

Lecture: Lagrange Duality*

Conner DiPaolo[†]

Monday, December 3, 2018

1 Lagrange Duality in Optimization

Lagrange duality allows us to turn one constrained optimization problem (called the *primal problem*) into another constrained optimization (called the *dual problem*) which is intimately related. In many cases of interest, the problems are equivalent to each other in that their optimal values are equal, and that solving one allows us to solve the other problem easily. This provides flexibility to the optimizer, since it allows her to choose between

- (a) solving the primal problem by itself,
- (b) solving the dual problem by itself, or
- (c) solving each problem together,

depending on which approach provides the most computationally efficient solution. We've seen the gradient descent algorithm for solving convex optimization problems already (an example of (a)), but there are other algorithms (like the mirror descent algorithm) which exploit (b) to solve the primal problem. We'll see an example later where solving the primal and dual problem *jointly* allows us to solve explicitly linearly constrained convex quadratic programs in a computationally efficient way.

1.1 Generic Optimization Problems

A generic optimization problem can be framed in the following way:

$$\begin{aligned} &\text{minimize: } f_0(\mathbf{x}) \\ &\text{subj. to: } f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ &\quad \quad \quad h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p, \end{aligned}$$

where \mathbf{x} varies over \mathbb{R}^n . This will be called the primal problem, and it's optimal value will be denoted p^* (following [BoydVandenberghe]).

Example 1.1. Consider the optimization problem

$$\begin{aligned} &\text{maximize: } \mathbf{c}^* \mathbf{x} \\ &\text{subj. to: } \mathbf{A} \mathbf{x} \leq \mathbf{b} \\ &\quad \quad \quad \mathbf{x} \text{ has integer components.} \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. This is a generic *integer program*, and is NP-hard. We'll form this in the standard form given above. Note that $x \in \mathbb{R}$ is integral if and only if $\sin(\pi x) = 0$. Moreover, the

*Some of this material is based on Boyd and Vandenberghe's seminal text *Convex Optimization*. If you'd like more details see that book. It's freely available online.

[†]Harvey Mudd College (cdipaolo@hmc.edu)

constraint $\mathbf{Ax} \leq \mathbf{b}$ can be broken into the row constraints $\mathbf{a}_i^* \mathbf{x} \leq \mathbf{b}_i$ for all the rows \mathbf{a}_i of \mathbf{A} . Thus this problem above is equivalent (in terms of optimal value and the actual minimizer \mathbf{x}^*) to

$$\begin{aligned} & \text{minimize: } \mathbf{c}^* \mathbf{x} \\ & \text{subj. to: } \mathbf{a}_i^* \mathbf{x} - \mathbf{b}_i \leq 0, \quad i = 1, \dots, m \\ & \quad \sin(\pi x) = 0. \end{aligned}$$

1.2 The Dual Problem

We'll start with some definitions. We'll let $I_{\leq} : \mathbb{R} \rightarrow \{0, \infty\}$ and $I_0 : \mathbb{R} \rightarrow \{0, \infty\}$ be the indicator functions of the non-positive reals and the point zero, respectively. That is,

$$I_{\leq}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \infty & \text{otherwise.} \end{cases}, \quad I_0(x) = \begin{cases} 0 & \text{if } x = 0 \\ \infty & \text{otherwise.} \end{cases}$$

Now let's give a variational characterization of these indicators (verify this!):

$$I_{\leq}(x) = \sup_{\lambda \geq 0} \lambda x, \quad I_0(x) = \sup_{\nu \in \mathbb{R}} \nu x$$

Observe what this means for optimization, however. Our original constrained optimization problem is equivalent to

$$\begin{aligned} p^* &= \inf_{\mathbf{x}} f_0(\mathbf{x}) + \sum_{i=1}^m I_{\leq}(f_i(\mathbf{x})) + \sum_{i=1}^p I_0(h_i(\mathbf{x})) \\ &= \inf_{\mathbf{x}} \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} \underbrace{f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x})}_{\text{called the Lagrangian } L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}. \end{aligned}$$

Thus

$$p^* = \inf_{\mathbf{x}} \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}),$$

which is just a representation where the inner supremum ensures that \mathbf{x} must lie in the constraint set, at which point we minimize over our objective in \mathbf{x} .

