Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 18

Weak Duality

Just as we have two eyes and two feet, duality is part of life.

Carlos Santana

Outline

1. Weak duality
   - Lagrangian
   - Minimax inequality
   - Weak duality
   - Geometry

2. Examples
   - Projection on the probability simplex
   - Sum of k largest elements
   - Dual of an LP

3. Take-aways

Constrained optimization problem

Consider an optimization problem in standard form

$$
\begin{aligned}
p^* = \min_{x \in \mathbb{R}^n} \quad & f_0(x) \\
\text{subject to:} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\
& h_i(x) = 0, \quad i = 1, \ldots, q,
\end{aligned}
\tag{1}
$$

and let $\mathcal{D}$ denote the domain of this problem, assumed to be nonempty.

We refer to the above problem as the *primal* problen.

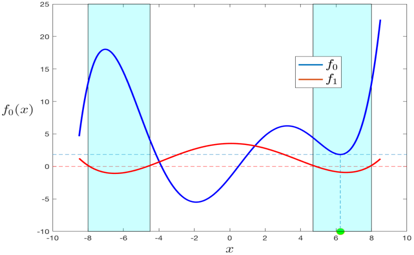
Note: we are not assuming convexity of f_0, $f_1, \ldots, f_m$ or of $h_1, \ldots, h_q$, for the time being.

A running example

To illustrate, we focus on a problem with a single inequality constraint, with f_0, f_1 defined as

$$f_0(x) := \begin{cases} \begin{array}{l} 0.0025x^5 - 0.00175x^4 - 0.212625x^3 \\ +0.3384375x^2 + 3.368x - 1.692 \end{array} & -10 \leq x \leq 10, \\ +\infty & \text{otherwise,} \end{cases}$$

$$f_1(x) := 0.0025x^4 - 0.0005x^3 - 0.2129x^2 + 0.0320x + 3.5340.$$



A one-dimensional problem: minimize a fifth-order polynomial on the domain $\mathcal{D} = [-10, 10]$, with one quadratic inequality constraint that requires x to belong to the union of two intervals (indicated in light blue). The (unique) optimal point is shown in green on the x-axis.

Lagrangian

Define a new function, called the *Lagrangian*, with values for $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^q$:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{q} \nu_i h_i(x).$$

Vectors λ and ν are referred to as *Lagrange multipliers*, or dual variables.

Example: for the previous problem, the Lagrangian is given by: for $x \in \mathcal{D} = [-10, 10]$ and $\lambda \in \mathbb{R}$:

$$\mathcal{L}(x, \lambda) = f_0(x) + \lambda f_1(x) = \text{ a polynomial of degree 5.}$$

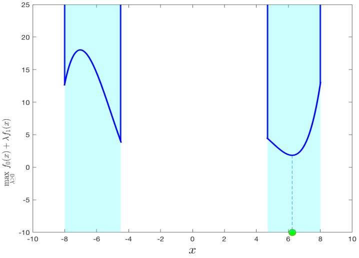
Problem in min-max form

Thanks to the Lagrangian we may express the problem in "min-max" form:

$$p^* = \min_x \max_{\lambda \geq 0,\, \nu} \mathcal{L}(x, \lambda, \nu).$$

The above is due to the fact that, for any x,

$$\max_{\lambda \geq 0,\, \nu} \mathcal{L}(x, \lambda, \nu) = \begin{cases} f_0(x) & \text{if } x \text{ is feasible,} \\ +\infty & \text{otherwise.} \end{cases}$$



We have encoded the problem as one without constraint, by re-defining the objective to be $+\infty$ outside the feasible set. The minimizer of the function (green) is optimal for the original problem.

Minimax inequality

For any sets X, Y and any function $F : X \times Y \to \mathbb{R}$:

$$\min_{x \in X} \max_{y \in Y} F(x, y) \geq \max_{y \in Y} \min_{x \in X} F(x, y).$$

Proof: for any $(x_0, y_0) \in X \times Y$:

$$h(y_0) \doteq \min_{x \in X} F(x, y_0) \leq F(x_0, y_0) \leq \max_{y \in Y} F(x_0, y) \doteq g(x_0).$$

Hence, $h(y_0) \leq g(x_0)$. Result follows from taking the max over $y_0 \in Y$, then the min over $x_0 \in X$.

Interpretation as a game

Assume you play game against an opponent: given the payoff matrix below, you pick a row $i \in \{1, \ldots, 5\}$ and the opponent a column $j \in \{1, \ldots, 6\}$. The payoff to you, the maximizing player, and cost to your opponent, the minimizing player, is M_{ij}, where M is the payoff matrix. Players play once, one after the other. The second player sees what the first does.

