Homework Out: March 26
Due Date: April 9, midnight

The HW contains some exercises (fairly simple problems to check you are on board with the concepts; dont submit your solutions), and problems (for which you should submit your solutions, and which will be graded). Some problems have sub-parts that are exercises. For this problem set, its OK to work with others. (Groups of 2, maybe 3 max.) That being said, please think about the problems yourself before talking to others. Please cite all sources you use, and people you work with. The expectation is that you try and solve these problems yourself, rather than looking online explicitly for answers. Submissions due at beginning of class on the due date. Please check the Piazza for details on submitting your *LaTeXed* solutions.

Exercises

1. **Lots of Flows.** Suppose you wanted to find an approximate solution to the following multicommodity 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 $\mathbb{P}_i$ 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 $\cup_i \mathbb{P}_i$.

   $$\begin{array}{ll} \sum_{P\in P_i} f_P = F_i & \forall i \in [k] \\ \Sigma_i \sum_{P\in P_i e\in P} f_P \leq 1 & \forall e \in E \end{array} \tag{1}$$

   (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 $(1 + \epsilon)$ capacity on each edge in time $\frac{O(k(m+n\ln n))}{\text{poly}(\epsilon)} \cdot (\sum_i F_i) = \frac{O(km(m+n\ln n))}{\text{poly}(\epsilon)}$.

2. **Strength in Convexity.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is called ℓ-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 η_t, then we can find $\hat{x} \in \mathbb{R}^n$ such that

   $$f(\hat{x}) - f(x^*) \leq O(\frac{G^2 \log T}{\ell \cdot T}).$$

   Again, assume that $||\nabla f(x)|| \leq G$. Note due to the assumption of strong convexity, we got better convergence: the dpendence on T is better, there's no dependence on $D = ||x_0 - x^*||$. Show that this analysis also works in the online case if each function is strongly convex. Bonus: try and remove the $\log T$ term in the offline case. Why does this not work in 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

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

   Show the following:

   (a) Give examples where $KL(p||q) \neq KL(q||p)$ where both divergences are finite.

   (b) Show $KL(p||q) \geq 0$ and $KL(p||q) = 0$ only if $p = q$.

(c) If U_N is the uniform distribution over N elements, then $KL(p||U_n) \leq \log_2 N - H(p)$ where $H(p)$ is the Shannon entropy of p.

(d) Show that the Bregman divergence $D_h(p||q) = KL(p||q)$ when $h(x) = \sum_i x_i \log_2 x_i$ is the negative entropy function.

And here are some useful facts about Bregman divergences. Recall that $D_h(x||y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle$ for a strinctly convex h.

(e) Given points $x_{1,2} \ldots, x_n$, show that the unique point c that minimizes the average distance $\frac{1}{n} \sum_i D_h(x_i||c)$ is the center of mass $c = \frac{1}{n} \sum_i x_i$.

4. **That's the Norm..** In lecture, we defined a differentiable convex function $f : K \to \mathbb{R}$ to be L Lipschitz with respect to $|| \cdot ||$ 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 α-strongly-convex with respect to $||\cdot||$ is equivalent to $||\nabla f(x) - \nabla f(y)||_* leq \beta ||x - y||$.

Problems

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 $S \subseteq V$ such that $|S|$ is odd, and $x(\partial S)$ is minimized, where $x(\delta(S)) = \sum_{i \in S, j \notin S} x_{ij}$. Then, comparing this min-odd-cut value to 1, we can find a violated constraint if one exists. Assume $|V|$ is even, otherwise the LP is infeasible.

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B).$$

Show that $f(A) = x(\partial(A))$ is submodular. Observe f is symmetric: $f(A) = f(V \setminus A)$.

(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 either contained within C or $\bar{C}$.

(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 have some matrix $M \in \mathbb{R}^{m \times n|}$; if the row player plays $x \in \Delta^m$ and the column player plays $y \in \Delta^n$, the payoff for the row player is $x^T My$.

If we define $C(x) = \min_{y \in \Delta^n} x^T My$ and $R(y) = \max_{x \in \Delta^m} x^T My$, the minimax theorem proves that (a) for all $x, y, C(x) \leq R(y)$, and (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 minimum in its definition, looking to find $\max_x \min_y x^T My$, which in this form is not 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 the above such that the optimal solution for the dual for the first LP is a solution to the second LP.

