x} \in \mathbb{R}^n \) such that

   \[
   f(\hat{x}) - f(x^*) \leq O\left(\frac{G^2 \log T}{\ell \cdot T}\right)
   \]

   Again, assume that \( \|\nabla f(x)\| \leq G \). Note due to the assumption of strong convexity, we got better convergence (the dependence on \( T \) is better, there is no dependence on \( D = \|x_0 - x^*\| \)). Show that this analysis also works in the online case, if each function is strongly convex. (As a bonus problem, remove the \( \log T \) term in the numerator in the offline case. Why does this not extend to the online case?)

3. **Divergent Views.** Given two discrete probability distributions \( p, q \) defined over a universe of \( N \) elements, the Kullback-Liebler divergence between the two is defined as

   \[
   KL(p\|q) := \sum_{i=1}^N p_i \log_2 \frac{p_i}{q_i}.
   \]

   Show the following properties:
4. That's the Norm. In lecture, we defined a differentiable convex function $f: K \rightarrow \mathbb{R}$ to be $L$-Lipschitz with respect to norm $\| \cdot \|$ if

$$\frac{|f(x) - f(y)|}{\|x - y\|} \leq L.$$

Show that this is equivalent to $\|\nabla f(x)\|_* \leq L$ for all $x \in K$.

Similarly, show that $f$ being $\alpha$-strongly-convex with respect to $\| \cdot \|$ is equivalent to $\|\nabla f(x) - \nabla f(y)\|_* \geq \alpha \|x - y\|$. And that $f$ being $\beta$-smooth with respect to $\| \cdot \|$ is equivalent to $\|\nabla f(x) - \nabla f(y)\|_* \leq \beta \|x - y\|$.

5. LPs for Arborescences. Consider the LP:

$$\min \quad c^T x$$
$$x(\partial(v)) = 1 \quad \forall v \in V,$$
$$x(\partial(S)) \geq 1 \quad \forall S \text{ s.t. } r \not\in S, S \neq \emptyset,$$
$$x \geq 0.$$

Using the ideas from Lectures #2 and #8 show that all vertices of this LP are integral, and correspond to all the arborescences of $G$. Moreover, give a separation oracle for the exponentially many contraints.

6. These are “Small” Numbers. For an integer $k$, define $\langle k \rangle = 1 + \lceil \log_2(|k| + 1) \rceil$; for a rational $p/q$ (with $p, q$ coprime, $q > 0$), define $\langle p/q \rangle = \langle p \rangle + \langle q \rangle$; for a matrix $R = (r_{ij})$ of rationals, define $\langle M \rangle = \sum_{i,j} \langle r_{ij} \rangle$. Let $\det(R)$ denote the determinant of $R$.

(a) If $R$ is an $n \times n$ matrix, show that $\langle \det(R) \rangle \leq \text{poly}(n, \text{size } R)$.

Now consider the LP $\min \{c^T x \mid Ax \geq b\}$, where $x \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

(b) If each of the numbers in $A$ and $b$ are rationals having size at most $S$, and if $x^*$ is a basic feasible solution, give an upper bound on the size of each entry of $x^*$. In particular, show that each entry $x^*_i$ is a rational number with size at most $K = \text{poly}(n \cdot S)$.

(c) If the LP has a finite optimum, and the numbers in $c$ have size at most $S$, infer that the optimal value of the LP has size $O((K + S)n)$.

Hint: Google for ‘Cramer’s rule’ and ‘Hadamard Inequality’.
Problems

Solve the first problem, and any three of the other four.

1. **Hmm, That’s Odd...** To solve the max-weight perfect matching problem, we need to optimize over the perfect matching polytope. In turn, this requires that we find a separation oracle for the odd cut constraints. I.e., given \( x \in \mathbb{R}^{|E|} \), we need to find a set \( S \subseteq V \) such that \( |S| \) is odd, and \( x(\partial(S)) := \sum_{i \in S, j \notin S} x_{ij} \). Then, comparing this min-odd-cut value to 1, we can find a violated constraint (if one exists). (Assume that \( |V| \) is even, else the LP has no feasible solution anyways.)

(a) A function \( f: 2^V \to \mathbb{R} \) is called **submodular** if for all \( A, B \subseteq V \), we have

\[
    f(A) + f(B) \geq f(A \cup B) + f(A \cap B).
\]

Show \( f(A) := x(\partial(A)) \) is submodular. Observe \( f \) is symmetric, i.e., \( f(A) = f(V \setminus A) \).

(b) If \((C, \bar{C})\) is the min-cut, i.e, if \( x(\partial(C)) \) is the least among all non-empty cuts, and \( |C| \) is even, then show that there exists a min-odd cut contained within \( C \), or within \( \bar{C} \).

