1 Motivation

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

We'll build $g(y)$, that preserves the uniform bound:

$$g(y) \leq f(x) \qquad \forall x \in S, \forall y \in \Omega$$

As a consequence:

$$\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)$$

We'll consider one of many possible ways to construct $g(y)$ in case, when we have a general mathematical programming problem with functional constraints:

$$\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \ i = 1, \ldots, m \\
& h_i(x) = 0, \ i = 1, \ldots, p
\end{split}$$

And the Lagrangian, associated with this problem:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) = f_0(x) + \lambda^\top f(x) + \nu^\top h(x)$$

We assume $\mathcal{D} = \bigcap_{i=0}^m \textbf{dom } f_i \cap \bigcap_{i=1}^p \textbf{dom } h_i$ is nonempty. We define the Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the original problem is not convex.

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0, \ \lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{= 0} \leq f_0(\hat{x})$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

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

A natural question is: what is the *best* lower bound that can be obtained from the Lagrange dual function? This leads to the following optimization problem:

$$
\begin{aligned}
g(\lambda, \nu) &\to \max_{\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p} \\
\text{s.t. } \lambda &\succeq 0
\end{aligned}
$$

The term “dual feasible”, to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$, now makes sense. It means, as the name implies, that (λ, ν) is feasible for the dual problem. We refer to (λ^*, ν^*) as dual optimal or optimal Lagrange multipliers if they are optimal for the above problem.

1.1 Summary

	Primal	Dual
Function	$f_0(x)$	$g(\lambda, \nu) = \min\limits_{x \in \mathcal{D}} L(x, \lambda, \nu)$
Variables	$x \in S \subseteq \mathbb{R}^n$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
Constraints	$f_i(x) \le 0,\ i = 1, \dots, m$ $h_i(x) = 0,\ i = 1, \dots, p$	$\lambda_i \ge 0, \forall i \in \overline{1, m}$
Problem	$\begin{aligned} & f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & f_i(x) \le 0,\ i = 1, \dots, m \\ & h_i(x) = 0,\ i = 1, \dots, p \end{aligned}$	$\begin{aligned} g(\lambda, \nu) & \to \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \\ \text{s.t. } & \lambda \succeq 0 \end{aligned}$
Optimal	x^* if feasible, $p^* = f_0(x^*)$	λ^*, ν^* if $\max$ is achieved, $d^* = g(\lambda^*, \nu^*)$

Least-squares solution of linear equations

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
\min \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

Solution

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

leading to $x = -(1/2)A^T \nu$. As a result, the dual function is articulated as

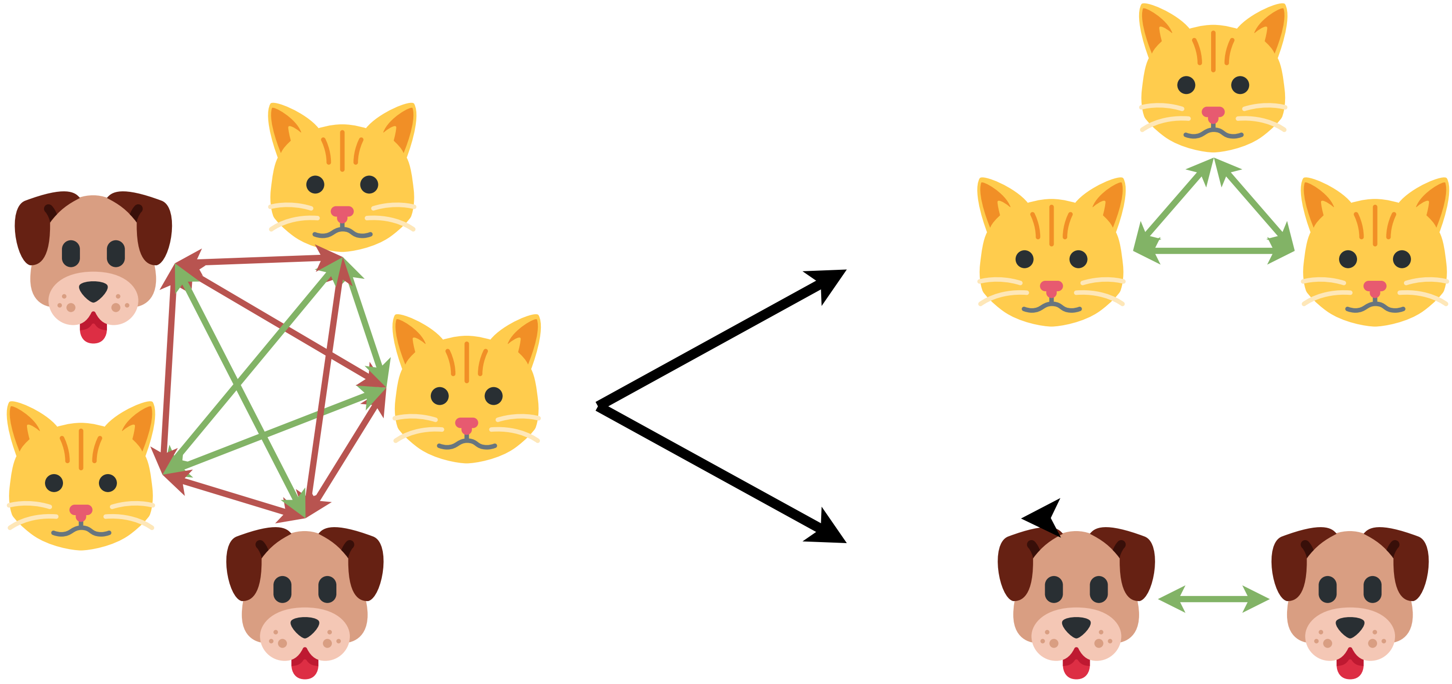
$$
g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,
$$

