Fenchel duality

Last time we began by showing that if we consider the unconstrained problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \ f(\boldsymbol{x}) + g(\boldsymbol{x}) \tag{1}$$

where f and g are both convex, we can derive the equivalent dual problem

$$\underset{\boldsymbol{\nu}}{\text{maximize}} -f^*(\boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}). \tag{2}$$

Recall from our first discussion of Lagrange duality that the dual problem provides a lower bound for the primal problem, or in the language of the problems above, we have

$$\inf_{\boldsymbol{x}} \ f(\boldsymbol{x}) + g(\boldsymbol{x}) \geq \sup_{\boldsymbol{\nu}} -f^*(\boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}).$$

Moreover, under certain conditions we have *strong duality*. In this setting, strong duality implies that the above inequality will hold with equality, i.e.,

$$\inf_{\boldsymbol{x}} \ f(\boldsymbol{x}) + g(\boldsymbol{x}) = \sup_{\boldsymbol{\nu}} -f^*(\boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}). \tag{3}$$

Fenchel's Duality Theorem tells us that under certain regularity assumptions on f and g, we have strong duality and (3) holds.[1] Specifically, if $\mathcal{D} = \operatorname{dom} f$ and $\mathcal{C}$ denotes the set of $\boldsymbol{x} \in \mathbb{R}^N$ where g is finite and continuous, then (3) holds whenever there exists an $\bar{\boldsymbol{x}} \in \mathcal{D} \cap \mathcal{C}$.

[1] In our discussion here as well as when reviewing Lagrangian duality, we have assumed that $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{x})$ is finite (so that these quantities are even defined). If the primal (or dual) is not bounded, then there is no solution to the optimization problem and strong duality will not hold.

Note that if $g(\boldsymbol{x})$ is the indicator for a convex set $\mathcal{C}$, then this is equivalent to a constrained optimization problem, and the conditions above are equivalent to the assumption that there is a strictly feasible point in $\operatorname{dom} f$, i.e., Slater's condition. In this case we can also write (3) more cleaning if we define a new function $h'_{\mathcal{C}}(\boldsymbol{\nu})$ which is related to the support function of $\mathcal{C}$, just with an infimum instead of a supremum:

$$h'_{\mathcal{C}}(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{\nu} \rangle = -\sup_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, -\boldsymbol{\nu} \rangle = -h_{\mathcal{C}}(-\boldsymbol{\nu}).$$

With this notation we can re-write (3) for the case of standard constrained optimization as

$$\inf_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x}) = \sup_{\boldsymbol{\nu}} h'_{\mathcal{C}}(\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}). \tag{4}$$

We will not do so here, but from this point you can actually show that if our constraints match the form that is assumed in our discussion of Lagrangian duality, then the right-hand side of (4) exactly corresponds to the Lagrangian dual problem. In this sense Lagrangian duality is just a special case of Fenchel duality.

Super-Easy Example

Before we look at serious applications of Fenchel duality, let's look at a very simple example just to get a feel for the computations involved. We will compute

$$\inf_{x\in[3,5]} x^3.$$

Of course, we know the answer already: it is 27, as the function above achieves its minimum value at $\widehat{x} = 3$. But let's verify the Fenchel duality theorem for this case.

We will take $f(x) = x^3$ which is convex over the non-negative reals, so we take $\mathcal{D} = \{x : x \geq 0\}$. The constraint set is the interval $\mathcal{C} = [3, 5]$. First we compute

$$h'_{\mathcal{C}}(\nu) = \inf_{x \in [3,5]} \; \nu x \;=\; \begin{cases} 3\nu, & \nu \geq 0, \\ 5\nu, & \nu < 0. \end{cases}$$

The conjugate of f is

$$f^*(\nu) = \sup_{x \geq 0} \; (\nu x - x^3).$$

For fixed $\nu \geq 0$, this expression is maximized at $x^\star = \sqrt{\nu/3}$; for $\nu < 0$ it is maximized at $x^\star = 0$. Thus

$$f^*(\nu) = \begin{cases} \frac{2}{3}\sqrt{\frac{\nu^3}{3}}, & \nu \geq 0, \\ 0, & \nu < 0. \end{cases}$$