(c) Give an algorithm to find a min-odd-cut in polynomial time.

2. **Zero-Sum Games using LP Duality.** Recall the zero-sum game setup: we’re given a matrix \( M \in \mathbb{R}^{m \times n} \); if the row player plays a strategy \( x \in \Delta_m \) and the column player plays strategy \( y \in \Delta_n \), the payoff to the row player is \( x^\top M y \).

If we define \( C(x) = \min_{y \in \Delta_n} x^\top M y \), and \( R(y) = \max_{x \in \Delta_m} x^\top M y \), the minimax theorem proves that (a) for all \( x, y \), \( C(x) \leq R(y) \), and moreover (b) there exist \( x^*, y^* \) such that \( C(x^*) = R(y^*) \).

(a) Show an LP to compute \( \max_x C(x) \), the optimal strategy for the row player. (Hint: be careful, the definition of \( C() \) has a min sitting in there, so you’re looking to find \( \max_x \min_y x^\top M y \), which certainly does not look like a linear program.)

(b) Show an LP to compute \( \min_y R(y) \), the optimal strategy for the column player.

(c) Show that you can, in fact, find LPs for both the above parts, such that the dual of the first LP is a solution to the second part.

(d) (Do not submit.) Use weak duality to infer the first part of the minimax theorem, and strong duality to infer the second part.

3. **Capacitated Max-Flow and Width Reduction.** Consider the directed \( s-t \) max-flow problem: in Lectures #12-13, we defined \( K = \{ f \mid f_P \geq 0, \sum_{P \in p} f_P = F \} \), with constraints

\[
    f_e/c_e \leq 1 \quad \forall e \in E,
\]

where define \( f_e := \sum_{P \in P} f_P \). We considered \( c_e = 1 \) in lecture; now we consider the general case. Given the weights \( p \in \Delta_m \) given by Hedge, the “average” constraint looks like

\[
    \sum_e p_e \cdot (f_e/c_e) = \sum_e f_e(p_e/c_e) \leq 1.
\]

(a) (Do not submit.) Suppose the oracle sends the entire \( F \) units of flow along a shortest path w.r.t. \( p_e/c_e \). Show that there are capacitated networks, and possible \( p \in \Delta_m \), where all \( F \) flow is routed along a path using the least-capacity edge. Hence the width of this oracle is at least \((F/c_{\min})\).
Note that $F$ may be as large as $\Omega(mc_{\text{max}})$, so with general capacities, this ratio could be $\gg m$. Now, we’ll investigate the idea of adding a little bit to the edge weights (e.g., setting $w_e := p_e + \varepsilon/m$), and how it reduces the width of the problem at the expense of giving slightly approximate solutions.

For the rest, assume $\varepsilon \leq 1/10$, say. You may also assume that the instance is feasible; i.e., there exists a flow $f^* \in K$ that satisfies all the edge capacities.

(b) Set the edge weights to be $w_e := p_e + \varepsilon/m$, compute the shortest path w.r.t. edge lengths $\frac{w_e}{c_e}$, and send all $F$ flow along it. If this shortest path is $P^*$, show

$$F \cdot \sum_{e \in P^*} \frac{p_e}{c_e} \leq \min_{f \in K} \sum_{e \in E} \frac{f_e}{c_e} w_e \leq 1 + \varepsilon.$$  

(c) Show that

$$\max_{e \in P^*} F/c_e \leq O\left(\frac{m \varepsilon}{\varepsilon^2}\right).$$

Hence the oracle width is $O(\frac{m}{\varepsilon^2})$.

(d) (Do not submit.) Using this oracle and the MW algorithm guarantee, give an $\tilde{O}(m^2/\varepsilon^3)$-time $(1 + \varepsilon)$-approximate max-flow algorithm for the capacitated case.

(e) Bonus: use these ideas to get an algorithm for the multi-commodity case from exercise #1 that works for capacitated graphs, but whose runtime does not depend on the magnitude of the capacities.

Note: This idea was used in the Christiano et al. paper, combined with the electrical flows, to bring the width down to $O\left(\sqrt{\frac{m}{\varepsilon}}\right)$.

4. Solving a Linear System. Given a positive-definite matrix $A \in \mathbb{R}^{n \times n}$ with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, the condition number for $A$ is $\kappa := \frac{\lambda_n}{\lambda_1}$. Given a vector $b \in \mathbb{R}^n$, the goal is to find a “near”-solution to the linear system $Ax = b$. Consider the function $f(x) = \frac{1}{2} x^\top Ax - bx$.

(a) (Do not submit) Show that $f$ is a convex function, with $\nabla f(x) = Ax - b$. Hence infer that the minimizer $x^*$ of $f(x)$ satisfies $Ax^* = b$. Moreover, show that $f$ is $\lambda_1$-strongly-convex and $\lambda_n$-smooth.

