Homework Assignment 3
Discuss Session Friday July 29 in Class

Reading. Read selected sections in Luenberger and Ye's *Linear and Nonlinear Programming Fourth Edition* Chapters 3, 5, 6, 8, 9, 10, and 14.

Theoretical Homework:

1. (10') In most real applications, the (first-order) Lipschitz constant β is unknown. Furthermore, we like to use a localized Lipschitz constant β^k at iteration k such that

$$f(\mathbf{x}^k + \alpha \mathbf{d}^k) - f(\mathbf{x}^k) - \nabla f(\mathbf{x}^k)^T(\alpha \mathbf{d}^k) \leq \frac{\beta^k}{2}\|\alpha \mathbf{d}^k\|^2,$$

where $\mathbf{d}^k$ is the steepest descent direction $-\nabla f(\mathbf{x}^k)$. The goal is to decide a step-size $\alpha \approx \frac{1}{\beta^k}$.

Consider the following *forward-backward tracking method.* In the following, assume that $\beta^k \geq 1$ and $\alpha_{\max} \geq 1/\beta^k$. Notice that if $\beta^k < 1$, we can enforce it to satisfy our assumption by replacing it with $\max\{1, \beta^k\}$.

Now start at a initial guess $\alpha > 0$,

   i) If $\alpha \leq \frac{2(f(\mathbf{x}^k) - f(\mathbf{x}^k + \alpha \mathbf{d}^k))}{\|\mathbf{d}^k\|^2}$, then doubling the step-size: $\alpha \leftarrow 2\alpha$, stop as soon as the inequality is reversed or $\alpha > \alpha_{\max} (> 0)$, and select the latest α such that the inequality ($\alpha \leq \frac{2(f(\mathbf{x}^k) - f(\mathbf{x}^k + \alpha \mathbf{d}^k))}{\|\mathbf{d}^k\|^2}$) holds and $\alpha \leq \alpha_{\max}$.

   ii) Otherwise halving the step-size: $\alpha \leftarrow \alpha/2$; stop as soon as $\alpha \leq \frac{2(f(\mathbf{x}^k) - f(\mathbf{x}^k + \alpha \mathbf{d}^k))}{\|\mathbf{d}^k\|^2}$ and return it.

   (a) (4') Let $\bar{\alpha}$ be a step-size generated by the scheme. Show that $\bar{\alpha} \geq \frac{1}{2\beta^k}$.

   (b) (3') Prove that the above scheme will terminate in finite steps.

   (c) (3') Show that $f(\mathbf{x}^k + \bar{\alpha}\mathbf{d}^k) \leq f(\mathbf{x}^k) - \frac{1}{4\beta^k}\|\mathbf{d}^k\|_2^2$.

2. (10') (L_2 Regularization and Logarithmic Barrier) Consider the optimization problem

$$\begin{array}{rl} \text{minimize}_{x_1, x_2} & (x_1 - x_2 + 1)^2 \\ \text{subject to} & x_1 \geq 0 \quad x_2 \text{ "free"}. \end{array}$$

Then we may combine the L_2-regularization and barrier together, that is, for any $\mu > 0$, consider

$$\text{minimize}_{x_1, x_2} \qquad (x_1 - x_2 + 1)^2 + \frac{\mu}{2}(x_1^2 + x_2^2) - \mu \log(x_1)$$

(a) (4') Develop explicit path formula in terms of μ. What is the limit solution as $\mu \to 0$?

(b) (3') Using $\mu = 1$ and $\mathbf{x}^0 = (1,\ 0)$, apply one step of SDM with step-size $1/5$ to compute the next iterate.

(c) (3') Using $\mu = 1$ and $\mathbf{x}^0 = (1,\ 0)$, apply one step of Newton's Method to compute the next iterate.

3. (20') (L_2 Path-Following) Consider a convex function $f : R^n \to R$ in C^2 that is twice continuously differentiable. Assume that its value is bounded from below and that it has a minimizer. For any given positive parameter $\mu > 0$, consider the regulated minimization problem

$$\text{minimize } f(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x}\|^2. \tag{1}$$

(a) (3') Write down the first-order optimality condition of (1). Is it sufficient for $\mathbf{x}$ to be a minimizer?

(b) (3') Show that the minimizer, denoted by $\mathbf{x}(\mu)$, of (1) is unique for each fixed μ.

(c) (5') Show that $f(\mathbf{x}(\mu))$ is an increasing function of μ (i.e., $f(\mathbf{x}(\mu)) \geq f(\mathbf{x}(\mu'))$ if $\mu \geq \mu' > 0$), and $\|\mathbf{x}(\mu)\|$ is a decreasing function of μ.