Thus

$$\max_{\nu \in \mathbb{R}} \left[h'_{\mathcal{C}}(\nu) - f^*(\nu)\right] \;=\; \max_{\nu \in \mathbb{R}} \begin{cases} 3\nu - \frac{2}{3}\sqrt{\frac{\nu^3}{3}}, & \nu \geq 0, \\ 5\nu, & \nu < 0. \end{cases}$$

It is easy to check that this expression is maximized at $\nu^\star = 27$ (coincidence), and that

$$\left. \left(3\nu - \frac{2}{3}\sqrt{\frac{\nu^3}{3}} \right) \right|_{\nu = 27} = 27.$$

Example: Resource allocation [Lue69]

The "law of diminishing returns" is a fundamental tenet of economics: as we put more and more resources into something, at some point, the incremental gains become less and less. You see this everywhere: what is the difference between spending \$5 on dinner, \$50 on dinner, \$500 on dinner? What are the differences between a \$50 bicycle, a \$500 bicycle, and a \$5000 bicycle?

What this means is that functions $f(x)$ that map resources to return are concave.

Suppose we have D dollars that we would like to allocate to N different activities in such a way that maximizes the return. The return of each activity is a (possibly different) concave function $f_n(x_n)$, where x_n is the amount of money invested. Our optimization problem is

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{maximize}} \;\; f(\boldsymbol{x}) = \sum_{n=1}^{N} f_n(x_n) \quad \text{subject to} \quad \begin{array}{l} \displaystyle\sum_{n=1}^{N} x_n = D \\ \boldsymbol{x} \geq \boldsymbol{0}, \end{array}$$

or equivalently

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; \tilde{f}(\boldsymbol{x}) = -f(\boldsymbol{x}) \quad \text{subject to} \quad \begin{array}{l} \displaystyle\sum_{n=1}^{N} x_n = D \\ \boldsymbol{x} \geq \boldsymbol{0}, \end{array}$$

This is a convex optimization problem in N variables, and of course its solution depends on what we actually choose for the return functions $f_n(x_n)$. However, by using Fenchel duality, we can recast this problem as an optimization in a single variable.

Since the natural domain of the f_n is $x \geq 0$, let's take

$$\mathcal{D} = \{\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\}, \quad \mathcal{C} = \{\boldsymbol{x} : x_1 + \cdots + x_N = D\}.$$

We start by computing

$$h'_{\mathcal{C}}(\boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{C}} \ \langle \boldsymbol{x}, \boldsymbol{\nu} \rangle.$$

Since $\mathcal{C}$ is itself an affine set, $h'_{\mathcal{C}}(\boldsymbol{\nu})$ is unbounded below for almost every $\boldsymbol{\nu}$ we plug in – the exception is if all of the entries of $\boldsymbol{\nu}$ are equal to one another. In this case,

$$\boldsymbol{\nu} = \lambda \mathbf{1}, \quad h_{\mathcal{C}}(\boldsymbol{\nu}) = D\lambda,$$

where $\mathbf{1}$ is an N-vector of all ones. Thus, we have

$$h'_{\mathcal{C}}(\boldsymbol{\nu}) = \begin{cases} D\mathbf{1}^{\mathrm{T}}\boldsymbol{\nu}, & \boldsymbol{\nu} \in \mathrm{Range}(\mathbf{1}) \\ -\infty, & \text{otherwise.} \end{cases}$$

Now we compute the conjugate $\tilde{f}^*(\boldsymbol{\nu})$ of $\tilde{f}(\boldsymbol{x}) = -f(\boldsymbol{x})$. Since $\tilde{f}$ is a sum of convex functions of independent variables,

$$\tilde{f}^*(\boldsymbol{\nu}) = \sum_{n=1}^{N} \tilde{f}_n^*(\nu_n),$$

where $\tilde{f}_n^*(\nu_n)$ is the conjugate of a function of a single variable:

$$\tilde{f}_n^*(\nu_n) = \sup_{x \geq 0} \left[\nu_n x - \tilde{f}_n(x) \right].$$

This means we can write the dual as

$$\max_{\boldsymbol{\nu}} \left[h'_{\mathcal{C}}(\boldsymbol{\nu}) - \tilde{f}^*(\boldsymbol{\nu}) \right] \ = \ \max_{\lambda \in \mathbb{R}} \left[D\lambda - \sum_{n=1}^{N} \tilde{f}_n^*(\lambda) \right].$$

That is, the expression to be minimized is a function **of a single variable** λ. All we need to know how to do is evaluate the conjugate functions $\tilde{f}_n^*$.

