Optimization Methods for Data Science, Machine Learning and AI
Prof. Yinyu Ye

Homework Assignment 4

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

1. Recall that the (local) second-order (SO), concordant second-order (CSO) and scaled concordant second-order (SCSO) Lipschitz conditions (LC) are defined as follows:

   $$\text{SOLC}: \ \|\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\| \leq \beta \|\mathbf{d}\|^2, \text{ where } \|\mathbf{d}\| \leq C \text{ for some } C > 0$$

   $$\text{CSOLC}: \ \|\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\| \leq \beta |\mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d}|, \text{ where } \|\mathbf{d}\| \leq C \text{ for some } C > 0,$$

   and

   $$\text{SCSOLC}: \ \|X(\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d})\| \leq \beta |\mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d}|,$$
   $$\text{where } \|X^{-1}\mathbf{d}\| \leq C \text{ for some } C > 0,$$

   and $X = \text{diag}(\mathbf{x} > \mathbf{0})$. Here we have implicitly assumed/required that $\mathbf{x}$ and $\mathbf{x}+\mathbf{d}$ are in the domain of f. Here the constant C should be independent of $\mathbf{x}$.

   For each of the following scalar functions, find the Lipschitz parameter β value of (SOLC), (CSOLC) and (SCSOLC). You can provide an upper bound on β or state that it doesn't exist.

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

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

   (c) $f(x) = x\log(x)$, $x > 0$

2. Consider the following questions:

   (a) Let $\phi(\mathbf{y})$, where $\mathbf{y} \in R^m$, be (regular) β-second-order (SO) Lipschitz and be δ-strongly convex, that is, for all $\mathbf{y}$ in the domain of ϕ, the largest eigenvalue of Hessian $\nabla^2\phi(\mathbf{y})$ is bounded above by $\beta > 0$ and the smallest eigenvalue of $\nabla^2\phi(\mathbf{y})$ is bounded below by $\delta > 0$. Prove that the function

   $$f(\mathbf{x}) = \phi(A\mathbf{x}),$$

where $A \in R^{m\times n}$, $n \geq m$, is a constant coefficient matrix with rank m, is <u>concordant</u> second-order Lipschitz for all $\mathbf{x} \in R^n$ such that $\mathbf{y} = A\mathbf{x}$ is in the domain of ϕ.

(b) Find the <u>concordant</u> Lipschitz bounds α for the following three functions (or show that a global constant doesn't exist):

- $f(\mathbf{x}) = \frac{1}{2}(x_1 + x_2)^2$
- $f(\mathbf{x}) = e^{x_1+x_2}$
- $f(\mathbf{x}) = (x_1 + x_2)\log(x_1 + x_2)$ where $x_1 + x_2 > 0$.

3. Prove the logarithmic approximation lemma for SDP. Let $D \in S^n$ and $|D|_\infty < 1$. Then,

$$Tr(D) \geq \log\det(I + D) \geq Tr(D) - \frac{|D|^2}{2(1 - |D|_\infty)}$$

where for any given symmetric matrix D, $|D|^2$ is the sum of all its squared eigenvalues, 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.