Advanced Topics in Control 2020: Large-Scale Convex Optimization

Solution to Exercise 3

Goran Banjac, Mathias Hudoba de Badyn, Andrea Iannelli,
Angeliki Kamoutsi, Ilnura Usmanova

April 2, 2020

---

1 Local Minima Are Global Minima in Convex Optimization

Assume on the contrary that $x^\star$ is not a global minimizer., i.e., there exists $\bar{x} \in \mathbb{R}^n$, such that $f(\bar{x}) < f(x^\star)$. Consider the point $x_\theta := \theta\bar{x} + (1-\theta)x^\star$ for $\theta := \min\{1, \frac{\delta}{\|x^\star - \bar{x}\|}\} \in [0,1]$. Then, we have $\|x_\theta - x^\star\| = \theta\|\bar{x} - x^\star\| \le \delta$. Thus, by the assumption on local optimality of $x^\star$, we get $f(x^\star) \le f(x_\theta)$. On the other hand, by convexity of f and since $f(\bar{x}) < f(x^\star)$, we have

$$f(x_\theta) \le \theta f(\bar{x}) + (1-\theta)f(x^\star) < \theta f(x^\star) + (1-\theta)f(x^\star) = f(x^\star).$$

Thus, we have arrived to a contradiction.

2 The Lasso Problem

Set $f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$. We have for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
f(x) &\ge \frac{1}{2}\|Ax - b\|_2^2 + \|A^\top b\|_\infty \|x\|_1 \\
&\ge \frac{1}{2}\|Ax - b\|_2^2 + b^\top Ax \\
&= \frac{1}{2}\|Ax\|_2^2 + \frac{1}{2}\|b\|_2^2 - b^\top Ax + b^\top Ax \\
&= \frac{1}{2}\|Ax\|_2^2 + \frac{1}{2}\|b\|_2^2 \ge \frac{1}{2}\|b\|_2^2,
\end{aligned}$$

where in the first line we used the assumption $\lambda \ge \|A^\top b\|_\infty$ and in the second line the inequality $b^\top Ax \le \|A^\top b\|_\infty \|x\|_1$. Since, $f(0) = \frac{1}{2}\|b\|_2^2$, we conclude that $x = 0$ is an optimal solution to the given optimization problem.

An alternative approach is to use the first-order optimality conditions for convex optimization. We have that $x = 0$ is an optimal solution if and only if $0 \in \partial f(0)$. The subgradient of f at 0 is given by

$$\partial f(0) = -A^\top b + \lambda[-1,1]^n = \{-A^\top b + \lambda x : x \in [-1,1]^n\} = \{-A^\top b + \lambda x : \|x\|_\infty \le 1\}.$$

Therefore,

$$\begin{aligned}
0 \in \partial f(0) \quad &\Leftrightarrow \quad \exists x \in \mathbb{R}^n \text{ with } \|x\|_\infty \le 1 : \; 0 = -A^\top b + \lambda x \\
&\Leftrightarrow \quad \lambda \ge \|A^\top b\|_\infty.
\end{aligned}$$

Actually, we have shown that the condition $\lambda \ge \|A^\top b\|_\infty$ is both sufficient and necessary for the optimality of $x = 0$.

3 Logistic Regression

(a) First note that since $y \in \{-1, 1\}$, we can write $p_y(x) = \frac{1}{1+e^{-y(w^\top x+b)}}$ and the likelihood becomes

$$l_w(x, y) = \prod_{i=1}^{N} p_{y_i}(x_i) = \prod_{i=1}^{N} \frac{1}{1 + e^{-y_i(w^\top x_i+b)}}$$

Maximizing $l_w(x, y)$ is the same as maximizing $\log(l_w(x, y))$, since the logarithm is monotonically increasing. Furthermore, maximizing $\log(l_w(x, y))$ is the same as minimizing $-\log(l_w(x, y))$, yielding

$$\begin{aligned} -\log l_w(x, y) &= -\log \left(\prod_{i=1}^{N} \frac{1}{1 + e^{-y_i(w^\top x_i+b)}} \right) \\ &= \log \left(\prod_{i=1}^{N} (1 + e^{-y_i(w^\top x_i+b)}) \right) \\ &= \sum_{i=1}^{N} \log \left(1 + e^{-y_i(w^\top x_i+b)} \right) \end{aligned}$$

This function is convex, since for each $i = 1, \ldots, N$, the i-th summand can be written as a composition $h(g_i(w, b))$, where $g_i(w, b) = y_i(w^\top x_i + b)$ is affine, and $g(x) = \log(1 + e^x)$ is convex ($g''(x) \geq 0$, for all x).