(b) Show that gradient descent on $f(\cdot)$ starting at some point $x_0 \in \mathbb{R}^n$ guarantees that

$$\|x_t - x^*\|_2 \leq \max\{\|\mu_1\|, \|\mu_n\|\}^t \cdot \|x_0 - x^*\|_2$$

where $\mu_1 \leq \ldots \leq \mu_n$ are the eigenvalues of $(I - \eta A)$.

(c) Show that $\|x_t - x^*\|_2 \leq \varepsilon \cdot \|x_0 - x^*\|_2$ after $O(\kappa \log \frac{1}{\varepsilon})$ iterations.

(d) (Do not submit.) Define $\|x\|_A = \sqrt{x^\top Ax}$. Show that $\|x_t - x^*\|_A \leq \varepsilon \cdot \|x_0 - x^*\|_A$ after $O(\kappa \log(\kappa/\varepsilon))$ iterations.

5. Gradient Descent, meet Linear Optimization. When we did constrained gradient descent (over convex body $K \subseteq \mathbb{R}^n$), we took a step along the negative gradient, and then projected back to $K$. Here is a different approach that uses linear programming. Given $x_t \in K$, the next iterate is:

$$y_t \leftarrow \arg\min_{y \in K} \{((\nabla f(x_t))^\top y)\}$$

$$x_{t+1} \leftarrow (1 - \eta_t)x_t + \eta_t y_t.$$
The minimizer in the first step can be found using linear optimization over \( K \), which may be much simpler than general convex optimization.

(a) Assume that the function \( f \) is \( \beta \)-smooth with respect to some norm \( \| \cdot \| \), \( R = \max_{a,b \in K} \| a - b \| \) (using the same norm), and \( \eta_t = \frac{2}{t+1} \). Show that for all \( t \geq 0 \):

\[
f(x_{t+1}) - f(x_t) \leq \eta_t (f(x^*) - f(x_t)) + \frac{\beta^2 \eta_t^2 R^2}{2}.
\]

(b) Hence show \( f(x_t) - f(x^*) \leq \frac{2\beta R^2}{t+1} \) via induction.

(c) (Do not submit.) Show that if \( K \) is a polytope with \( m \leq n \) non-trivial constraints (and \( n \) non-negativity constraints), then we can choose \( y_t \) to be a vertex of this polytope which will have at most \( m \) non-zero coordinates. Hence if we start with \( x_0 = \vec{0} \), \( x_t \) will have support over at most \( mt \) coordinates. E.g., for \( K = \Delta_n = \{ x \in \mathbb{R}^n \mid \sum_i x_i = 1, x_i \geq 0 \} \), \( x_t \) will have at most \( t \) non-zeroes. Moreover, \( f(x_t) - f(x^*) \leq \frac{2\beta}{t+1} \) in this case.

Now let us use this to solve a problem. Given a topic matrix \( A \in \mathbb{R}^{m \times n} \) where each column is a possible “topic” and each row is a “word”, and given a “document” \( y \in \mathbb{R}^m \), we want to find some set of topics would generate this document. Formally, we want to find \( x \in \mathbb{R}^n \) s.t.

\[
y \text{ is close to } Ax \quad \text{and} \quad x \text{ is sparse}.
\]

Since \( \| x \|_0 \) minimization is hard, we often use \( \| x \|_1 \)-minimization instead.

\[
\min_{x \in \mathbb{R}^n} \| y - Ax \|_2^2 \quad \text{s.t.} \quad \| x \|_1 = 1.
\]

In this setting \( n \gg m \) (there are billions of topics, but only \( \approx 10^5 \) words), but we don’t want to take \( \text{poly}(n) \) time! However, we have an oracle that given \( w \in \mathbb{R}^m \) can output \( \arg\min_{j \in [n]} \langle w, A \cdot j \rangle \) in \( Z \geq m \) time.

(d) Show that each iterate \( x_t \) can be computed in \( O(Zt + t^2) \), by keeping track of just the non-zero entries in \( x_t \).

(e) Show that if each column length \( \| A \cdot j \|_2 \leq L \) then \( f(x) = \| y - Ax \|_2^2 \) is \( 2L^2 \)-smooth w.r.t. the \( \ell_1 \)-norm. (You may use Exercise 4 without proof.) Observe that the \( \ell_1 \)-diameter \( R \) of the polytope \( K = \{ x \mid \| x \|_1 = 1 \} \) is 1.

Hence we can get an \( \varepsilon \)-approximate solution in time \( O(ZL^2/\varepsilon + L^4/\varepsilon^2) \).