EECS 127/227AT **Fall 2020**
Discussion 6

1 The flexibility of duality

In class and in our two textbooks the emphasis when Lagrange duality is introduced is on dualizing the indicator functions $I_{\mathbb{R}_-}$ and $I_{\{0\}}$ (each of which is a convex and lower semicontinuous function). However, the general scheme behind Lagrange duality is much broader and one can dualize by expressing any lower semicontinuous convex function in terms of its Fenchel conjugate. This flexibility is very important in creating more insightful and useful duals.

In this question we will illustrate this flexibility of duality in a toy problem. In the next question in this discussion set we will see a more interesting example of the use of this flexibility. Recall from Question 3 of Discussion 5 that we know that that $I^{**}_{\mathbb{R}_-} = I_{\mathbb{R}_-}$ and $I^{**}_{\{0\}} = I_{\{0\}}$, because $\mathbb{R}_- := \{x \in \mathbb{R} : x \leq 0\}$ and $\{0\}$ are closed convex subsets of the real line.

(a) Verify that the Fenchel conjugate of $I_{\mathbb{R}_-}$ is $I_{\mathbb{R}_+}$. Conclude that we can write

$$I_{\mathbb{R}_-}(x) = \sup_{z \geq 0} \; zx. \tag{1}$$

(It is of course directly obvious, without going through Fenchel conjugates, that the equality claimed in equation (1) is true. However, we aim to make a much broader point, for which it is useful to take a less direct route to this equality.)

(b) Verify that the Fenchel conjugate of $I_{\{0\}}$ is 0. Conclude that we can write

$$I_{\{0\}}(x) = \sup_{z \in \mathbb{R}} \; zx. \tag{2}$$

(Again, it is directly obvious, without going through Fenchel conjugates, that the equality claimed in equation (2) is true, but we are deliberately showing this in a more complicated way.)

What is going on in traditional Lagrange duality is that we recognize that the constrained convex optimization problem:

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{subject to:} \quad & f_i(x) \leq 0, i = 1, \ldots, m, \\ & h_i(x) = 0, i = 1, \ldots, p, \end{aligned}$$

is the same as the unconstrained optimization problem

$$\min_x f_0(x) + \sum_{i=1}^{m} I_{\{f_i(x) \leq 0\}} + \sum_{j=1}^{p} I_{\{h_j(x)=0\}},$$

namely

$$\min_x f_0(x) + \sum_{i=1}^{m} I_{\mathbb{R}_-}(f_i(x)) + \sum_{j=1}^{p} I_{\{0\}}(h_j(x)),$$

and then we use equation (1) to replace each term of the form $I_{\mathbb{R}_-}(f_i(x))$ (the corresponding dual variable is denoted λ_i and must satisfy $\lambda_i \geq 0$), and equation (2) to replace each term of the form $I_{\{0\}}(h_j(x))$ (the corresponding dual variable is denoted ν_j and is unconstrained). This gives us a min-max problem (min over the primal variables and max over the dual variables) which we dualize to a max-min problem, as discussed in class and in the two textbooks. In the max-min problem, after we take the min (over the primal variables) we get a function of the dual variables (which is what we call the dual objective function) and the problem of maximizing the dual objective function subject to the constraints on the dual variables is what is called the dual problem (as we saw, strictly speaking we have to consider the problem of minimizing the negative of the dual objective function to get a convex optimization problem)

However, we can play the same game in a more sophisticated way to create many other dual problems for a given primal problem, as we will now illustrate in a toy problem (see also the next question in this discussion set).

First, let us note that in both the general convex optimization problem and its unconstrained version above, it is implicitly assumed that the minimization is over x in the intersection of the domains of all the functions $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$. This is important to keep in mind for what follows.

(c) Consider the following convex optimization problem

$$\min_x \ \frac{1}{2}x^2 - \log x.$$

Here the logarithm is to the natural base, the function $\frac{1}{2}x^2$ has domain $\mathbb{R}$ and the function $-\log x$ has domain $\mathbb{R}_{++}$. Also note that it is implicitly assumed that the minimization is over the set of x that lie in the domains of all the functions involved, which in this case is $\mathbb{R}_{++}$.

Find the optimal value of the problem and an optimal point.

(d) Dualize $-\log x$ to write the convex optimization problem of the preceding part of the question as a min-max problem. For this, recall that in Question 2(a) of Discussion 5 you have shown that the Fenchel conjugate of $\phi(x) := -\log x$, with $\mathrm{dom}(\phi) = \mathbb{R}_{++}$, is given by

$$\phi^*(z) = \begin{cases} -1 - \log|z| & \text{if } z < 0, \\ \infty & \text{otherwise.} \end{cases}$$

Further, note that ϕ is a lower semicontinuous convex function, so it equals the Fenchel conjugate of its Fenchel conjugate.