emerging as a concave quadratic function within the domain $\mathbb{R}^p$. According to the lower bound property (5.2), for any $\nu \in \mathbb{R}^p$, the following holds true:

$$
-(1/4)\nu^T A A^T \nu - b^T \nu \le \inf\{x^T x \mid Ax = b\}.
$$

Which is a simple non-trivial lower bound without any problem solving.

Two-way partitioning problem



We are examining a (nonconvex) problem:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array}$$

🔽 **Solution**

The matrix W belongs to S_n. The constraints stipulate that the values of x_i can only be 1 or -1, rendering this problem analogous to finding a vector, with components ± 1, that minimizes $x^T W x$. The set of feasible solutions is finite, encompassing 2^n points, thereby allowing, in theory, for the resolution of this problem by evaluating the objective value at each feasible point. However, as the count of feasible points escalates exponentially, this approach is viable only for modest-sized problems (for instance, when $n \leq 30$). Generally, and especially when n exceeds 50, the problem poses a formidable challenge to solve.

This problem can be construed as a two-way partitioning problem over a set of n elements, denoted as $\{1, \ldots, n\}$: A viable x corresponds to the partition

$$\{1, \ldots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.$$

The coefficient W_{ij} in the matrix represents the expense associated with placing elements i and j in the same partition, while $-W_{ij}$ signifies the cost of segregating them. The objective encapsulates the aggregate cost across all pairs of elements, and the challenge posed by problem is to find the partition that minimizes the total cost.

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise}, \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$\nu = -\lambda_{\min}(W) \mathbf{1}$$

which is dual feasible, since

$$W + \mathbf{diag}(\nu) = W - \lambda_{\min}(W) I \succeq 0.$$

This renders a simple bound on the optimal value p^*

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

The code for the problem is available here 🧑‍💻

2 Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$p^* \geq d^*$$

While the difference between them is often called **duality gap:**

$$p^* - d^* \geq 0$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always concave, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality happens if duality gap is zero:

$$p^* = d^*$$

Notice: both p^* and d^* may be $+\infty$.

- Several sufficient conditions known!
- "Easy" necessary and sufficient conditions: unknown.

🤔 **Question**

In the Least-squares solution of linear equations example above calculate the primal optimum p^* and the dual optimum d^* and check whether this problem has strong duality or not.

3 Useful features

- **Construction of lower bound on solution of the direct problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

- **Sometimes it is easier to solve a dual problem than a primal one.**

  In this case, if the strong duality holds: $g(y^*) = f_0(x^*)$ we lose nothing.

- **Obtaining a lower bound on the function's residual.**

  $f_0(x) - f_0^* \leq f_0(x) - g(y)$ for an arbitrary $y \in \Omega$ (suboptimality certificate). Moreover, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

- **Dual function is always concave**

  As a pointwise minimum of affine functions.

