Advanced Topics in Control 2020: Large-Scale Convex Optimization

Solution to Exercise 4

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

April 3, 2020

---

1 Problem 1

Let us recall the following two optimization problems:

$$\min_{x} \quad c^\top x + \frac{1}{\mu}\sum_{i=1}^{m} \log\left(1 + e^{\mu(a_i^\top x - b_i)}\right), \tag{1}$$

and:

$$\begin{aligned} \min_{x,y} \quad & c^\top x + \frac{1}{\mu}\sum_{i=1}^{m} \log\left(1 + e^{\mu y_i}\right), \\ \text{s.t.} \quad & Ax - b \leq y, \end{aligned} \tag{2}$$

(a) It is noted first that the two objective functions in (1) and (2) differ for the change of variable $Ax - b = y$ (to be interpreted entry-wise, i.e. $a_i^\top x - b_i{=}y_i$). Therefore, the two problems are equivalent if the lower bound $Ax - b \leq y$ on y in (2) is tight (i.e. the inequality constraints hold as equality). To show this, let us introduce a vector ϵ with non-negative entries such that the optimizers of (2) satisfy $Ax^\star - b + \epsilon = y^\star$. Note that the exponential function $e^{\mu y_i}$ is monotonic with respect to y_i, so the objective function of (2) will be higher for any strictly positive entry of ϵ. Therefore, $\epsilon \equiv 0$ and the optimal value $x^\star$ is such that the inequality constraint $Ax - b \leq y$ will hold as equality at the optimal value, i.e. $Ax^\star - b = y^\star$.

    This technique of variable decoupling can be useful to write optimization problems in different but equivalent ways which might be better suited for analysis.

(b) The Lagrangian is:

$$L(x, y, z) = c^\top x + \frac{1}{\mu}\sum_{i=1}^{m} \log\left(1 + e^{\mu y_i}\right) + z^\top (Ax - b - y).$$

    The (Lagrange) dual function g is by definition the minimum value of the Lagrangian over the optimization variables of the primal problem, in this case x and y:

$$\begin{aligned} g(z) &= \inf_{x,y} \ L(x, y, z) \\ &= \inf_{x,y} \ -z^\top (b + y) + (A^\top z + c)^\top x + \frac{1}{\mu}\sum_{i=1}^{m} \log\left(1 + e^{\mu y_i}\right), \quad (z \geq 0). \end{aligned}$$

The minimum over x is unbounded below unless the term multiplying x is zero, i.e. $A^\top z + c{=}0$. To find the minimum over y we can separate the objective function and solve m separate problem:

$$\inf_{y_i} \frac{1}{\mu} \log\left(1 + e^{\mu y_i}\right) - z_i y_i$$

We can solve this by setting the derivatives of the objective function with respect to y_i to zero:

$$\frac{e^{\mu y_i}}{e^{\mu y_i} + 1} = z_i, \quad \rightarrow \quad y_i = \frac{1}{\mu} \log\left(\frac{z_i}{1 - z_i}\right)$$

Therefore the solution for (b) is:

$$\inf_{y_i} \log\left(1 + e^{\mu y_i}\right) - z_i y_i = \begin{cases} -\frac{1}{\mu}\left(z_i \log z_i + (1 - z_i)\log(1 - z_i)\right), & \text{if } 0 \le z_i \le 1 \\ -\infty, & \text{else} \end{cases}$$

where the interpretation $0 \log 0 = 0$ holds.

The expression for the Lagrange dual problem is obtained by maximizing the dual function over the variables of the dual problem, i.e. z:

$$\begin{aligned} \max_z g(z) = \quad \max_z \quad & -b^\top z - \frac{1}{\mu} \sum_{i=1}^m \left(z_i \log z_i + (1 - z_i)\log(1 - z_i)\right), \\ \text{s.t.} \quad & A^\top z + c = 0, \\ & 0 \le z \le 1 \end{aligned}$$

