# Duality for convex optimization

Duality in constrained convex optimization[1] means that we can recast the "primal" problem

$$\underset{\boldsymbol{x}\in\mathcal{C}\cap\mathcal{D}}{\text{minimize}} \;\; f(\boldsymbol{x}) \tag{1}$$

as a search over hyperplanes. We have already seen in the context of non-smooth constrained optimization that a necessary and sufficient condition for $\boldsymbol{x}^\star$ to solve (1) is that

$$\mathbf{0} \in \partial f(\boldsymbol{x}^\star) + \mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^\star),$$

where $\mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^\star)$ is the **normal cone** of $\mathcal{C}$ at $\boldsymbol{x}^\star$. That is, there is a $\boldsymbol{\nu}^\star$ such that

$$\boldsymbol{\nu}^\star \in \partial f(\boldsymbol{x}^\star) \quad \text{and} \quad -\boldsymbol{\nu}^\star \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^\star).$$

When $\boldsymbol{x}^\star$ is on the boundary of $\mathcal{C}$, this means that $\boldsymbol{\nu}^\star$ defines a supporting hyperplane for both the sublevel set

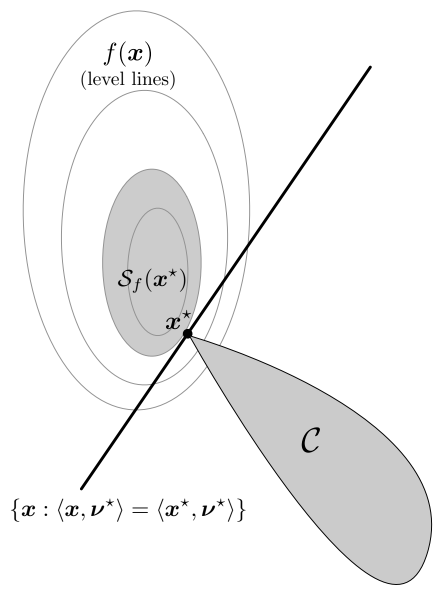
$$\mathcal{S}_f(\boldsymbol{x}^\star) = \{\boldsymbol{x} \,:\, f(\boldsymbol{x}) \le f(\boldsymbol{x}^\star)\}$$

and the constraint set $\mathcal{C}$ and it **separates** $\mathcal{S}_f(\boldsymbol{x}^\star)$ from $\mathcal{C}$, i.e.,

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{\nu}^\star\rangle &\le \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star\rangle, \quad \text{for all } \boldsymbol{x} \in \mathcal{S}_f(\boldsymbol{x}^\star),\\ \langle \boldsymbol{x}, \boldsymbol{\nu}^\star\rangle &\ge \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star\rangle, \quad \text{for all } \boldsymbol{x} \in \mathcal{C}. \end{aligned}$$

[1] As usual, we assume that $\mathcal{C}$ is a closed convex set. We are also making the domain $\mathcal{D}$ of $f$ explicit here, just as we did when we discussed the Fenchel conjugate, because it will be useful to do so in some of our examples below. We are arranging things so that $\mathcal{D}$ does not effectively add additional constraints; the unconstrained minimizer should be in the interior of $\mathcal{D}$. We will also assume that $f$ is closed (meaning that it has closed sublevel sets).

Here is a picture:




A proof of this follows quickly from the definitions of the subdifferential and the normal cone:

1. Since $\boldsymbol{\nu}^\star \in \partial f(\boldsymbol{x}^\star)$, we have $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^\star) + \langle \boldsymbol{x} - \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle$. So if $f(\boldsymbol{x}) \leq f(\boldsymbol{x}^\star)$ it must be that $\langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \leq \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle$.
2. Since $-\boldsymbol{\nu}^\star \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^\star)$, we know that for all $\boldsymbol{x} \in \mathcal{C}$ we have $\langle \boldsymbol{x} - \boldsymbol{x}^\star, -\boldsymbol{\nu}^\star \rangle \leq 0$ which is the same as $\langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \geq \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle$.

The **Fenchel dual** of (1) searches for this supporting/separating hyperplane directly. It is given by

$$\underset{\boldsymbol{\nu} \in \mathcal{C}^* \cap \mathcal{D}^*}{\text{maximize}} \; d(\boldsymbol{\nu}) := -f^*(\boldsymbol{\nu}) + h'_{\mathcal{C}}(\boldsymbol{\nu}), \tag{2}$$

where $h'_{\mathcal{C}}(\boldsymbol{\nu})$ is related to the support function. Specifically,

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

where $\mathcal{C}^*$ is the domain of $h_{\mathcal{C}}$ and $\mathcal{D}^*$ is the domain of the Fenchel conjugate $f^*$.

