CME 307 / MS&E 311 Summer 2021-2022
Optimization To be discussed on August 5, 2022
Prof. Yinyu Ye

Homework Assignment 4 Sample Solution

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:

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

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

and

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

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

Solution: Basic comments:

- The (local) here actually only means for a bounded region of d instead of arbitrary d. But it's global in terms of x. But we are accepting solutions that talks about local constants for x. Although in general, proving non-existence in global sense (for x) is also not that difficult.

- By saying that you can provide an upper bound on β, we just mean that you don't need to provide the tightest β.

The solution below is talking about **global** constants for x.

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

Note that $f'(x) = x^2 + 1$, $f''(x) = 2x$.

The SOLC condition holds for $\beta = 1$. To see this, we observe that for all $x > 0$, and d such that $x + d > 0$,

$$|f'(x+d) - f'(x) - f''(x) \cdot d| = d^2$$

Hence $f(x)$ is 1-SOLC.

The CSOLC does not hold for any β. To see this, simply notice that the LHS is still d^2, while the RHS becomes $2|x|\beta d^2$. By taking $x \to 0$, we see that no β will satisfy the CSOLC.

The SCSOLC holds for $\beta = 1/2$. For all $x > 0$, and d such that $x + d > 0$, we have that

$$|x(f'(x+d) - f'(x) - f''(x) \cdot d)| = xd^2 = \frac{1}{2}|d^2 f''(x)|$$

Hence $f(x)$ is 1/2-SCSOLC.

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

Note that $f'(x) = -x^{-1}$, $f''(x) = x^{-2}$, and that

$$|f'(x+d) - f'(x) - f''(x)d| = \frac{d^2}{x^2(x+d)}$$

The SOLC does not hold for any $\beta > 0$. To see this, simply notice that for any $d > 0$ (no matter how small it is), by taking $x \to 0+$, the LHS goes to $+\infty$ while the RHS βd^2 remains finite, and hence no β satisfies this inequality.

The CSOLC does not hold for any $\beta > 0$. To see this, simply notice that the RHS is $\beta d^2/x^2$, and hence LHS $\leq$ RHS $\Rightarrow 1/(x+d) \leq \beta$. By taking both x and d going to 0, we see that β can not be finite.

The SCSOLC holds for $\beta = 2$ if $|x^{-1}d| \leq \frac{1}{2}$. To see that, for all $x > 0$ and d such that $|x^{-1}d| \leq 1/2$, we have $1 + \frac{d}{x} \geq \frac{1}{2}$. It follows that

$$|x(f'(x+d) - f'(x) - f''(x)d)| = \frac{d^2}{x(x+d)} = \frac{d^2}{x^2\left(1 + \frac{d}{x}\right)} \leq 2\frac{d^2}{x^2} = 2|d^2 f''(x)|.$$

Hence f is 2-SCSOLC provided $|x^{-1}d| \leq \frac{1}{2}$.

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

Note that $f'(x) = 1 + \log x$, $f''(x) = 1/x$, and that for any d such that $x + d > 0$,

$$|f'(x+d) - f'(x) - f''(x)d| = \frac{d}{x} - \log\left(1 + \frac{d}{x}\right).$$

Recall that $\frac{x}{1+x} \leq \log(1+x) \leq x$ for all $x > -1$.

The SOLC does not hold for any $\beta > 0$. To see this, notice that by the L'Hospital rule, we have for any fixed $x > 0$,

$$\lim_{d\to 0} \frac{|f'(x+d) - f'(x) - f''(x)d|}{d^2} = \frac{1}{2x^2},$$

which is unbounded as x goes to 0.

The CSOLC does not hold for any $\beta > 0$. To see this, again notice that by the L'Hospital rule, we have for any fixed $x > 0$,

$$\lim_{d\to 0} \frac{|f'(x+d) - f'(x) - f''(x)d|}{d^2/x} = \frac{1}{2x},$$

which is again unbounded as x goes to 0.

The SCSOLC holds for $\beta = 2$ if $|x^{-1}d| \leq \frac{1}{2}$. To see this, notice that when $|x^{-1}d| \leq 1/2$, we have

$$\frac{|x||f'(x+d) - f'(x) - f''(x)d|}{d^2/x} = \frac{d/x - \log(1 + d/x)}{d^2/x^2} \leq 2.$$

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 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 concordant 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 concordant 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 $x_1 + x_2 > 0$.