🧑‍🏫 **Projection onto probability simplex**

To find the Euclidean projection of $x \in \mathbb{R}^n$ onto probability simplex $\mathcal{P} = \{z \in \mathbb{R}^n \mid z \succeq 0, \mathbf{1}^\top z = 1\}$, we solve the following problem:

$$\begin{split}
&\frac{1}{2}\|y - x\|_2^2 \to \min_{y \in \mathbb{R}^n \succeq 0} \\
\text{s.t. } &\mathbf{1}^\top y = 1
\end{split}$$

Hint: Consider the problem of minimizing $\frac{1}{2}\|y - x\|_2^2$ subject to subject to $y \succeq 0, \mathbf{1}^\top y = 1$. Form the partial Lagrangian

$$
L(y, \nu) = \frac{1}{2}\|y - x\|_2^2 + \nu(\mathbf{1}^\top y - 1),
$$

leaving the constraint $y \succeq 0$ implicit. Show that $y = (x - \nu\mathbf{1})_+$ minimizes $L(y, \nu)$ over $y \succeq 0$.

🤔 **Projection on the Euclidian Ball**

Find the projection of a point x on the Euclidian ball

$$
\begin{split}
& \frac{1}{2}\|y - x\|_2^2 \to \min_{y \in \mathbb{R}^n} \\
\text{s.t. } & \|y\|_2^2 \leq 1
\end{split}
$$

4 Slater's condition

🤓 **Theorem**

If for a convex optimization problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$ (existance of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.

🧑‍🏫 **An example of convex problem, when Slater's condition does not hold**

$$
\min \{ f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0 \},
$$

The only point in the budget set is: $x^* = 0$. However, it is impossible to find a non-negative $\lambda^* \geq 0$, such that

$$
\nabla f_0(0) + \lambda^* \nabla f_1(0) = 1 + \lambda^* x = 0.
$$

🧑‍🏫 **A nonconvex quadratic problem with strong duality**

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{split}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{split}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

🔽 **Solution**

Lagrangian and dual function

$$
L(x, \lambda) = x^\top A x + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I)x + 2b^\top x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^\top (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
$$

Dual problem:

$$
\begin{split}
& -b^\top (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & A + \lambda I \succeq 0
\end{split}
$$

$$
\begin{split}
& -\sum_{i=1}^n \frac{(q_i^\top b)^2}{\lambda_i + \lambda} - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & \lambda \geq -\lambda_{min}(A)
\end{split}
$$

4.1 Reminder of KKT statements:

Suppose we have a general optimization problem (from the chapter)

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
\tag{1}
$$

and convex optimization problem (see corresponding chapter), where all equality constraints are affine: $h_i(x) = a_i^T x - b_i, i \in 1, \ldots p$

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& Ax = b,
\end{split}
\tag{2}
$$

The Lagrangian is

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

The KKT system is:

$$
\begin{split}
& \nabla_x L(x^*, \lambda^*, \nu^*) = 0 \\
& \nabla_\nu L(x^*, \lambda^*, \nu^*) = 0 \\
& \lambda_i^* \geq 0, i = 1, \ldots, m \\
& \lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m \\
& f_i(x^*) \leq 0, i = 1, \ldots, m
\end{split}
\tag{3}
$$

🤓 **KKT becomes necessary**

If x^* is a solution of the original problem Equation 1, then if any of the following regularity conditions is satisfied:

- **Strong duality** If $f_1, \ldots f_m, h_1, \ldots h_p$ are differentiable functions and we have a problem Equation 1 with zero duality gap, then Equation 3 are necessary (i.e. any optimal set x^*, λ^*, ν^* should satisfy Equation 3)
- **LCQ** (Linearity constraint qualification). If $f_1, \ldots f_m, h_1, \ldots h_p$ are affine functions, then no other condition is needed.
- **LICQ** (Linear independence constraint qualification). The gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at x^*
- **SC** (Slater's condition) For a convex optimization problem Equation 2 (i.e., assuming minimization, f_i are convex and h_j is affine), there exists a point x such that $h_j(x) = 0$ and $g_i(x) < 0$.

Than it should satisfy Equation 3

🤓 **KKT in convex case**

If a convex optimization problem Equation 2 with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so x^* is optimal if and only if there are (λ^*, ν^*) that, together with x^*, satisfy the KKT conditions.

5 Applications

5.1 Connection between Fenchel duality and Lagrange duality

🧑‍🏫 **Example**

$$
\begin{aligned}
f_0(x) = \sum_{i=1}^n f_i(x_i) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } a^\top x = b &
\end{aligned}
$$

The dual problem is thus

$$
\begin{aligned}
-b\nu - \sum_{i=1}^n f_i^*(-\nu a_i) &\to \max_{\nu \in \mathbb{R}} \\
\text{s.t. } \nu \geq -\lambda_{\text{min}}(A) &
\end{aligned}
$$

with (scalar) variable $\nu \in \mathbb{R}$. Now suppose we have found an optimal dual variable ν^* (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.). It is very easy to recover the optimal value for the primal problem.

