

Homework Assignment 4

Discuss Session Friday March 10 in Class

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

Solve the following problems:

1. Recall the (local) second-order (SO) and scaled second-order (SSO) Lipschitz conditions (LC):

$$\text{SOLC : } \|\nabla f(\mathbf{x} + \mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\| \leq \beta \|\mathbf{d}\|^2, \text{ where } \|\mathbf{d}\| \leq .5$$

and

$$\text{SSOLC : } \|X(\nabla f(\mathbf{x} + \mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d})\| \leq \beta \mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d}, \text{ where } \|X^{-1}\mathbf{d}\| \leq .5, X = \text{diag}(\mathbf{x}).$$

Find parameter β values (or upper bounds) of (SOLC) and (SSOLC) for each of the following scalar functions:

(a) $f(x) = \frac{1}{3}x^3 + x, x > 0$

(b) $f(x) = \log(x), x > 0.$

(c) $f(x) = \log(1 + e^{-x}), x > 0$

2. Consider again the Logistic Regression, where we like to determine x_0 and \mathbf{x} to maximize

$$\left(\prod_{i, c_i=1} \frac{1}{1 + \exp(-\mathbf{a}_i^T \mathbf{x} - x_0)} \right) \left(\prod_{i, c_i=-1} \frac{1}{1 + \exp(\mathbf{a}_i^T \mathbf{x} + x_0)} \right).$$

which is equivalent to maximize the log-likelihood probability

$$- \sum_{i, c_i=1} \log(1 + \exp(-\mathbf{a}_i^T \mathbf{x} - x_0)) - \sum_{i, c_i=-1} \log(1 + \exp(\mathbf{a}_i^T \mathbf{x} + x_0)).$$

Or to minimize the log-logistic-loss

$$f(\mathbf{x}, x_0) = \sum_{i, c_i=1} \log(1 + \exp(-\mathbf{a}_i^T \mathbf{x} - x_0)) + \sum_{i, c_i=-1} \log(1 + \exp(\mathbf{a}_i^T \mathbf{x} + x_0)).$$

- (a) Write down the Hessian matrix function of \mathbf{x} and x_0 .
 - (b) (Computation Team Work) Apply any Quasi-Newton (e.g., slide 18 of Lecture #13 or L&Y Chapter 10) and Newton methods to solve the problem using the data in HW2 for SVM (may or may not with regulation), randomly generate data sets, and/or benchmark data sets you can find. Compare the two methods with each other and with the previous methods used in HW3.
3. Consider the LP problem

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) = x_1 + x_2 \\ & \text{subject to} && x_1 + x_2 + x_3 = 1, \\ & && (x_1, x_2, x_3) \geq 0. \end{aligned}$$

- (a) What is the analytic center of the feasible region with the logarithmic barrier function?
- (b) Find the central path point $\mathbf{x}(\mu) = (x_1(\mu), x_2(\mu), x_3(\mu))$.
- (c) Show that as μ decreases to 0, $\mathbf{x}(\mu)$ converges to the unique optimal solution.
- (d) (Computation Team Work) Draw \mathbf{x} part of the primal-dual potential function level sets:

$$\psi_6(\mathbf{x}, \mathbf{s}) \leq 0 \quad \text{and} \quad \psi_6(\mathbf{x}, \mathbf{s}) \leq -10,$$

and

$$\psi_{12}(\mathbf{x}, \mathbf{s}) \leq 0 \quad \text{and} \quad \psi_{12}(\mathbf{x}, \mathbf{s}) \leq -10;$$

respectively in the primal feasible region (on a plane).

Hint: Sample interior points in the primal and dual feasible regions. To plot the \mathbf{x} part of the level set of potential function, say $\psi_6(\mathbf{x}, \mathbf{s}) \leq 0$, in primal feasible region F_p , you plot

$$\{\mathbf{x} \in F_p : \min_{\mathbf{s} \in F_d} \psi_6(\mathbf{x}, \mathbf{s}) \leq 0\}$$

where F_d represents the dual feasible region. This can be approximately done by sampling as follows.

You randomly generate N interior feasible points of the primal \mathbf{x}^p and the dual $(\mathbf{y}^q, \mathbf{s}^q)$, respectively. For each primal point \mathbf{x}^p , you find if it is true that

$$\min_{q=1, \dots, N} \psi_6(\mathbf{x}^p, \mathbf{s}^q) \leq 0.$$

Then, you plot those \mathbf{x}^p who give an "yes" answer.

(f) Changing the objective to $f(\mathbf{x}) = x_1$ and repeat questions (b)–(d).

4. Questions (a) and (b) of Problem 7, Section 5.9, the text book L&Y.

Hint: Use the fact that, for any feasible pair $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ of LP,

$$(\mathbf{x} - \mathbf{x}(\mu))^T (\mathbf{s} - \mathbf{s}(\mu)) = 0,$$

the optimality of the central path solutions.