Example: Norm minimization

Here we look at an example of how we can apply Fenchel duality to provide an alternative characterization of a common constrained optimization program. Consider the "norm minimization" problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{x}\| \quad \text{subject to} \ \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y},$$

where the vector $\boldsymbol{y} \in \mathbb{R}^M$ and matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ are given and we assume that rank$(\boldsymbol{A})$ is at most N so that we are guaranteed that there is at least one feasible solution. Here the norm in the objective function is an arbitrary norm. We will derive the dual for the general case, which could then be specialized to tell us something about least squares (for the ℓ_2 norm), "Basis Pursuit" (for the ℓ_1 norm), or the result for any choice of norm.

We have already calculated $f^*(\boldsymbol{\nu})$ for the case where $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$. In this case $f^*(\boldsymbol{\nu})$ is the indicator function for the set $\{\boldsymbol{\nu} : \|\boldsymbol{\nu}\|_* \leq 1\}$. If we plug this into (4) we see that the objective function will be $-\infty$ unless $\|\boldsymbol{\nu}\|_* \leq 1$, and thus we can equivalently write the dual problem as

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \ h'_{\mathcal{C}}(\boldsymbol{\nu}) \quad \text{subject to} \ \ \|\boldsymbol{\nu}\|_* \leq 1,$$

where $\mathcal{C} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$. This, it remains to calculate $h'_{\mathcal{C}}(\boldsymbol{\nu})$.

We first note that if $\langle \boldsymbol{u}, \boldsymbol{\nu} \rangle \neq 0$ for some $\boldsymbol{u} \in \text{Null}(\boldsymbol{A})$, then $h'_{\mathcal{C}}(\boldsymbol{\nu}) = -\infty$. To see this, note that if $\boldsymbol{u} \in \text{Null}(A)$ then for any $\boldsymbol{x} \in \mathcal{C}$ and $t \in \mathbb{R}$, $\boldsymbol{A}(\boldsymbol{x} + t\boldsymbol{u}) = \boldsymbol{y}$. Thus, $\boldsymbol{x} + t\boldsymbol{u} \in \mathcal{C}$, and $\langle \boldsymbol{x} + t\boldsymbol{u}, \boldsymbol{\nu} \rangle = \langle \boldsymbol{x}, \boldsymbol{\nu} \rangle + t\langle \boldsymbol{u}, \boldsymbol{\nu} \rangle$, which is unbounded since t can be arbitrary.

What remains is to calculate $h'_{\mathcal{C}}(\boldsymbol{\nu})$ for $\boldsymbol{\nu}$ that are orthogonal to $\text{Null}(\boldsymbol{A})$. Recall that this is equivalent to the assumption that $\boldsymbol{\nu} \in \text{Range}(\boldsymbol{A}^{\mathrm{T}})$. For any such $\boldsymbol{\nu}$ we can write $\boldsymbol{\nu} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$ for some $\boldsymbol{w} \in \mathbb{R}^M$, in which case

$$\begin{aligned}\inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{\nu} \rangle &= \inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \rangle \\ &= \inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{w} \rangle \\ &= \inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{y}, \boldsymbol{w} \rangle \\ &= \langle \boldsymbol{y}, \boldsymbol{w} \rangle.\end{aligned}$$

Thus, if we replace $\boldsymbol{\nu}$ in our dual problem with $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$ and optimize over $\boldsymbol{w}$ instead, we arrive at the dual problem

$$\underset{\boldsymbol{w}}{\text{maximize}} \ \langle \boldsymbol{y}, \boldsymbol{w} \rangle \quad \text{subject to} \ \ \|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}\|_* \leq 1.$$

As an example, if we consider the ℓ_1-norm minimization problem (also known as "Basis Pursuit") where $\|\cdot\| = \|\cdot\|_1$, the dual becomes

$$\underset{\boldsymbol{w}}{\text{maximize}} \ \langle \boldsymbol{y}, \boldsymbol{w} \rangle \quad \text{subject to} \ \ \|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}\|_\infty \leq 1.$$

Note that this is a standard linear program. This can be a useful observation from a computational perspective, but later in the course we will show how Fenchel duality for this problem can also be used to provide a theoretical characterization of the properties (e.g., sparsity) of the solution $\boldsymbol{x}^\star$ of the primal problem.

References

[Lue69] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, 1969.