(c) We first plug the minimizer $x^\star$ of the primal LP in the objective function of the original problem (1). This allows us to say:

$$\begin{aligned} q^\star \le & c^\top x^\star + \frac{1}{\mu} \sum_{i=1}^m \log\left(1 + e^{\mu(a_i^\top x^\star - b_i)}\right) \\ = & p^\star + \frac{1}{\mu} \sum_{i=1}^m \log\left(1 + e^{\mu(a_i^\top x^\star - b_i)}\right) \end{aligned} \tag{3}$$

Note now that, since the primal LP has the constraint $Ax \le b$, then the optimal solution must satisfy $a_i^\top x - b_i \le 0$. Therefore, the right hand side in (3) can be bounded as:

$$\begin{aligned} q^\star \le & p^\star + \frac{1}{\mu} \sum_{i=1}^m \log\left(1 + e^{\mu(a_i^\top x^\star - b_i)}\right) \\ \le & p^\star + \frac{m \log(2)}{\mu} \end{aligned} \tag{4}$$

where the facts: $e^y \le 1$ for $y \le 0$; and $\log(2) \ge \log(1 + t)$ for $t \le 1$, have been used.

We then plug the minimizer $z^\star$ of the dual LP in the objective function of the Lagrange dual problem derived in point (b). Since the dual problem provides a lower bound on the optimal value of the primal, we can say:

$$q^\star \ge -b^\top z^\star - \frac{1}{\mu} \sum_{i=1}^m \left(z_i^\star \log z_i^\star + (1 - z_i^\star)\log(1 - z_i^\star)\right) \tag{5}$$

We can now use the fact that $0 \le z_i \le 1 \ \forall i = 1, .., m$ and that $u \log(u) \le 0$ on [0,1] to conclude that:

$$\begin{aligned} q^\star & \ge -b^\top z^\star - \frac{1}{\mu} \sum_{i=1}^m \left(z_i^\star \log z_i^\star + (1 - z_i^\star)\log(1 - z_i^\star)\right) \\ & \ge -b^\top z^\star \\ & \ge p^\star \end{aligned} \tag{6}$$

where the last inequality follows from strong duality for LP, i.e. $p^\star = -b^\top z^\star$. Equations (4) and (6) prove the statement.

2 Problem 2

Let us recall that we want to study the following optimization problem:

$$\begin{aligned} \min_{x_1,x_2} \quad & e^{-x_2}, \\ \text{s.t.} \quad & \sqrt{x_1^2+x_2^2} - x_1 \leq 0. \end{aligned} \tag{7}$$

(a) It is easy to see that the feasible set F is:

$$F = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\},$$

Thus the constraints define an affine set. Noting that the objective function is convex, it is then also shown that the problem is convex. Finally, since x_2=0 and the objective function does not depend on x_1, it follows that the optimal value of (7) is $p^\star = 1$.

(b) To compute the duality gap, we solve the dual problem of (7) and we find its optimum $d^\star$. By definition, the duality gap will be Δ=$p^\star$-$d^\star$.

The Lagrangian of the problem is

$$L(x_1, x_2, \lambda) = e^{-x_2} + \lambda\left(\sqrt{x_1^2+x_2^2} - x_1\right).$$

The associated dual function is

$$g(\lambda) = \inf_{x_1,x_2} L(x_1, x_2, \lambda), \tag{8}$$

and the dual problem is:

$$\begin{aligned} \max_{\lambda} \quad & g(\lambda), \\ \text{s.t.} \quad & \lambda \geq 0. \end{aligned}$$

Note that, for $\lambda \geq 0$, the dual function is always positive because L is given by the sum of two positive function, therefore $g(\lambda) \geq 0$. We can actually prove that the infimum in problem (8) is 0 regardless of λ. To see this, observe that for any $\epsilon > 0$, if we take $x_2 = -\log\epsilon$, $x_1 = \frac{x_2^2-\epsilon^2}{2\epsilon}$ we can rewrite the Lagrangian as:

$$\begin{aligned} L(x_1, x_2, \lambda) =& \epsilon + \lambda\left(\sqrt{\frac{(x_2^2-\epsilon^2)^2}{4\epsilon^2} + x_2^2} - \frac{x_2^2-\epsilon^2}{2\epsilon}\right), \\ =& \epsilon + \lambda\left(\sqrt{\frac{(x_2^2+\epsilon^2)^2}{4\epsilon^2}} - \frac{x_2^2-\epsilon^2}{2\epsilon}\right) = \epsilon + \lambda\epsilon = \epsilon(1+\lambda). \end{aligned}$$