One might ask, what if we re-arranged the optimization procedure this gives us? That is, instead of minimizing over \mathbf{x} after maximizing over the *Lagrange multipliers* $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$, we first minimize over \mathbf{x} , giving the *dual function*

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

before performing the maximization over $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$

$$d^* = \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

As the notation suggests, the maximization problem giving us d^* is called the *dual problem* of the primal optimization giving p^* . Notationally, this dual problem *looks* intimately related to the primal problem, but we can show a strong result, namely that the solution to the dual problem allows us to bound the primal problem.

Theorem 1.2. For any $\lambda \geq 0$ and $\nu \in \mathbb{R}^p$,

$$g(\lambda, \nu) \leq p^*.$$

In particular,

$$d^* = \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} g(\lambda, \nu) \leq p^*.$$

Proof. Note that for any $\lambda \geq 0$ and $\nu \in \mathbb{R}^p$, if we consider a fixed $\mathbf{x} \in \mathbb{R}^n$,

$$L(\mathbf{x}, \lambda, \nu) \leq \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} L(\mathbf{x}, \lambda, \nu)$$

Taking the infimum of both sides with respect to \mathbf{x} gives

$$g(\lambda, \nu) = \inf_{\mathbf{x}} L(\mathbf{x}, \lambda, \nu) \leq \inf_{\mathbf{x}} \sup_{\substack{\lambda \geq 0 \\ \nu \leq 0}} L(\mathbf{x}, \lambda, \nu) = p^*.$$

■

Example 1.3. Consider the optimization problem

$$\begin{aligned} &\text{minimize: } (x - 1)^2 \\ &\text{subj. to: } x^2 \geq 4 \end{aligned}$$

over the real line. Clearly the optimal $x^* = 2$, with the associated $p^* = (x^* - 1)^2 = 1$. This problem has Lagrangian

$$L(x, \lambda) = (x - 1)^2 + \lambda(4 - x^2) = (1 - \lambda)x^2 - 2x + 4\lambda + 1.$$

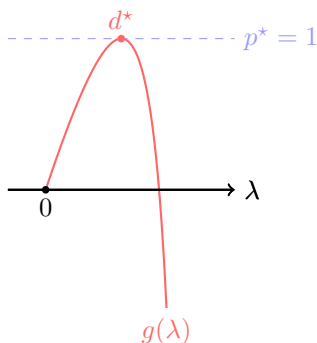
To find the dual function, we compute (verify this!)

$$g(\lambda) = \inf_x L(x, \lambda) = \begin{cases} 4\lambda - \frac{1}{1-\lambda} + 1 & 0 \leq \lambda < 1 \\ -\infty & \text{otherwise.} \end{cases}$$

The theory says, then, that for $\lambda \in [0, 1)$,

$$4\lambda - \frac{1}{1-\lambda} + 1 = g(\lambda) \leq p^* = 1.$$

This is verified visually by the following plot.



We can also see that $d^* = 1 = p^*$, satisfied by $\lambda^* = 1/2$.

The above example exhibits no *duality gap*: $p^* - d^* = 0$. In these situations, solving the dual problem exactly gives us the optimal value of the primal problem. The situation that $d^* = p^*$ is known as *strong duality*; the other case $d^* < p^*$ is known as *weak duality*. In the strong duality situation, using what's known as the KKT (Karush-Kuhn-Tucker) conditions, we can sometimes easily recover the optimal parameters \mathbf{x}^* from this dual solution as well.

1.3 Strong Duality and Slater's Condition

As proposed above, strong duality can be an extremely helpful condition to exhibit, since it allows us to make a choice of which optimization problem we want to solve to achieve at our solution. The question, then, is if there are easy checks we can make to an optimization problem to see if strong duality holds. It turns out that this is the case most of the time if our problem is convex.

An optimization problem is called *convex* if it can be written in the form

$$\begin{aligned} &\text{minimize: } f_0(\mathbf{x}) \\ &\text{subj. to: } f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ &\quad \mathbf{Ax} = \mathbf{b} \end{aligned}$$

where f_i are all convex, $\mathbf{A} \in \mathbb{R}^{p \times n}$ and $\mathbf{b} \in \mathbb{R}^p$. This corresponds to minimizing a convex function over a convex set.