We quickly see that

$$d(\boldsymbol{\nu}) \leq f(\boldsymbol{x}) \quad \text{for all } \boldsymbol{\nu} \in \mathcal{C}^* \cap \mathcal{D}^* \text{ and } \boldsymbol{x} \in \mathcal{C} \cap \mathcal{D},$$

as for every $\boldsymbol{\nu}, \boldsymbol{x}$ we have

$$-f^*(\boldsymbol{\nu}) = -\sup_{\boldsymbol{x}' \in \mathcal{D}} \left( \langle \boldsymbol{x}', \boldsymbol{\nu} \rangle - f(\boldsymbol{x}') \right) \leq -\langle \boldsymbol{x}, \boldsymbol{\nu} \rangle + f(\boldsymbol{x}),$$

and

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

What do we have if we happen to find a $\boldsymbol{\nu}^\star$ such that $d(\boldsymbol{\nu}^\star) = f(\boldsymbol{x}^\star)$ where $\boldsymbol{x}^\star$ is a solution to the primal (1)? In this case,

$$\begin{aligned} f(\boldsymbol{x}^\star) &= -f^*(\boldsymbol{\nu}^\star) + h'_{\mathcal{C}}(\boldsymbol{\nu}^\star) \\ &= \inf_{\boldsymbol{x} \in \mathcal{D}} \left( f(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \right) + \inf_{\boldsymbol{x} \in \mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \\ &\leq f(\boldsymbol{x}^\star) - \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle + \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle \\ &= f(\boldsymbol{x}^\star). \end{aligned}$$

Since we ended up with the same thing we started with, the inequality above must be an equality, and we know that

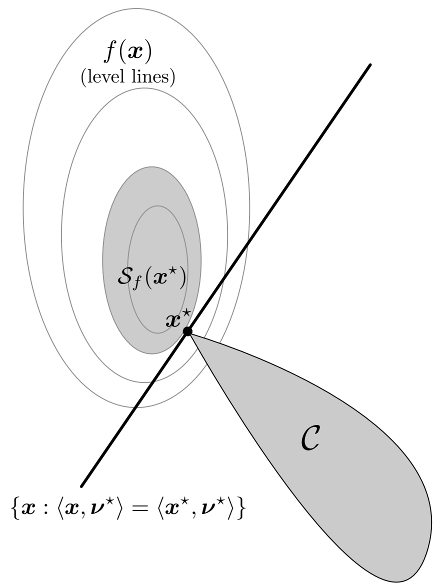
$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathcal{D}} \left( f(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \right) \quad \Rightarrow \quad \boldsymbol{\nu}^\star \in \partial f(\boldsymbol{x}^\star), \tag{3}$$

and

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \quad \Rightarrow \quad -\boldsymbol{\nu}^\star \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^\star), \tag{4}$$

since $\langle \boldsymbol{x}, \boldsymbol{\nu}^\star \rangle \geq \langle \boldsymbol{x}^\star, \boldsymbol{\nu}^\star \rangle$ for all $\boldsymbol{x} \in \mathcal{C}$. So this $\boldsymbol{\nu}^\star$ does indeed define a supporting hyperplane for both $\mathcal{C}$ and the sublevel set $\mathcal{S}_f(\boldsymbol{x}^\star)$ while also separating these sets.

Once again:




The **Fenchel duality theorem** states that there exists such $\boldsymbol{\nu}^\star$ whenever the relative interior of $\mathcal{C} \cap \mathcal{D}$ is nonempty ($\text{relint}(\mathcal{C} \cap \mathcal{D}) \neq \emptyset$). Note that this is the same as Slater's condition, which told us when we had strong Lagrangian duality. (In fact, you can show that Lagrangian duality is a special case of Fenchel duality: if $\mathcal{C}$ is an intersection of sublevel sets of differentiable functions and if $f$ is also differentiable, then the Fenchel dual becomes the Largrange dual.) We will forgo a proof of this theorem here, but we will discuss it in a more general form in the next section.

Finally, note that if we have strong duality and a dual solution $\boldsymbol{\nu}^\star$, we can recover the primal solution via the unconstrained optimization problem in (3) or minimizing a linear function over $\mathcal{C}$ as in (4).