(d) (6') Show that As $\mu \to 0^+$ (i.e., μ decreases to 0), $\mathbf{x}(\mu)$ converges to the minimizer of $f(\mathbf{x})$ with the minimal Euclidean norm.

(e) (3') Consider the specific example

$$\text{minimize}_{x_1,x_2}\ (x_1 - x_2 - 1)^2,$$

where the optimal solution set is unbounded. Write out the explicit path formula of $\mathbf{x}(\mu) = (x_1(\mu), \ldots, x_n(\mu))$ in terms of μ. What is the limit solution as $\mu \to 0$?

4. (15') (Affine-Scaling Interior-Point SD) Consider the conic constrained optimization problem

$$\min_{x} \quad f(\mathbf{x}) \qquad \text{s.t.} \quad \mathbf{x} \geq \mathbf{0} \tag{2}$$

where we assume the objective function f is first-order β-Lipschitz. Starting from $\mathbf{x}^0 = \mathbf{e} > 0$, consider the affine-scaling interior-point method as follows: at iterate $\mathbf{x}^k > 0$ let diagonal scaling matrix D be

$$D_{ii} = \min\{1,\ x_i^k\}$$

and

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k D^2 \nabla f(\mathbf{x}^k),$$

with step-size

$$\alpha^k = \min\left\{\frac{1}{\beta},\ \frac{1}{2\|D\nabla f(\mathbf{x}^k)\|_\infty}\right\}.$$

(a) (3') Show that $-D^2\nabla f(\mathbf{x}^k)$ is a descent direction.

(b) (3') Show that $\mathbf{x}^{k+1} > \mathbf{0}$ for all $k = 0, 1, ...$

(c) (6') Show that

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \leq -\min\left\{\frac{1}{2\beta}\|D\nabla f(\mathbf{x}^k)\|_\infty^2, \frac{1}{4}\|D\nabla f(\mathbf{x}^k)\|_\infty\right\}$$

(d) (3') Derive a iterative complexity bound for $\|D\nabla f(\mathbf{x}^k)\|_\infty \leq \epsilon$.

Computational Homework (group of 1-3 people):

5. (10') There is a simple nonlinear least squares approach for Sensor Network Localization:

$$\min \quad \sum_{(ij)\in N_x} \left(\|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij}^2\right)^2 + \sum_{(kj)\in N_a} \left(\|\mathbf{a}_k - \mathbf{x}_j\|^2 - d_{kj}^2\right)^2 \tag{3}$$

which is an unconstrained nonlinear minimization problem.

(a) (5') Apply the Steepest Descent Method, starting with either the origin or a random solution as the initial solution for model (3), to solve the SNL instances you created in Problem 9 of HW1. Does it work?

(b) (5') Apply the same Steepest Descent Method, starting from the SOCP or SDP solution (which may not have errors) as the initial solution for model (3), to solve the same instances in (a). Does it work? Does the SOCP or SDP initial solution make a difference?

6. (30') (Multi-Block ADMM)

Part I Implement the ADMM to solve the divergence example:

$$\begin{array}{ll} \text{minimize} & 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 \\ \text{subjectto} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0} \end{array}$$

(a) (5') Try $\beta = 0.1$, $\beta = 1$, and $\beta = 10$, respectively. Does the choice of β make a difference?

(b) (5') Add the objective function to minimize

$$0.5(x_1^2 + x_2^2 + x_3^2)$$

to the problem, and retry $\beta = 0.1$, $\beta = 1$, and $\beta = 10$, respectively. Does the choice of β make a difference?

(c) (5') Set $\beta = 1$ and apply the randomly permuted updating-order of $\mathbf{x}$ (discussed in class) to solving each of the two problems in (a) and (b). Does the iterate converge?

Part II Generate some (feasible) convex QP problems with linear equality constraints, say 30 variables and 10 constraints (i.e., $A \in R^{10\times 30}$),

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q \mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{array}$$

(d) (5') Divide the variables of $\mathbf{x}$ into 5 blocks and apply the ADMM with $\beta = 1$. Does it converge? (You may construct 5 different blocks and conduct the experiments.)

(e) (5') Apply the randomly permuted updating-order of the 5 blocks in each iteration of the ADMM. Does it converge? Convergence performance?

(f) (5') Consider the following scheme – random-sample-without-replacement: in each iteration of ADMM, randomly sample 6 variables for update, and then randomly select 6 variables from the remaining 24 variable for update, and... , till all 30 variables are updated; then update the multipliers as usual. Does it converge? Convergence performance?