Theorem 1.4 (Slater's Condition). If our problem is convex and there exists an $\mathbf{x} \in \text{dom} f_0$ so that

$$f_i(\mathbf{x}) < 0, \quad i = 1, \dots, m, \quad \mathbf{Ax} = \mathbf{b},$$

then strong duality holds.

1.4 Examples of Dual Problems (LP, QP)

These will follow Boyd and Vandenberghe quite closely.

Example 1.5. Consider the linear program

$$\begin{aligned} &\text{minimize: } \mathbf{c}^* \mathbf{x} \\ &\text{subj. to: } \mathbf{Ax} = \mathbf{b} \\ &\quad \mathbf{x} \geq 0. \end{aligned}$$

The Lagrangian is

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{c}^* \mathbf{x} - \boldsymbol{\lambda}^* \mathbf{x} + \boldsymbol{\nu}^* (\mathbf{Ax} - \mathbf{b}).$$

For fixed $\boldsymbol{\lambda}, \boldsymbol{\nu}$, this is an affine function in \mathbf{x} and hence we can find the dual function easily (verify this!)

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^* \boldsymbol{\nu} & \text{if } \mathbf{A}^* \boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{c} = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Such a $\boldsymbol{\lambda} \geq 0$ exists to satisfy the finite-dual condition $\mathbf{A}^* \boldsymbol{\nu} = -\mathbf{c} + \boldsymbol{\lambda}$ so long as $\mathbf{A}^* \boldsymbol{\nu} \geq -\mathbf{c}$. Since the non-negative multiplier $\boldsymbol{\lambda}$ is not irrelevant to the dual function

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^* \boldsymbol{\nu} & \text{if } \mathbf{A}^* \boldsymbol{\nu} \geq -\mathbf{c} \\ -\infty & \text{otherwise,} \end{cases}$$

we can write the dual problem to the original (primal) linear program above as

$$\begin{aligned} &\text{minimize: } \mathbf{b}^* \boldsymbol{\nu} \\ &\text{subj. to: } \mathbf{A}^* \boldsymbol{\nu} \geq -\mathbf{c}, \end{aligned}$$

which is just another linear program. When does strong duality hold?

Example 1.6. Suppose we have a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ which is symmetric. We want to partition a group of items

$$\{1, 2, \dots, n\} = \{i : \mathbf{x}_i = 1\} \cup \{i : \mathbf{x}_i = -1\} = A_+ \cup A_-$$

into two sets A_+ and A_- , which are identified from a vector $\mathbf{x} \in \{\pm 1\}^n$. There is a profit \mathbf{A}_{ij} associated with keeping i and j in the same partition, and a cost $-\mathbf{A}_{ij}$ associated with putting them in the same partition.

Thus the total cost associated with a partition is (up to an additive constant $\text{tr } \mathbf{A}$) equal to $\mathbf{x}^* \mathbf{A} \mathbf{x}$. We wish to find the partition which minimizes the cost. This corresponds to the optimization problem

$$\begin{aligned} & \text{minimize: } \mathbf{x}^* \mathbf{A} \mathbf{x} \\ & \text{subj. to: } \mathbf{x}_i^2 = 1, i = 1, 2, \dots, n. \end{aligned}$$

This problem is nonconvex and no apparent solution besides brute force iteration seems available, but passing to the dual (which turns out to be convex) allows us to efficiently find a lower bound to the problem. Exercise: prove $p^* \leq \text{tr } \mathbf{A}$.

The Lagrangian

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^* \mathbf{A} \mathbf{x} + \sum_{i=1}^n \nu_i (\mathbf{x}_i^2 - 1) = \mathbf{x}^* (\mathbf{A} + \text{diag}(\boldsymbol{\nu})) \mathbf{x} - \mathbf{1}^* \boldsymbol{\nu}.$$

Thus the dual function

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{1}^* \boldsymbol{\nu} & \text{if } \mathbf{A} + \text{diag}(\boldsymbol{\nu}) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