(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: $K = \{f | f_P \geq 0, \sum_{P \in \mathcal{P}} f_P = F\}$, with constraints

$$f_e / c_e \leq 1 \qquad \forall e \in E,$$

where $f_e = \sum_{P:e\in P} f_P$. In lecture we considered $c_e = 1$, now we consider the general case. Given the weights $p \in \Delta^m$ given by Hedge, the "average" constraint looks like

$$\sum_e p_e(f_e/c_e) = \sum_e f_e p_e/c_e \leq 1.$$

(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 $\Omega(mc_{\max})$ so with general capacities this ratio could be much larger than m. Now, we'll try the idea of bumping all of the edge weights up a bit to reduce the width of the problem at the expense of some approximation in our solution. For the remainder of the problem, assume $\epsilon \leq \frac{1}{10}$. You may assume the instance is feasible (one can send a flow of value F while satisfying all capacity constraints).

(b) Set the weights to be $w_e = p_e + \epsilon/m$. Compute the shortest path P^* w.r.t. edge lengths $\frac{w_e}{c_e}$, and send all F flow along it. Show

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

(c) Show that

$$\max_{e\in P^*} F/c_e \leq O(m/\epsilon)$$

and so the width of the oracle is $O(m/\epsilon)$.

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

(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 those capacities.

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

4. **Solving a Linear System..** Given a positive-definite $A \in \mathbb{R}^{n\times n}$ with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, the condition number of 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 $f(x) = \frac{1}{2}x^T AX - bx$.

(a) (Do not submit) Show that f is convex, and $\nabla f(x) = Ax - b$. Hence infer that the minimizer x^* of f satisfies $Ax^* = b$. Also show that f is λ_1-strongly convex and λ_n-smooth.

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

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

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

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

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

5. **Gradient Descent, meet Linear Optimization..** When we did constrained GD over a convex K, we took a step along the negative gradient and projected back to K. Here is a different approach using LPs. Give $x_t \in K$ the next iterate is

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

$$x_t \leftarrow (1 - \eta_t)x_t + \eta_t y_t.$$

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 β-smooth w.r.t. some norm $|| \cot ||$, and $R = \max_{a,b \in K} ||a - b||$ in the same norm, and $\eta_t = \frac{2}{t+1}$. Show that for all $t \geq 0$:

$$f(x_{t+1}) - f(x_t) \leq \eta_t \left(f(x^*) - f(x_t)\right) + \frac{\beta}{2}\eta_t^2 R^2$$

(b) Use induction to show that $f(x_t) - f(x^*) \leq \frac{2\beta R^2}{t+1}$.

(c) (Do not submit.) Suppose K is a polytope with n non-negativity and m other constraints. Show that we can choose y_t to be a vertex of K with at most m non-zero coordinates. So, if we start with $x_0 - \bar{o}$, x_t will have support over at most mt coordinates. Show the previous problems objective guarantee for this x_t.

Now we use this algorithmic setup to solve the "topic model" problem. Consider a topic matrix $A \in \mathbb{R}^{m \times n}$, with columns being associated with topics and rows being associated with words. Let $y \in \mathbb{R}^m$ be a document, and we'd like to find a set of topics x such that y was likely generated by those topics:

y is close to Ax and x is sparse.

Since $||x||_0$ minimization is hard, we often use $||x||_1$ minimization instead:

$$\min_{x \in \mathbb{R}^n} ||y - A_x||_2^2 \text{ s.t. } ||x||_1 = 1.$$

In this setting $n \gg m$, many many many topics but many fewer words, and we'd ideally like to avoid running in poly(n) time. Suppose we have an oracle that given $w \in \mathbb{R}^m$ outputs $\arg\min_{j \in [n]} \langle w, A_j \rangle$ in $Z \geq m$ time.

(d) Show that each iterate x_t can be compute in $O(Zt + t^2)$ time, by keeping track of the nonzero entries in x_t.

(e) Show that if each column length $||A_j||_2 \leq L$ then $f(x) = ||y - Ax||_2^2$ is $2L^2$-smooth w.r.t. the ℓ_1 norm. You may use exercise 4 without proof. Observe that the ℓ_1-diameter R of the polytope $K = \{x | ||x||_1 = 1\}$ is 1.

So, we get an ϵ-approximate solution in time $O(ZL^2/\epsilon + L^4/\epsilon^2)$.