Consequently, the infimum of the Lagrangian in (8) is 0, and thus the dual problem is trivially:

$$\max_{\lambda} g(\lambda) = \begin{aligned} \max_{\lambda} \quad & 0, \\ \text{s.t.} \quad & \lambda \geq 0, \end{aligned}$$

And the optimal dual is $d^\star = 0$. Therefore, the duality gap $\Delta = p^\star - d^\star$=1.

Alternatively, we can find the infimum of the Lagrangian by setting to zero its gradient (with respect to x_1 and x_2). Note that $\nabla_{x_1,x_2} L$ is a vector with two components, and *both* of them must be zero at the stationary point:

$$\frac{\partial L}{\partial x_1} = \frac{x_1}{\sqrt{x_1^2 + x_2^2}} - 1 = 0, \tag{9a}$$

$$\frac{\partial L}{\partial x_2} = \frac{x_2}{\sqrt{x_1^2 + x_2^2}} - e^{-x_2} = 0, \tag{9b}$$

These equations are both satisfied only if $x_1 \to +\infty$ and $x_2 \to +\infty$. Specifically, x_1 needs to go to infinity faster than x_2, so that in the limit $\frac{x_1}{\sqrt{x_1^2+x_2^2}} \to 1$ and $\frac{x_2}{\sqrt{x_1^2+x_2^2}} \to 0$. The infimum of the Lagrangian thus results to be 0 ($e^{-x_2} \to 0, \left(\sqrt{x_1^2 + x_2^2} - x_1\right) \to x_1 - x_1 = 0$). Note that setting $x_2 = -\log \epsilon$, $x_1 = \frac{x_2^2 - \epsilon^2}{2\epsilon}$ and taking $\epsilon \to 0$ is a way to achieve the desired limit behaviour for x_1 and x_2, and indeed satisfies Eq. (9).
In Example 3 of the Summary it was shown a similar strategy to compute the infimum of L and consisting of taking the limit to infinity of the optimization variables at different *speeds*.

(c) The existence of a nonzero duality gap could have not been concluded a priori. One could observe that, since the constraint in (7) always holds as an equality constraint, Slater's condition is not satisfied. However, this is only a sufficient condition (together with convexity) for strong duality and thus cannot be used to infer weak duality. In other words, one can only say that, since the Slater's condition is not satisfied, the duality gap might not be zero. Note that Example 4 of the Summary showed an example where constraint qualification did not hold (like here), but strong duality did hold, thus confirming that nothing can be said on strong duality in this circumstance.

3 Problem 3

Let us recall that we want to study the following optimization problem:

$$\begin{aligned} \min_{x,y} \quad & \|y\|_2 + \gamma \|x\|_1, \\ \text{s.t.} \quad & Ax - b = y, \end{aligned} \tag{10}$$

(a) The Lagrangian of (10) is by definition:

$$L(x, y, z) = \|y\|_2 + \gamma \|x\|_1 + z^\top (Ax - b - y).$$

We can rewrite the Lagrangian in the following form, which will be useful to later compute the dual function:

$$L(x, y, z) = -b^\top z + \left(\|y\|_2 - y^\top z\right) + \left(\gamma \|x\|_1 + z^\top Ax\right).$$

The dual function $g(z)$ is then:

$$g(z) = \inf_{x,y} L(x, y, z) = \begin{cases} -b^\top z, & \text{if } \|z\|_2 \leq 1, \|A^\top z\|_\infty \leq \gamma, \\ -\infty, & \text{else} \end{cases}$$

Where the two inequality constraints guarantee that there exist minimizers $\tilde{x}$ and $\tilde{y}$ of the Lagrangian such that the dual function $g(z)$ is bounded. In other words, if z satisfies the inequalities, it is guaranteed that:

$$\inf_y \left(\|y\|_2 - z^\top y \right) = \|\tilde{y}\|_2 - z^\top \tilde{y} = 0, \tag{11a}$$

$$\inf_x \left(\gamma \|x\|_1 + z^\top A x \right) = \gamma \|\tilde{x}\|_1 + z^\top A \tilde{x} = 0. \tag{11b}$$