Solution:

(a) The key is to notice that $\nabla f(x) = A^T \nabla\phi(Ax)$ and $\nabla^2 f(x) = A^T \nabla^2\phi(Ax)A$. Then since $\phi(x)$ is second-order Lipschitz, we have that for all x, d such that $Ax, A(x+d)$ in the domain of ϕ,

$$\|\nabla\phi(Ax + Ad) - \nabla\phi(Ax) - \nabla^2\phi(Ax)Ad\| \leq \beta\|Ad\|^2, \quad \text{where } \|Ad\| \leq O(1)$$

Hence

$$\begin{aligned} &\|\nabla f(x+d) - \nabla f(x) - \nabla^2 f(x)d\| \\ &= \|A^T(\nabla\phi(Ax + Ad) - \nabla\phi(Ax) - \nabla^2\phi(Ax)Ad)\| \leq \|A^T\|_2\beta\|Ad\|^2. \end{aligned}$$

Because ϕ is strongly convex, we have that for all x,

$$|d^T\nabla^2 f(x)d| = |(Ad)^T\nabla^2\phi(Ax)(Ad)| \geq \delta\|Ad\|^2$$

It follows that

$$\|\nabla f(x+d) - \nabla f(x) - \nabla^2 f(x)d\| \leq \|A^T\|_2\frac{\beta}{\lambda}(d^T\nabla^2 f(x)d), \quad \text{where } \|Ad\| \leq O(1)$$

Because A is of full row rank, it is equivalent to say $\|d\| \leq O(1)$. Hence the Concordant second-order Lipschitz holds for f.

(b) Although it's not that difficult to talk about global constants in terms of x as in problem 1, we show how to make use of part (a) to obtain local constants. In particular, we notice that changing β and δ to local constants $\beta(y)$ and $\delta(y)$ leads to the same result (with β and δ changed to local ones, i.e. $\beta(y)$ and $\delta(y)$, of course).

– $f(x) = (x_1 + x_2)^2/2$. In this case, $A^T = [1, 1]$, and hence $\|A^T\| = \sqrt{2}$, Furthermore, $\delta = 1$ and $\beta = 0$. Hence we can set $\alpha = 0$.

– $f(x) = e^{x_1+x_2}$. In this case, again $A^T = [1, 1]$ and $\|A^T\| = \sqrt{2}$. Furthermore, $\delta(y) = e^y$ and $\beta(y) = O(e^y)$. Hence we can set $\alpha = O(1)$. Notice that here we used the local version of (a) (see the comment above at the beginning of (b)) to obtain a global constant α.

- $f(x) = (x_1 + x_2)\log(x_1 + x_2)$, where $x_1 + x_2 > 0$. Once again, $A^T = [1, 1]$ and hence $\|A\| = \sqrt{2}$. Furthermore, $\delta = 1/y$ and $\beta(y) = O(1/y^2)$, and hence we can choose $\alpha = O(1/(x_1 + x_2))$.

  Remark: Globally, by computing the LHS and RHS exactly, we can easily see that it's not CSOLC by taking $x_1 + x_2 \to \infty$.

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.

Solution:

Suppose that the eigenvalues of D are λ_j, $j = 1, \ldots, n$. Then we have

$$\log\det(I + D) = \sum_{j=1}^{n} \log(1 + \lambda_j) \leqslant \sum_{j=1}^{n} \lambda_j = \text{trace}(D) \tag{1}$$

and

$$\text{trace}(D) - \frac{|D|^2}{2(1 - |D|_\infty)} = \sum_{j=1}^{n} \lambda_j - \frac{\sum_{j=1}^{n} \lambda_j^2}{2(1 - \max_j |\lambda_j|)} \leqslant \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{n} \frac{\lambda_j^2}{2(1 - |\lambda_j|)} \tag{2}$$

Hence it suffices to prove that $\forall |x| < 1$, we have $\log(1 + x) \geqslant x - \frac{x^2}{2(1-|x|)}$.

To see this, simply notice that by Taylor's series, we have $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + x^5/5 - \ldots$. On the other hand, we also have $\frac{x^2}{2(1-|x|)} = \frac{x^2}{2}(1 + |x| + |x|^2 + \ldots) = x^2/2 + |x|^3/2 + |x|^4/2 + \cdots \geq x^2/2 - x^3/3 + x^4/4 - \ldots$. Comparing term by term, we immediately see that $\log(1 + x) \geq x - \frac{x^2}{2(1-|x|)}$, which completes our proof.