On the other hand, the logistic regression problem is not strongly convex. Indeed, $\frac{d^2}{dx^2} \log(1+ e^x) = \frac{e^x}{(1+e^x)^2} \to 0$ as $x \to -\infty$ and $x \to +\infty$. Therefore, it does not exist any positive lower bound (recall the second oder characterization of strongly convex functions given in Summary 3).

(b) We have

$$\begin{aligned} \sum_{i=1}^{N} \log(1 + e^{-y_i(w^\top x_i+b)}) &= \sum_{i:y_i=-1} \log(1 + e^{w^\top x_i+b}) + \sum_{i:y_i=1} \log(1 + e^{-(w^\top x_i+b)}) \\ &= \sum_{i:y_i=-1} \log(1 + e^{w^\top x_i+b}) + \sum_{i:y_i=1} \log(\frac{1 + e^{w^\top x_i+b}}{e^{w^\top x_i+b}}) \\ &= \sum_{i:y_i=-1} \log(1 + e^{w^\top x_i+b}) + \sum_{i:y_i=1} \log(1 + e^{w^\top x_i+b}) - \sum_{i:y_i=1} (w^\top x_i + b) \\ &= \sum_{i=1}^{N} \log(1 + e^{w^\top x_i+b}) - \sum_{i:y_i=1} (w^\top x_i + b). \end{aligned}$$

From here we can go over the new labels $\hat{y}_i = 1$ and $\hat{y}_i = 0$.

$$\begin{aligned} \sum_{i=1}^{N} \log(1 + e^{-y_i(w^\top x_i+b)}) &= \sum_{i=1}^{N} \log(1 + e^{w^\top x_i+b}) - \sum_{i:\hat{y}_i=1} (w^\top x_i + b) \\ &= \sum_{i=1}^{N} \log(1 + e^{w^\top x_i+b}) - \sum_{i=1}^{N} \hat{y}_i(w^\top x_i + b). \end{aligned}$$

(c) First note that the objective function

$$\sum_{i=1}^{N} \underbrace{\log(1 + e^{-y_i(x_i^\top w+b)})}_{=:f_i(w,b)}$$

is strictly positive everywhere, since this is the case for all terms f_i. We want to show that the infimum is 0. To this end, let $((w_n, b_n))_n$ be a sequence defined by $(w_n, b_n) = n(\bar{w}, \bar{b})$. Let i with $y_i = -1$. Then $f_i(w_n, b_n) = \log(1+e^{n(x_i^\top \overline{w}+\bar{b})}) \to 0$ as $n \to \infty$, since $x_i^\top \overline{w} + \bar{b} < 0$. Similarly, for i with $y_i = 1$, then $f_i(w_n, b_n) = \log(1 + e^{-n(x_i^\top \overline{w}+\bar{b})}) \to 0$ as $n \to \infty$, since $x_i^\top \overline{w} + \bar{b} > 0$.

Hence the infimum is 0 and moreover is not attained by any (w, b) since the objective is strictly positive for all (w, b).