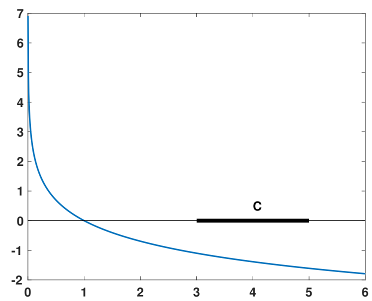
**Moral:** *Optimization of a convex function $f(\boldsymbol{x})$ over a closed convex set $\mathcal{C}$ can be thought of as a search for a hyperplane the simultaneously supports $\mathcal{C}$, supports a sublevel set of $f$, and separates these sets.*

## Fenchel duality examples

### Easy 1D 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

$$\underset{x\in[3,5]}{\text{minimize}} \ -\log(x).$$




Of course, we know the answer already: it is $-\log 5 \approx -1.6094$, as the function above achieves its minimum value at $x^\star = 5$. But let's look at what Fenchel duality says anyway.

The natural domain for $f(x) = -\log x$ is the positive reals, so we take $\mathcal{D} = \{x : x > 0\}$. The constraint set is the interval $\mathcal{C} = [3, 5]$. First we compute

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

So the domain of $h'_{\mathcal{C}}$ is $\mathcal{C}^* = \mathbb{R}$.

The Fenchel conjugate of $f$ is

$$f^*(\nu) = \sup_{x>0} \ (\nu x + \log x).$$

For $\nu \geq 0$, $\nu x + \log x$ diverges as $x \to \infty$. So the domain of the conjugate is $\mathcal{D}^* = \{\nu \ : \ \nu < 0\}$. For $\nu < 0$, $\nu x + \log x$ is maximized at $x_\nu = -1/\nu$ where it takes the value $-1 - \log(-\nu)$. Thus

$$f^*(\nu) = \begin{cases} 1 - \log(-\nu), & \nu < 0 \\ \infty & \nu \geq 0. \end{cases}$$

Combing these we have the Fenchel dual function on $\mathcal{C}^* \cap \mathcal{D}^*$,

$$d(\nu) = h'_{\mathcal{C}}(\nu) - f^*(\nu) = \begin{cases} 5\nu + \log(-\nu) + 1 & \nu < 0 \\ -\infty & \nu \geq 0. \end{cases}$$

The solution (maximizer) of the dual is at

$$\nu^\star = \frac{-1}{5}, \quad d(\nu^\star) = \log(1/5) = -\log(5).$$

Because strong duality holds, (3) tells us that $x^\star \in \arg\min_{x \in \mathcal{D}} f(x) - x\nu^\star$. This is exactly the optimization we solved to find $x_{\nu^\star}$ so we know

$$x^\star = x_{\nu^\star} = 5,$$

which is what we expected. We can check that

$$f(x^\star) = -\log 5 = d(\nu^\star).$$

**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 $h(x)$ 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 $h_n(x_n)$, where $x_n$ is the amount of money invested. Our optimization problem is a maximization of a concave function:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{maximize}} \;\; h(\boldsymbol{x}) = \sum_{n=1}^{N} h_n(x_n) \quad \text{subject to} \quad \begin{aligned} \sum_{n=1}^{N} x_n &= D \\ \boldsymbol{x} &\geq \boldsymbol{0}. \end{aligned}$$

One example, one which we will look at on the homework, is to take $h_n(x) = x/(s_n + x)$ for some constants $s_n > 0$. The natural domain for these functions is $x \geq 0$, so in this case the natural domain for the function we are optimizing is $\mathcal{D} = \{\boldsymbol{x} \,:\, \boldsymbol{x} \geq \boldsymbol{0}\}$.

The equivalent convex minimization problem is

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

where

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} f_n(x_n) = -h(\boldsymbol{x}), \quad f_n(x_n) = -h_n(x_n).$$

This is a convex optimization problem in $N$ variables, and of course its solution depends on what we actually choose for the return functions $h_n(x_n)$. 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

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

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

$$\mathcal{C}^* = \{\boldsymbol{\nu} \ : \ \boldsymbol{\nu} = \lambda \mathbf{1} \text{ for some } \lambda \in \mathbb{R}\}.$$

and we have

$$h'_{\mathcal{C}}(\boldsymbol{\nu}) = \begin{cases} D\lambda, & \boldsymbol{\nu} = \lambda \mathbf{1} \\ -\infty, & \text{otherwise.} \end{cases}$$

The function $f(\boldsymbol{x})$ is a separable sum of functions of the individual $x_n$, so the Fenchel conjugate can be written this way as well (recall this property of conjugate functions from earlier in the notes):

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

where $f_n^*(\nu_n)$ is the conjugate of a function of a single variable:

$$f_n^*(\nu_n) = \sup_{x \geq 0} \left( \nu_n x - f_n(x) \right).$$

This means we can write the dual problem as[2]

$$\underset{\boldsymbol{\nu}\in\mathcal{C}^*}{\text{maximize}}\ h'_{\mathcal{C}}(\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}) \ \Rightarrow\ \underset{\lambda\in\mathbb{R}}{\text{maximize}}\ D\lambda - \sum_{n=1}^{N} f_n^*(\lambda)$$

That is, the expression to be minimized is a function **of a single variable** $\lambda$. All we need to know how to do is evaluate the conjugate functions $f_n^*$.

