EE364b Spring 2023 Homework 4

Due Sunday 5/7 at 11:59pm via Gradescope

4.1 (4 points) Consider the problem

$$\begin{array}{ll} \text{minimize} & (x_1 - b_1)^2 + (x_2 - b_2)^2 \\ \text{subject to} & x_1 = x_2, \end{array}$$

where $x_1, x_2, b_1, b_2 \in \mathbf{R}$ are scalars. Here x_1 and x_2 are local variables, which need to satisfy the consensus constraint $x_1 = x_2$.

(a) (1 point) Derive the dual decomposition updates for x_1, x_2 and λ where λ is the dual variable that corresponds to the constraint $x_1 = x_2$. (See page 10 of dual decomposition lecture slides)

(b) (1 point) Find the value of the optimal dual parameter λ^* as a function of b_1 and b_2.

(c) (1 point) Show that the dual decomposition method yields dual iterates $\lambda^{(k)}$ that obey

$$\lambda^{(k+1)} - \lambda^* = (1 - \alpha)(\lambda^{(k)} - \lambda^*)$$

where α is the fixed step size in the dual subgradient method update, and k is the iteration counter.

(d) (1 point) Show that the iterates $x_1^{(k)}, x_2^{(k)}, \lambda^{(k)}$ converge to their optimal values for a small enough step size α.

4.2 *(4 points) Distributed ridge regression.* Consider the constrained ℓ_2-regularized least-squares ('ridge regression') problem

$$\begin{array}{ll} \text{minimize} & f(z) = (1/2) \left\| \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\|_2^2 + \mu \left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2^2 \\ \text{subject to} & x_1 + x_2 \succeq 0, \end{array}$$

with optimization variable $z = (x_1, x_2) \in \mathbf{R}^n \times \mathbf{R}^n$ and $\mu > 0$. We can think of x_i as the local variable for system i, with $x_1 + x_2 \succeq 0$ serving as a complicating constraint.

(a) *(2 points) Primal decomposition.* Explain how to solve this problem using primal decomposition, using the subgradient method for the master problem.

(b) *(2 points) Dual decomposition.* Explain how to solve this problem using dual decomposition, using the projected subgradient method for the master problem. Are we guaranteed that the primal variables $x_i^{(k)}$ converge to optimal values, and why?

4.3 *(5 points) Kelley's cutting-plane algorithm.* We consider the problem of minimizing a convex function $f : \mathbf{R}^n \to \mathbf{R}$ over some convex set C, assuming we can evaluate $f(x)$ and find a subgradient $g \in \partial f(x)$ for any x. Suppose we have evaluated the function and a subgradient at $x^{(1)}, \dots, x^{(k)}$. We can form the piecewise-linear approximation

$$\hat{f}^{(k)}(x) = \max_{i=1,\dots,k} \left(f(x^{(i)}) + g^{(i)T}(x - x^{(i)}) \right),$$

which satisfies $\hat{f}^{(k)}(x) \le f(x)$ for all x. It follows that

$$L^{(k)} = \inf_{x \in C} \hat{f}^{(k)}(x) \le p^\star,$$

where $p^\star = \inf_{x \in C} f(x)$. Since $\hat{f}^{(k+1)}(x) \ge \hat{f}^{(k)}(x)$ for all x, we have $L^{(k+1)} \ge L^{(k)}$.

In Kelley's cutting-plane algorithm, we set $x^{(k+1)}$ to be any point that minimizes $\hat{f}^{(k)}$ over $x \in C$. The algorithm can be terminated when $U^{(k)} - L^{(k)} \le \epsilon$, where $U^{(k)} = \min_{i=1,\dots,k} f(x^{(i)})$.

(a) (3 points) Use Kelley's cutting-plane algorithm to minimize the piecewise-linear function

$$f(x) = \max_{i=1,\dots,m} (a_i^T x + b_i)$$

that we have used for other numerical examples, with C the unit cube, *i.e.*, $C = \{x \mid \|x\|_\infty \le 1\}$. Generate the same data we used before using

```
n = 20; % number of variables
m = 100; % number of terms
randn('state',1);
A = randn(m,n);
b = randn(m,1);
```

You can start with $x^{(1)} = 0$ and run the algorithm for 40 iterations. Plot $f(x^{(k)})$, $U^{(k)}$, $L^{(k)}$ and the constant $p^\star$ (on the same plot) versus k.

(b) (2 points) Repeat for $f(x) = \|x - c\|_2$, where c is chosen from a uniform distribution over the unit cube C. (The solution to this problem is, of course, $x^\star = c$.)