The dual problem is then:

$$\begin{aligned} \max_z g(z) = \quad \max_z \quad & -b^\top z, \\ \text{s.t.} \quad & \|z\|_2 \leq 1, \\ & \|A^\top z\|_\infty \leq \gamma. \end{aligned}$$

(b) The KKT conditions for the problem are:

a-Primal feasibility: $Ax - b - y = 0$.

b-Dual feasibility: $\|z\|_2 \leq 1, \|A^\top z\|_\infty \leq \gamma$.

c-Stationarity of the Lagrangian: y satisfies Eq. ($11a$) and x satisfies Eq. ($11b$).

These conditions must be fulfilled by $x = x^\star$ and $y = Ax^\star - b$. Note first that, from the definition of r and KKT-a, it follows $r = \frac{Ax^\star - b}{\|Ax^\star - b\|_2} = \frac{y}{\|y\|_2}$. From KKT-c and Eq. (11a) it then holds $z = r$. The two sought relationships then follow from: KKT-b, which leads to $\|A^\top r\|_\infty \leq \gamma$; from KKT-c and Eq. (11b), which leads to $r^\top A x^\star + \gamma \|x^\star\|_1 = 0$.

4 Problem 4

By definition of conjugate function, it holds:

$$g^*(s) = \sup_x (s^\top x - g(x)),$$

Let us introduce $h(y) = f(y + c)$. Then, we can write $g(x) = h(Lx)$, and thus an equivalent expression for g^* is:

$$g^*(s) = \sup_x (s^\top x - h(Lx)) = -\inf_x (h(Lx) - s^\top x),$$

Define $l_s(x) = -s^\top x$. An equivalent expression for g^* is:

$$g^*(s) = -\inf_x (h(Lx) + l_s(x)), \tag{12}$$

We recognize that g^* is defined in (12) as a convex composite minimization problem. Therefore, we can apply Fenchel duality and consider the dual of (12). Moreover, constraint qualification is satisfied because of the assumption on g (i.e. relint dom $g \neq \emptyset$) and dom $l_s = \mathbb{R}^n$. By Fenchel strong duality it thus holds:

$$\begin{aligned} \inf_x (h(Lx) + l_s(x)) &= \sup_\mu (-h^*(\mu) - l_s^*(-L^\top \mu)), \\ &= -\inf_\mu (h^*(\mu) + l_s^*(-L^\top \mu)) \end{aligned}$$

Therefore, g^* can also be written as:

$$g^*(s) = -\inf_x (h(Lx) + l_s(x)) = \inf_\mu (h^*(\mu) + l_s^*(-L^\top \mu)), \tag{13}$$

To complete the proof, we need to explicitly write the conjugates of h and l_s.

$$
\begin{aligned}
h^*(\mu) =& \sup_y(\mu^\top y - h(y)) = \sup_y(\mu^\top x - f(y+c)) = \sup_y(\mu^\top(v-c) - f(v)),\\
=& \sup_y(\mu^\top v - f(v)) - \mu^\top c = f^*(\mu) - \mu^\top c
\end{aligned}
$$

where the change of variables $v = y + c$ was used. As for l_s, it holds:

$$
\begin{aligned}
l_s^*(\nu) =& \sup_y(\nu^\top y + s^\top y) = \sup_y((\nu+s)^\top y),\\
=& \mathcal{I}_0(\nu + s)
\end{aligned}
$$

where the indicator function $\mathcal{I}_0(\nu + s)$ is defined as:

$$
\mathcal{I}_0(\nu + s) = \begin{cases} 0, & \text{if } \nu + s = 0 \\ \infty, & \text{else} \end{cases}
$$

By substituting the conjugates in (13) one obtains:

$$
g^*(s) = \inf_\mu(f^*(\mu) - \mu^\top c + \mathcal{I}_0(-L^\top\mu + s)) = \inf_\mu(f^*(\mu) - \mu^\top c : s = L^\top\mu),
$$

which completes the proof.