As always, we can recover the primal solution from the dual solution. in this case, once we compute the optimal $\lambda$ above, we can take $x_n^\star$ as the maximizer of $\lambda x - f_n(x)$.

## Minimum norm solution to a system of linear equations.

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:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}}\ \|\boldsymbol{x}\| \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y},$$

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 $\text{rank}(\boldsymbol{A}) = M$ so that we are guaranteed that there is at least one feasible solution no matter what $\boldsymbol{y}$ is.

We will take the norm in the objective function as an arbitrary norm and derive the dual for the general case, which can then be specialized to tell us something about problems like least squares (for the $\ell_2$

[2]The domain $\mathcal{D}^*$ will depend on what the $f_n$ actually are. But honestly, it is not entirely necessary to restrict the optimization to $\mathcal{C}^*\cap\mathcal{D}^*$ as $d(\boldsymbol{\nu}) = -\infty$ outsize this set anyway.

norm), "basis pursuit" (for the $\ell_1$ norm), or any other norm of your choosing.

We have already calculated $f^*(\boldsymbol{\nu})$ for the case where $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ in our section on the Fenchel conjugate above. We saw that $\mathcal{D}^* = \{\boldsymbol{\nu} : \|\boldsymbol{\nu}\|_* \leq 1\}$ and $f^*$ was the indicator function for the unit ball of the dual norm,

$$f^*(\boldsymbol{\nu}) = \iota_{\mathcal{B}_{\|\cdot\|_*}}(\boldsymbol{\nu}) = \begin{cases} 0 & \|\boldsymbol{\nu}\|_* \leq 1, \\ \infty & \|\boldsymbol{\nu}\|_* > 1. \end{cases}$$

We now compute the support function $h'_{\mathcal{C}}(\boldsymbol{\nu})$ for

$$\mathcal{C} = \{\boldsymbol{x} \ : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}.$$

Note that if $\langle \boldsymbol{u}, \boldsymbol{\nu} \rangle \neq 0$ for some $\boldsymbol{u} \in \mathrm{Null}(\boldsymbol{A})$, then $h'_{\mathcal{C}}(\boldsymbol{\nu}) = -\infty$. For if $\boldsymbol{u} \in \mathrm{Null}(A)$ 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.

If $\boldsymbol{\nu}$ is orthogonal to $\mathrm{Null}(\boldsymbol{A})$, then $\boldsymbol{\nu} \in \mathrm{Row}(\boldsymbol{A})$ and there is a $\boldsymbol{w} \in \mathbb{R}^M$ such that $\boldsymbol{\nu} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$. We have

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

Using the above, the dual problem

$$\underset{\boldsymbol{\nu} \in \mathcal{C}^* \cap \mathcal{D}^*}{\text{maximize}} \ h'_{\mathcal{C}}(\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}),$$

with $\mathcal{C}^* = \text{Row}(\boldsymbol{A})$ and $\mathcal{D}^* = \{\boldsymbol{\nu} \ : \ \|\boldsymbol{\nu}\|_* \leq 1\}$, is equivalent to an optimization over $\boldsymbol{w} \in \mathbb{R}^M$,

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

As an example, if we consider the $\ell_1$-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 can be recast as a linear program, meaning that it can be solved with off-the-shelf software. Also, as we will see on the homework, we can use the Fenchel dual above to provide a theoretical characterization of the properties (e.g., sparsity) of the solution $\boldsymbol{x}^\star$ of the primal problem.

### Fenchel duality for competing objectives

Fenchel duality exists for a slightly more general types of optimization problems than (1). Consider the following minimization problem with competing objectives

$$\underset{\boldsymbol{x}\in\mathcal{C}\cap\mathcal{D}}{\text{minimize}} \; f(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{5}$$

where $f$ is a convex function with domain $\mathcal{D}$ and $g$ is a convex function with domain $\mathcal{C}$. The optimization is trying to balance minimizing $f$ against minimizing $g$. In previous the section we simply took $g(\boldsymbol{x}) = \iota_{\mathcal{C}}(\boldsymbol{x})$ and so we were balancing the minimization of $f$ versus staying in $\mathcal{C}$.