4 ℓ_1-, ℓ_∞- and ℓ_4-Norm Approximation Problems

(a) (1) Minimize $\|Ax - b\|_\infty$ is equivalent to the following LP.

$$\begin{cases} \min_{x,t} & t \\ \text{s.t.} & Ax - b \leq t\mathbf{1} \\ & Ax - b \geq -t\mathbf{1}, \end{cases}$$

where the optimization variables are the vector $x \in \mathbb{R}^n$ and the scalar $t \in \mathbb{R}$, and $\mathbf{1}$ is the vector with all entries equal to 1. To see the equivalence of the two programs, note that if for $i = 1, \ldots, m$, $a_i^\top$ is the i-th row of the matrix A, then

$$\begin{aligned} \min_{x,t}\{t :\ -t\mathbf{1} \leq Ax - b \leq t\mathbf{1}\} &= \min_x \min_t\{t :\ -t \leq a_i^\top x - b_i \leq t,\ i = 1\ldots, m\} \\ &= \min_x \min_t\{t :\ |a_i^\top x - b_i| \leq t,\ i = 1\ldots, m\} \\ &= \min_x \min_t\{t :\ \max_{i=1,\ldots,m} |a_i^\top x - b_i| \leq t\} \\ &= \min_x \min_t\{t :\ \|Ax - b\|_\infty \leq t\} \\ &= \min_x \|Ax - b\|_\infty, \end{aligned}$$

where the last equality is because for fixed $x \in \mathbb{R}^n$, we have that $\min_t\{t :\ \|Ax - b\|_\infty \leq t\} = \|Ax - b\|_\infty$.

(2) Minimize $\|Ax - b\|_1$ is equivalent to the following LP.

$$\begin{cases} \min_{x,s} & \mathbf{1}^\top s \\ \text{s.t.} & Ax - b \leq s \\ & Ax - b \geq -s, \end{cases}$$

where the optimization variables are the vector $x \in \mathbb{R}^n$ and the vector $s \in \mathbb{R}^m$. To see the equivalence of the two programs, note that if for $i = 1, \ldots, m$, $a_i^\top$ is the i-th row of the matrix A, then

$$\begin{aligned} \min_{x,s}\{\mathbf{1}^\top s :\ -s \leq Ax - b \leq s\} &= \min_x \min_s\{\sum_{i=1}^m s_i :\ |a_i^\top x - b_i| \leq s_i,\ i = 1\ldots, m\} \\ &= \min_x \sum_{i=1}^m |a_i^\top x - b_i| \\ &= \min_x \|Ax - b\|_1, \end{aligned}$$

where the second equality holds because the objective of the LP is separable, so for fixed $x \in \mathbb{R}^n$, the optimum over s is achieved by choosing $s_i = |a_i^\top x - b_i|$, for each $i = 1, \ldots, m$.

(3) Minimize $\|Ax - b\|_1$ subject to $\|x\|_\infty \leq 1$ is equivalent to the following LP.

$$\begin{cases} \min_{x,s} & \mathbf{1}^\top s \\ \text{s.t.} & -s \leq Ax - b \leq s \\ & -\mathbf{1} \leq x \leq \mathbf{1}, \end{cases}$$

with variables $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$.

(4) Minimize $\|x\|_1$ subject to $\|Ax - b\|_\infty \leq 1$ is equivalent to the following LP.

$$\begin{cases} \min_{x,s} & \mathbf{1}^\top s \\ \text{s.t.} & -s \leq x \leq s \\ & -\mathbf{1} \leq Ax - b \leq \mathbf{1}, \end{cases}$$

with variables $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$.

(5) Minimize $\|Ax - b\|_1 + \|x\|_\infty$ is equivalent to the following LP.

$$\begin{cases} \min_{x,s,t} & \mathbf{1}^\top s + t \\ \text{s.t.} & -t\mathbf{1} \leq x \leq t\mathbf{1} \\ & -s \leq Ax - b \leq s, \end{cases}$$

with variables $x \in \mathbb{R}^n$, $s \in \mathbb{R}^m$ and $t \in \mathbb{R}$.

(b) Minimize $\|Ax - b\|_4$. is equivalent to the following QCQP.

$$\begin{cases} \min_{x,s,y} & \sum_{i=1}^m s_i^2 \\ \text{s.t.} & a_i^\top x - b_i = y_i, \ i = 1, \ldots, m \\ & y_i^2 \leq s_i, \ i = 1, \ldots, m, \end{cases}$$

with variables $x \in \mathbb{R}^n$ and $s, y \in \mathbb{R}^m$. Moreover, note that the two constraints can be merged into a single quadratic constraint. This would eliminate the variable y.