7	-8	-7	-8	3	5
9	-5	10	-2	-10	5
-8	1	10	9	7	-2
9	10	0	6	9	3
3	10	6	10	4	-7

Payoff matrix representing the payoff to the maximizing player. It is equal to the cost to the minimizing player. This is thus a "zero-sum" game.

Question: Do you prefer to play first, or second? What is your payoff in each case?

Game interpretation (cont'd)

7	-8	-7	-8	3	5	-8
9	-5	10	-2	-10	5	-10
-8	1	10	9	7	-2	-8
9	**10**	**0**	**6**	**9**	**3**	**0**
3	10	6	10	4	-7	-7

If the maximizing player plays first, it will select a row (in **bold**) that maximizes the worst-case (minimum) payoff (in blue); the second player chooses the smallest element in that row. The payoff is

$$d^* = \max_i \min_j M_{ij} = \mathbf{0}.$$

7	-8	-7	-8	**5**	3
9	-5	10	-2	**5**	-10
-8	1	10	9	**-2**	7
9	10	0	6	**3**	9
3	10	6	10	**-7**	4
9	10	10	9	5	**3**

If the minimizing player plays first, it will select a column (in **bold**) that minimizes the worst-case (maximum) cost (in red); the second player accordingly chooses the largest element in that row. The payoff is

$$p^* = \min_j \max_i M_{ij} = \mathbf{3}.$$

It is always better to play **second** in this game, since the second player can adapt to the decision of the first; the first player must account for the worst-case.

Weak duality

Applying the minimax inequality to the Lagrangian, we obtain:

$$p^* = \min_x \max_{\lambda \geq 0,\, \nu} \mathcal{L}(x, \lambda, \nu) \geq d^* \doteq \max_{\lambda \geq 0,\, \nu} \min_x \mathcal{L}(x, \lambda, \nu).$$

- The problem on the right is called the dual problem; it involves maximizing (over $\lambda \geq 0, \nu$) the dual function:

$$g(\lambda, \nu) \doteq \min_x \mathcal{L}(x, \lambda, \nu).$$

- Since g is the pointwise minimum of affine (hence, concave) functions, g is concave.
- Hence the dual problem, a concave maximization problem over a convex set ($\mathbb{R}_+^m \times \mathbb{R}$), is convex!

Geometry

Making the problem 2D

Consider the problem, with variable $x \in \mathbb{R}^n$:

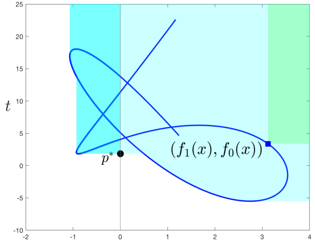
$$p^* = \min_x \ f_0(x) \ : \ f_1(x) \leq 0.$$

Define the 2D set of "achievable" values:

$$\mathcal{A} = \left\{ (u, t) \in \mathbb{R}^2 \ : \ \exists \, x \in \mathbb{R}^n, \ \ u \geq f_1(x), \ \ t \geq f_0(x) \right\}.$$

We can visualize the problem as a 2D problem:

$$p^* = \min_{u,t} \ t \ : \ (u, t) \in \mathcal{A}, \ \ u \leq 0.$$



For our example: set $\mathcal{A}$, generated by plotting the set $\{(f_1(x), f_0(x)) \ : \ x \in \mathbb{R}^n\}$, including the NE quadrant (green) at each point. Feasible points correspond to where the curve intersects the set of pairs (u, t), with $u \leq 0$ (dark blue).

Geometry

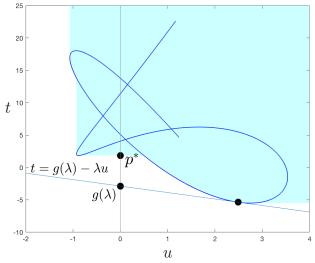
We have

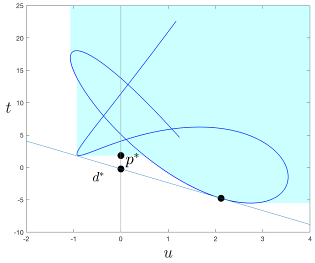
$$p^* = \min_{(u,t)\in\mathcal{A}} \max_{\lambda\geq 0} t + \lambda u \geq d^* = \max_{\lambda\geq 0} g(\lambda),$$

where

$$g(\lambda) = \min_{x} f_0(x) + \lambda f_1(x) = \min_{(u,t)\in\mathcal{A}} t + \lambda u.$$

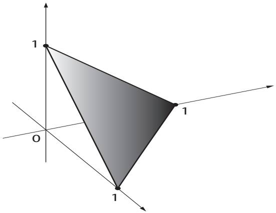
For a given λ, the function $g(\lambda)$ is a lower bound on p^*. The dual problem consists in finding the best such lower bound.