This gives the dual problem

$$\begin{aligned} & \text{minimize: } \mathbf{1}^* \boldsymbol{\nu} \\ & \text{subj. to: } \mathbf{A} + \text{diag}(\boldsymbol{\nu}) \succeq 0, \end{aligned}$$

which is a convex semidefinite program. It turns out strong duality does not hold here. Taking specific values of $\boldsymbol{\nu}$ gives helpful lower bounds on the original problem without the need to solve exactly. For example, if $\boldsymbol{\nu} = -\lambda_{\min}(\mathbf{A}) \mathbf{1}$ then

$$\mathbf{A} + \text{diag}(\boldsymbol{\nu}) = \mathbf{A} - \lambda_{\min}(\mathbf{A}) \mathbf{1} \succeq 0$$

is dual feasible, and hence

$$p^* \geq -\mathbf{1}^* \boldsymbol{\nu} = n \lambda_{\min}(\mathbf{A}).$$

This bound is quite helpful when \mathbf{A} is positive definite. In conclusion,

$$n \lambda_{\min}(\mathbf{A}) \leq p^* \leq \text{tr}(\mathbf{A}).$$

1.5 Application: Solutions to Linearly Constrained Quadratic Programs

Consider the problem

$$\begin{aligned} & \text{minimize: } \frac{1}{2} \mathbf{x}^* \mathbf{Q} \mathbf{x} + \mathbf{c}^* \mathbf{x} \\ & \text{subj. to: } \mathbf{A} \mathbf{x} = \mathbf{b} \end{aligned}$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is positive definite, $\mathbf{c} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. For the optimal \mathbf{x} , $\mathbf{A} \mathbf{x} = \mathbf{b}$ by assumption. Moreover, strong duality holds trivially if the problem is feasible, and so we can solve by the dual to find the original problem. The Lagrangian

$$L(\mathbf{x}, \boldsymbol{\nu}) = \frac{1}{2} \mathbf{x}^* \mathbf{Q} \mathbf{x} + \mathbf{c}^* \mathbf{x} + \boldsymbol{\nu}^* (\mathbf{A} \mathbf{x} - \mathbf{b}),$$

and for any $\boldsymbol{\nu}$ the dual function must satisfy

$$\partial_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{Q} \mathbf{x} + \mathbf{c} + \mathbf{A}^* \boldsymbol{\nu} = 0.$$

Thus we have two equations

$$\mathbf{A} \mathbf{x} = \mathbf{b}, \quad \mathbf{Q} \mathbf{x} + \mathbf{A}^* \boldsymbol{\nu} = -\mathbf{c}$$

from which we can solve the primal and dual problem *jointly* as the solution to a single linear system

$$\begin{bmatrix} \mathbf{Q} & \mathbf{A}^* \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \end{bmatrix}.$$

This system is used heavily with interior point methods for solving arbitrary systems.

1.6 Open Problem: Eigenvalues of Non-Negative Matrices via Convex Optimization

If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive and irreducible then the Collatz-Wielandt formula [AnantharamBorkar] says that the Perron-Frobenius eigenvalue

$$\lambda = \inf_{\mathbf{x} > 0} \max_{1 \leq i \leq n} \frac{(\mathbf{A}\mathbf{x})_i}{x_i}.$$

Write $\mathbf{z} \in \mathbb{R}^n$ as $z_i = \log x_i \in \mathbb{R}$ so that

$$\alpha := \log \lambda = \inf_{\mathbf{x} > 0} \max_{1 \leq i \leq n} \log \frac{(\mathbf{A}\mathbf{x})_i}{x_i} = \inf_{\mathbf{z}} \max_{1 \leq i \leq n} \log \left(\sum_{j=1}^n \mathbf{a}_{ij} e^{z_j} \right) - z_i$$

since the logarithm is increasing and this transformation is a bijection. Here, $\mathbf{a}_i \in \mathbb{R}^n$ is the i -th row of \mathbf{A} . Define the function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$f_i(\mathbf{z}) = \log \left(\sum_{j=1}^n \mathbf{a}_{ij} e^{z_j} \right) - z_i = \log \left(\sum_{j=1}^n e^{z_j + \log \mathbf{a}_{ij}} \right) - z_i.$$

Convexity of the log-sum-exp function ensures that each f_i are convex, and since the maximum

$$f(\mathbf{z}) = \max_{1 \leq i \leq n} f_i(\mathbf{z})$$

of convex functions is convex we know that the problem of finding the log-Perron eigenvalue

$$\alpha = \inf_{\mathbf{z}} f(\mathbf{z})$$

is a convex optimization problem. Can we solve this optimization problem faster than the power method, either in primal or dual form?