5. Problem 12, Section 6.8, the text book L&Y, where for any given symmetric matrix D , $|D|^2$ is the sum of all its eigenvalue squares, and $|D|_\infty$ is its largest absolute eigenvalue.

Hint: $\det(I + D)$ equals the product of the eigenvalues of $I + D$. Then the proof follows from Taylor's expansion.

6. Optimization with log-sum-exponential functions arises from smooth approximation for non-smooth optimization. Consider the non-smooth optimization problem:

$$\min_{\mathbf{x}} \max_{1 \leq i \leq m} (\mathbf{a}_i^T \mathbf{x} + b_i), \quad (1)$$

given $\mathbf{a}_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$.

(a) Derive an equivalent LP problem and write down its dual.

(b) Suppose we approximate the objective function $\max_{1 \leq i \leq m} (\mathbf{a}_i^T \mathbf{x} + b_i)$ with a smooth function and consider a different optimization problem:

$$\min_{\mathbf{x}} \log \left(\sum_{i=1}^m \exp(\mathbf{a}_i^T \mathbf{x} + b_i) \right) \quad (2)$$

Let z_1 and z_2 be the optimal values of (1) and (2), prove that

$$0 \leq z_2 - z_1 \leq \log m.$$

Hint: Use the monotone property of both functions $\log(\cdot)$ and $\exp(\cdot)$.

(c) Suppose we use a different function for approximation:

$$\min_{\mathbf{x}} \frac{1}{\gamma} \log \left(\sum_{i=1}^m \exp(\gamma(\mathbf{a}_i^T \mathbf{x} + b_i)) \right), \quad (3)$$

for some $\gamma > 0$. Suppose the optimal value to (3) is z_3 , derive a bound for $z_3 - z_1$ similar as above. What happens as $\gamma \rightarrow \infty$?

7. (20pts Computation Team Work) Implement the ADMM to solve the divergence example in Lecture &16.

- (a) Try $\beta = 0.1$, $\beta = 1$, and $\beta = 10$, respectively. Does the choice of β make a difference?
- (b) 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) Set $\beta = 1$ and apply the random permutation updating order of \mathbf{x} (discussed in class) to solving each of the two problems in (a) and (b). Does the iterate converge?

8. (20pts Computation Team Work) All pieces of Computation Team Work stated above, and one more described below.

You understand the SDP relaxation for SNL very well now, e.g., the two sensors and three anchors formulation: Find

$$Z = \begin{pmatrix} I & X \\ X^T & Y \end{pmatrix} \in S^4$$

to meet the constraints in the standard form:

$$\begin{aligned} (1; 0; 0; 0)(1; 0; 0; 0)^T \bullet Z &= 1, \\ (0; 1; 0; 0)(0; 1; 0; 0)^T \bullet Z &= 1, \\ (1; 1; 0; 0)(1; 1; 0; 0)^T \bullet Z &= 2, \\ (a_i; -1; 0)(a_i; -1; 0)^T \bullet Z &= d_{1i}^2, \quad i = 1, 2, \\ (a_i; 0; -1)(a_i; 0; -1)^T \bullet Z &= d_{2i}^2, \quad i = 2, 3, \\ (0; 0; 1; -1)(0; 0; 1; -1)^T \bullet Z &= \hat{d}_{12}^2, \\ Z &\succeq 0 \in S^4. \end{aligned}$$

where no objective function is present.

The question is: could we add an suitable (or regulative) objective function to improve

localization quality:

$$\begin{array}{ll}
\min & C \bullet X \\
\text{s.t.} & (1; 0; 0; 0)(1; 0; 0; 0)^T \bullet Z = 1, \\
& (0; 1; 0; 0)(0; 1; 0; 0)^T \bullet Z = 1, \\
& (1; 1; 0; 0)(1; 1; 0; 0)^T \bullet Z = 2, \\
& (a_i; -1; 0)(a_i; -1; 0)^T \bullet Z = d_{1i}^2, \quad i = 1, 2, \\
& (a_i; 0; -1)(a_i; 0; -1)^T \bullet Z = d_{2i}^2, \quad i = 2, 3, \\
& (0; 0; 1; -1)(0; 0; 1; -1)^T \bullet Z = \hat{d}_{12}^2, \\
& Z \succeq 0 \in S^4.
\end{array}$$

Try the following regulative objectives and construct the corresponding objective matrix C , and see which works best:

- Minimize the trace of Z
- Maximize the trace of Z ,
- Minimize the sum of all the non-edge distance squares (a non-edge is a edge whose distance information is unknown)
- Maximize the sum of the non-edge distance squares

You may use your generated data as in your HW2.

You may use CVX (or DSDPMLB.m code described in class) to solve your SDP problems.