4.4 *(5 points) Maximum and minimum volume ellipsoids.* Consider the convex set

$$\mathcal{C} = \{x \mid Ax \preceq b\},$$

where $A \in \mathbf{R}^{n \times d}$ and $b \in \mathbf{R}^n$. (The data files `Amatrix` and `bvector` are available on Canvas.) The set $\mathcal{C}$ is a convex polytope (i.e. a bounded polyhedron). This question considers the best inner/outer approximations of $\mathcal{C}$ using ellipsoids.

(a) (1 point) An *extreme point* of a convex set $\mathcal{S}$ is a point $\bar{x}$ such that for every $x_1, x_2 \in \mathcal{S}$, $x_1, x_2 \neq \bar{x}$, there does not exist $\alpha \in (0,1)$ such that $\bar{x} = \alpha x_1 + (1-\alpha)x_2$. In other words, $\bar{x}$ cannot be formed as the convex combination of any other points in $\mathcal{S}$.

The extreme points of a polyhedron are called *vertices.* Vertices admit the following useful characterization. Let a_i, $i \in \{1, \ldots, n\}$, be the rows of A and recall that the i'th constraint of $Ax \preceq b$ is active if and only if $\langle a_i, x\rangle = b_i$. Then, $\bar{x} \in \mathcal{C}$ is a vertex of $\mathcal{C}$ if and only if the set of active constraints

$$\{a_i : \langle a_i, x\rangle = b_i\},$$

contains d linearly independent vectors.

Compute the vertices of $\mathcal{C}$ and report the number. *Hint: You can brute-force the set of vertices by checking all $\binom{n}{d}$ sets of d constraints for linear independence. If $\mathcal{I}$ is such a set, then you can use $A_{\mathcal{I}}^{-1} b_{\mathcal{I}}$ to identify a potential vertex, where $A_{\mathcal{I}}$ is the sub-matrix formed by the rows index by $\mathcal{I}$. Don't forget to check that the potential vertex is feasible.*

(b) (2 points) Find the center of the maximum volume ellipsoid in $\mathcal{C}$ and the center of the minimum volume ellipsoid containing $\mathcal{C}$. What is the ratio of their respective volumes, i.e. $\mathrm{vol}(\mathcal{E}_{\text{small}})/\mathrm{vol}(\mathcal{E}_{\text{big}})$, where $\mathrm{vol}(\mathcal{E}_{\text{small}})$ is the maximum volume ellipsoid inside $\mathcal{C}$ and $\mathrm{vol}(\mathcal{E}_{\text{big}})$ is the minimum volume ellipsoid containing $\mathcal{C}$. You may use CVX/CVXPY. *Hint: See 364a slides for calculating the maximum/minimum volume ellipsoid. You may find it useful to remember that maximum of a concave function over a polytope is achieved by at least one vertex of the polytope.*

(c) (2 points) Denote the two centers (vectors in $\mathbf{R}^d$) in part (a) by x_{small} and x_{big} respectively. Let $g \in \mathbf{R}^d$ be the all-ones vector. We will consider the cuts $g^T(x - x_{\text{small}}) \geq 0$ and $g^T(x - x_{\text{big}}) \geq 0$. Estimate the volume ratios

$$R_{\text{small}} := \frac{\mathbf{vol}(\{g^T(x - x_{\text{small}}) \geq 0)\} \cap \mathcal{C})}{\mathbf{vol}(\mathcal{C})},$$

and

$$R_{\text{big}} := \frac{\mathbf{vol}(\{g^T(x - x_{\text{big}}) \geq 0)\} \cap \mathcal{C})}{\mathbf{vol}(\mathcal{C})},$$

by generating $M = 10^6$ i.i.d. uniformly distributed random vectors in $[-1,+1]^d$ (i.e., $x =$ `2*rand(d,1)-1` for M trials) and transforming them as $x' = O^\top x + v$, where we provide O and v in the data files `Omatrix` and `vvector` on Canvas.

Hint: Let $M_{\mathcal{C}}$ be number of random vectors that satisfy $Ax \preceq b$. Let M_{small} be the number of random vectors that satisfy $Ax \preceq b$ and $g^T(x - x_{\text{small}}) \geq 0$.

Similarly, let $M_{\rm big}$ be the number of random vectors that satisfy $Ax \preceq b$ and $g^T(x - x_{\rm big}) \geq 0$. The volume ratios can be estimated by

$$R_{\rm small} \approx \frac{M_{\rm small}}{M_{\mathcal{C}}}\,,$$

and

$$R_{\rm big} \approx \frac{M_{\rm big}}{M_{\mathcal{C}}}\,.$$