(e) Dualize the min-max problem found in the preceding part of the question to a max-min problem, find the corresponding dual objective and state the corresponding dual problem.

(f) Find the optimal value of the dual problem found in the preceding part of this question and also an optimal point for this optimization problem. Verify that strong duality holds.

(g) **(Optional)**

Returning to the original primal optimization problem in part (c) of this question, dualize the component $\frac{1}{2}x^2$ in the objective to get a different min-max formulation of the primal problem, find the corresponding max-min problem, the corresponding dual objective function, and the corresponding dual optimal point and dual optimal value. Show that strong duality once again holds.

2 Least absolute deviations with ℓ_∞ regularization

Let $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$ and $\mu > 0$. Consider the problem

$$p^* = \min_x \ \|Ax - y\|_1 + \mu\|x\|_\infty.$$

In the *least absolute deviations* problem we aim to find the vector of regression coefficients to minimize the sum of the absolute values of the residuals, i.e. given A and y, we seek to find x to minimize $\|Ax - y\|_1$. This is considered a more robust criterion than the least squares criterion because it pays less attention to outliers in the data (the least squares criterion, by squaring large residuals, tends to overemphasize outliers). Here we are considering an ℓ_∞ regularization of the least absolute deviations problem, by introducing a penalty on the ℓ_∞ norm of the vector of regression coefficient, scaled by the hyperparameter μ. Such ℓ_∞ regularization is quite unusual, compared to ℓ_1 or ℓ_2 regularization. It does not promote sparsity, but tends to make all the regression coefficients of roughly the same size, so that many of them will be nonzero but they will all be small.

For $j \in \{1, \ldots, n\}$, denote by $a_j \in \mathbb{R}^m$ the j-th column of A, so that $A = [a_1, \ldots, a_n]$, and define

$$\|A\|_1 := \sum_{j=1}^n \|a_j\|_1.$$

You should check for yourself that this notation makes sense, in that $\|A\|_1$, as defined above, is the induced ℓ_1 norm of the matrix A, namely

$$\|A\|_1 = \max_{x:\|x\|_1=1} \ \|Ax\|_1$$

(a) Express the problem as an LP.

(b) Show that a dual to the problem can be written as

$$\begin{array}{ll} d^* = \max\limits_u & -u^T y \\ \text{subject to:} & \|u\|_\infty \le 1, \\ & \|A^T u\|_1 \le \mu. \end{array}$$

Hint: Use the fact that for any vector z we have

$$\max_{u\,:\,\|u\|_1\le 1} \ u^T z = \|z\|_\infty, \quad \max_{u\,:\,\|u\|_\infty\le 1} \ u^T z = \|z\|_1.$$

You should check for yourself that these formulas are correct. They correspond to the fact that the ℓ_1 and the ℓ_∞ norms are duals of each other.

Remark: When there are many conditions, as in this problem, it may be more convenient to write optimization problems with the conditions written in line. Thus here we aim to show that the dual problem can be written as

$$d^* := \max_u \ -u^T y \ : \ \|u\|_\infty \le 1, \ \|A^T u\|_1 \le \mu.$$

(c) It can be shown that strong duality holds, i.e. $d^* = p^*$. Assuming this, show that the condition "$\|A^T u\|_1 < \mu$ for every u with $\|u\|_\infty \le 1$" ensures that $x = 0$ is optimal.

(d) Assuming once again that strong duality holds, show that the condition in the previous part of this question holds if $\mu > \|A\|_1$.

Remark: The point being made in this result is that if the hyperparameter is chosen to be too big, i.e. if we penalize the ℓ_∞ norm of the vector of regression parameters too much, then the solution to the ℓ_∞ regularized least absolute deviations problem will end up simply ignoring the data.

3 KKT conditions

Consider the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & (x_1 - 1)^2 + (x_2 - 1)^2 \leq 2, \\ & (x_1 - 1)^2 + (x_2 + 1)^2 \leq 2, \end{aligned}$$

where $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top \in \mathbb{R}^2$.

(a) Sketch the feasible region and the level sets of the objective function.

(b) Is this a convex optimization problem?

(c) Write the Lagrangian for this problem in the traditional framework of convex duality.

(d) Using Slater's condition, verify that strong duality holds in this problem in the context of the traditional Lagrangian duality.

(e) Write the KKT conditions for this optimization problem.

(f) Prove from the KKT conditions that $(x_1^*, x_2^*) = (0, 0)$ is a primal optimal point and $(\lambda_1^*, \lambda_2^*) = (0, 0)$ is a dual optimal point. Use this to find the primal optimal value p^*.

(g) Find all the solutions of the KKT equations.