Example

Projection on the probability simplex



The probability simplex in $\mathbb{R}^n$ is the set of discrete probabilities

$$\Delta^n \doteq \left\{ x \in \mathbb{R}^n \; : \; x \geq 0, \;\; \sum_{i=1}^{n} x_i = 1 \right\}.$$

The problem of projecting a given vector $z \in \mathbb{R}^n$ onto the simplex arises in many contexts. The projection problem writes

$$\min_x \; \frac{1}{2}\|x - z\|_2^2 \; : \; x \geq 0, \;\; \sum_{i=1}^{n} x_i = 1.$$

Projection on the probability simplex

Dual problem

Lagrangian:

$$\mathcal{L}(x, \nu) = \frac{1}{2}\|x - z\|_2^2 + \nu(1 - \mathbf{1}^\top x) \;:\; x \geq 0.$$

Dual function:

$$g(\nu) = \min_{x \geq 0} \; \mathcal{L}(x, \nu) = \frac{1}{2} z^\top z + \nu - \frac{1}{2} \sum_{i=1}^{n} \max(0, z_i + \nu)^2, \tag{2}$$

where we use the fact that, for a given $\beta \in \mathbb{R}$:

$$\min_{\xi \geq 0} \; \frac{1}{2}\xi^2 - \beta\xi = -\frac{1}{2} \max(0, \beta)^2.$$

The function g can be optimized by brute-force line search, or (faster) bisection methods.

By dualizing the equality constraint, we made the problem (2) easy (decoupled)!

Projection on the probability simplex

Strong duality

For every $\nu \in \mathbb{R}$, the solution to the problem

$$\min_{x \geq 0} \; \mathcal{L}(x, \nu)$$

is unique, and characterized by the zero-gradient condition $\nabla_x \mathcal{L}(x, \nu) = 0$, leading to

$$x_i^*(\nu) = \max(0, z_i + \nu), \quad i = 1, \ldots, m.$$

In addition, the dual function g is smooth, and at its maximum its gradient is zero:

$$0 = \nabla_\nu \, g(\nu^*) = 1 - \sum_{i=1}^n \max(0, z_i + \nu^*) = 1 - \sum_{i=1}^n x_i^*(\nu^*),$$

which proves that the point $x^*(\nu^*)$ is feasible for the primal problem.

Projection on the probability simplex

Strong duality

Further, after some algebra, exploiting $\mathbf{1}^\top x^*(\nu^*) = 1$, it can be shown that

$$\frac{1}{2}\|x^*(\nu^*) - z\|_2^2 = g(\nu^*) = d^*,$$

which proves that $x^*(\nu^*)$ attains the dual lower bound, hence it is optimal, and "strong duality" holds, that is:

$$p^* = d^*.$$

This is an example where we are able to recover a primal feasible point from the dual and prove that strong duality holds, so that solving the dual solves the original problem. We will see later how to generalize this approach.

Example

Sum of k largest elements

For given $w \in \mathbb{R}^n$, and $k \in \{1, \dots, n-1\}$, we define

$$s_k(w) = \sum_{i=1}^{k} w_{[i]},$$

where $w_{[i]}$ is the i-th largest element in w.

The function s_k is convex, due to the pointwise maximum rule:

$$\begin{aligned} s_k(w) &= \max_{\mathcal{I}} \sum_{i \in \mathcal{I}} w_i \ : \ \mathcal{I} \subseteq \{1, \dots, n\}, \ \ \mathbf{Card}\, \mathcal{I} \le k \\ &= \max_{u \in \{0,1\}^n} \ u^\top w \ : \ \mathbf{1}^\top u = k. \end{aligned}$$

Weak and strong duality

By weak duality (third line):

$$
\begin{aligned}
s_k(w) &= \max_{u \in \{0,1\}^n} \; u^\top w \;:\; \mathbf{1}^\top u = k \\
&= \max_{u \in \{0,1\}^n} \min_{\nu} \; u^\top w + \nu(k - \mathbf{1}^\top u) \\
&\leq \min_{\nu} \max_{u \in \{0,1\}^n} \; u^\top w + \nu(k - \mathbf{1}^\top u) \\
&= \min_{\nu} \; k\nu + \sum_{i=1}^{n} \max(0, w_i - \nu),
\end{aligned}
$$

exploiting in the last line that for any vector z

$$
\max_{u \in \{0,1\}} \; u^\top z = \sum_{i=1}^{n} \max(0, z_i).
$$

We observe that if ν is set to the $(k+1)$-th largest element in w, then we recover $s_k(w)$. Hence equality (strong duality) holds on the second line, and we obtained the dual form:

$$
s_k(w) = \min_{\nu} \; k\nu + \sum_{i=1}^{n} \max(0, w_i - \nu).
$$

Application

Diversification in resource allocation

Consider an asset allocation problem where $w \geq 0$ is a vector containing the amountinvested in the different assets:

$$\max_{w \in \mathcal{W}} \; r^\top w \;\;:\;\; w \geq 0, \;\; s_k(w) \leq \theta \sum_{i=1}^{n} w_i,$$

where $\theta \in [0, 1]$, and

- $r \in \mathbb{R}^n$ contains the expected return on investment for each asset;
- The polytope $\mathcal{W}$ encodes other constraints on w (such as, upper bound on its elements);
- The constraint on $s_k(w)$ means that no more than a fraction θ of the total budget $\mathbf{1}^\top w$ is ascribed to the k largest investments.

The above problem is an LP, provided we are willing to express the constraint on $s_k(w)$ as an exponential list of ordinary affine inequalities in w:

$$\forall\, \mathcal{I} \subseteq \{1, \ldots, n\}, \;\; \textbf{Card}\, \mathcal{I} \leq k \;\;:\;\; \sum_{i \in \mathcal{I}} w_i \leq \theta \sum_{i=1}^{n} w_i.$$

Using the dual form

The previous naïve approach is not practical, as there are n-choose-k constraints.

The constraint $s_k(w) \leq \theta(\mathbf{1}^\top w)$ holds if and only if there exist ν such that

$$k\nu + \sum_{i=1}^{n} \max(0, w_i - \nu) \leq \theta \sum_{i=1}^{n} w_i.$$

The above is a convex, perfectly manageable constraint. It can even be represented in linear inequality form, by introducing n slack variables

$$k\nu + \sum_{i=1}^{n} s_i \leq \theta \sum_{i=1}^{n} w_i, \quad s \geq 0, \quad s \geq w - \nu\mathbf{1}.$$

Thus, at the price of augmenting the number of variables, we avoided dealing with an exponential number of constraints.

Geometrically: the set corresponding to the constraint on $s_k(w)$ is a polytope in $\mathbb{R}^n$, with 2^n facets; it is the projection of another polytope in $\mathbb{R}^{2n+1}$ that has $2n+1$ facets only.

Dual of a linear program

- Consider the following optimization problem with linear objective and linear inequality constraints (a so-called linear program in standard inequality form)

$$\begin{aligned} p^* = \min_x \quad & c^\top x \\ \text{s.t.:} \quad & Ax \leq b, \end{aligned} \tag{3}$$

  where $A \in \mathbb{R}^{m,n}$ is a matrix of coefficients, and the inequality $Ax \leq b$ is to be intended elementwise.

- The Lagrangian for this problem is

$$\mathcal{L}(x, \lambda) = c^\top x + \lambda^\top (Ax - b) = (c + A^\top \lambda)^\top x - \lambda^\top b.$$

- In order to determine the dual function $g(\lambda)$ we next need to minimize $\mathcal{L}(x, \lambda)$ w.r.t. x. But $\mathcal{L}(x, \lambda)$ is affine in x, hence this function is unbounded below, unless the vector coefficient of x is zero (i.e., $c + A^\top \lambda = 0$), and it is equal to $-\lambda^\top b$ otherwise. That is,

$$g(\lambda) = \begin{cases} -\infty & \text{if } c + A^\top \lambda \neq 0 \\ -\lambda^\top b & \text{if } c + A^\top \lambda = 0. \end{cases}$$

Dual of a linear program

- The dual problem then amounts to maximizing $g(\lambda)$ over $\lambda \geq 0$:

$$\begin{aligned} d^* = \max_{\lambda} \quad & -\lambda^\top b \\ \text{s.t.:} \quad & c + A^\top \lambda = 0, \\ & \lambda \geq 0. \end{aligned} \tag{4}$$

- From weak duality, we have that $d^* \leq p^*$.
- We may also rewrite the dual problem into an equivalent minimization form, by changing the sign of the objective, which results in

$$\begin{aligned} -d^* = \min_{\lambda} \quad & b^\top \lambda \\ \text{s.t.:} \quad & A^\top \lambda + c = 0, \\ & \lambda \geq 0, \end{aligned}$$

and this is again an LP, in standard conic form.

Take-aways

Weak duality:

- We consider a non-convex minimization problem, and refer to it as the "primal" problem.
- Weak duality is a process by which we find a lower bound on the optimal value of the primal.
- It is based on expressing the primal problem in a min-max form, and applying the minimax inequality.
- The lower bound is the value of an optimization problem, referred to as the dual.
- The dual problem is a convex problem, even if the primal is not.

Coming up next:

- can we make duality strong?
- How can we recover a primal point from the dual problem?
- What are applications of duality?