The Fenchel dual for this general problem is

$$\underset{\boldsymbol{\nu}\in\mathcal{C}^*\cap\mathcal{D}^*}{\text{maximize}} \; -g^*(-\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}), \tag{6}$$

where $f^*, g^*$ are conjugates for $f, g$ and $\mathcal{C}^*, \mathcal{D}^*$ are their respective domains. The Fenchel duality theorem states that if $\operatorname{relint}(\mathcal{C}\cap\mathcal{D})$ is nonempty, and $f(\boldsymbol{x}) + g(\boldsymbol{x})$ is bounded below, then we have strong duality in that

$$f(\boldsymbol{x}^\star) + g(\boldsymbol{x}^\star) = -g^*(-\boldsymbol{\nu}^\star) - f^*(\boldsymbol{\nu}^\star),$$

where $\boldsymbol{x}^\star$ is a solution to (5) and $\boldsymbol{\nu}^\star$ is a solution to (6).

We will not give a detailed proof of this here; if you are interested, excellent expositions can be found in [Roc70, Chapter 31] and [Lue69, Chapter 7.12]. But we can sketch how it comes about using the Lagrangian. The primal (5) can be recast as the constrained problem

$$\underset{\boldsymbol{x},\boldsymbol{z}}{\text{minimize}} \; f(\boldsymbol{x}) + g(\boldsymbol{z}) \quad \text{subject to} \quad \boldsymbol{x} = \boldsymbol{z}.$$

The Lagrangian for this problem is

$$L(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\nu}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \langle \boldsymbol{z}-\boldsymbol{x},\boldsymbol{\nu}\rangle,$$

and so the dual function is

$$\begin{aligned} d(\boldsymbol{\nu}) &= \inf_{\boldsymbol{x},\boldsymbol{z}} \; f(\boldsymbol{x}) + g(\boldsymbol{z}) + \langle \boldsymbol{z}-\boldsymbol{x},\boldsymbol{\nu}\rangle \\ &= \inf_{\boldsymbol{x}\in\mathcal{D}} \; f(\boldsymbol{x}) - \langle \boldsymbol{x},\boldsymbol{\nu}\rangle \; + \inf_{\boldsymbol{z}\in\mathcal{C}} \; g(\boldsymbol{z}) + \langle \boldsymbol{z},\boldsymbol{\nu}\rangle \\ &= -\underbrace{\sup_{\boldsymbol{x}\in\mathcal{D}} \langle \boldsymbol{x},\boldsymbol{\nu}\rangle - f(\boldsymbol{x})}_{f^*(\boldsymbol{\nu})} - \underbrace{\sup_{\boldsymbol{z}\in\mathcal{C}} \langle \boldsymbol{z},-\boldsymbol{\nu}\rangle - g(\boldsymbol{z})}_{g^*(-\boldsymbol{\nu})}, \end{aligned}$$

It should be easy to see from the (two infs in the second line) above that

$$d(\boldsymbol{\nu}) = -g^*(\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}) \le f(\boldsymbol{x}) + g(\boldsymbol{x}),$$

for all $\boldsymbol{\nu} \in \mathcal{C}^* \cap \mathcal{D}^*$ and all $\boldsymbol{x} \in \mathcal{C} \cap \mathcal{D}$. Strong duality says that we have $d(\boldsymbol{\nu}^\star) = f(\boldsymbol{x}^\star)$. As before, given a dual solution $\boldsymbol{\nu}^\star$, we can recover a primal solution $\boldsymbol{x}^\star$ by maximizing $\langle \boldsymbol{x},\boldsymbol{\nu}^\star\rangle - f(\boldsymbol{x})$ or $\langle \boldsymbol{x},-\boldsymbol{\nu}^\star\rangle - g(\boldsymbol{\nu}^\star)$.

Finally, we note that the Fenchel duality theorem takes a particularly nice form when optimizing the difference between a convex function $f$ and concave function $h$ (just take $h(x) = -g(x)$ above). Under similar conditions as above, we have

$$\min_{\boldsymbol{x}\in\mathcal{C}\cap\mathcal{D}} f(\boldsymbol{x}) - h(\boldsymbol{x}) = \max_{\boldsymbol{\nu}\in\mathcal{C}^*\cap\mathcal{D}^*} h^*(\boldsymbol{\nu}) - f^*(\boldsymbol{\nu}), \tag{7}$$

where $h^*(\boldsymbol{\nu})$ is the Fenchel conjugate for a concave function

$$h^*(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x},\boldsymbol{\nu}\rangle - h(\boldsymbol{x}).$$

In fact, (7) is usually how Fenchel duality is talked about.

# References

[Bec17] A. Beck. *First-order methods in optimization*. SIAM, 2017.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

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

[Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.