Let $f : E \to \mathbb{R}$ and $g : G \to \mathbb{R}$ — function, defined on the sets E and G in Euclidian Spaces V and W respectively. Let $f^* : E_* \to \mathbb{R}, g^* : G_* \to \mathbb{R}$ be the conjugate functions to the f and g respectively. Let $A : V \to W$ — linear mapping. We call Fenchel - Rockafellar problem the following minimization task:

$$
f(x) + g(Ax) \to \min_{x \in E \cap A^{-1}(G)}
$$

where $A^{-1}(G) := \{x \in V : Ax \in G\}$ — preimage of G. We'll build the dual problem using variable separation. Let's introduce new variable $y = Ax$. The problem could be rewritten:

$$
\begin{aligned}
f(x) + g(y) &\to \min_{x \in E, \; y \in G} \\
\text{s.t. } Ax = y &
\end{aligned}
$$

Lagrangian

$$
L(x, y, \lambda) = f(x) + g(y) + \lambda^\top (Ax - y)
$$

Dual function

$$
\begin{aligned}
g_d(\lambda) &= \min_{x \in E, \; y \in G} L(x, y, \lambda) \\
&= \min_{x \in E} \left[f(x) + (A^*\lambda)^\top x \right] + \min_{y \in G} \left[g(y) - \lambda^\top y \right] = \\
&= -\max_{x \in E} \left[(-A^*\lambda)^\top x - f(x) \right] - \max_{y \in G} \left[\lambda^\top y - g(y) \right]
\end{aligned}
$$

Now, we need to remember the definition of the conjugate function:

$$
\sup_{y \in G} \left[\lambda^\top y - g(y) \right] = \begin{cases} g^*(\lambda), & \text{if } \lambda \in G_* \\ +\infty, & \text{otherwise} \end{cases}
$$

$$
\sup_{x \in E} \left[(-A^*\lambda)^\top x - f(x) \right] = \begin{cases} f^*(-A^*\lambda), & \text{if } \lambda \in (-A^*)^{-1}(E_*) \\ +\infty, & \text{otherwise} \end{cases}
$$

So, we have:

$$
\begin{aligned}
g_d(\lambda) &= \min_{x \in E, y \in G} L(x, y, \lambda) = \\
&= \begin{cases} -g^*(\lambda) - f^*(-A^*\lambda) & \text{if } \lambda \in G_* \cap (-A^*)^{-1}(E_*) \\ -\infty, & \text{otherwise} \end{cases}
\end{aligned}
$$

which allows us to formulate one of the most important theorems, that connects dual problems and conjugate functions:

🤓 **Fenchel - Rockafellar theorem**

Let $f : E \to \mathbb{R}$ and $g : G \to \mathbb{R}$ — function, defined on the sets E and G in Euclidian Spaces V and W respectively. Let $f^* : E_* \to \mathbb{R}, g^* : G_* \to \mathbb{R}$ be the conjugate functions to the f and g respectively. Let $A : V \to W$ — linear mapping. Let $p^*, d^* \in [-\infty, +\infty]$ - optimal values of primal and dual problems:

$$p^* = f(x) + g(Ax) \to \min_{x \in E \cap A^{-1}(G)}$$

$$d^* = f^*(-A^*\lambda) + g^*(\lambda) \to \min_{\lambda \in G_* \cap (-A^*)^{-1}(E_*)},$$

Then we have weak duality: $p^* \geq d^*$. Furthermore, if the functions f and g are convex and $A(\mathbf{relint}(E)) \cap \mathbf{relint}(G) \neq \varnothing$, then we have strong duality: $p^* = d^*$. While points $x^* \in E \cap A^{-1}(G)$ and $\lambda^* \in G_* \cap (-A^*)^{-1}(E_*)$ are optimal values for primal and dual problem if and only if:

$$\begin{aligned} -A^*\lambda^* &\in \partial f(x^*) \\ \lambda^* &\in \partial g(Ax^*) \end{aligned}$$

Convex case is especially important since if we have Fenchel - Rockafellar problem with parameters (f, g, A), than the dual problem has the form $(f^*, g^*, -A^*)$.

5.2 Sensitivity analysis

Let us switch from the original optimization problem

$$\begin{aligned} &f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } &f_i(x) \leq 0, \ i = 1, \ldots, m \\ &h_i(x) = 0, \ i = 1, \ldots, p \end{aligned} \tag{P}$$

To the perturbed version of it:

$$\begin{aligned} &f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } &f_i(x) \leq u_i, \ i = 1, \ldots, m \\ &h_i(x) = v_i, \ i = 1, \ldots, p \end{aligned} \tag{Per}$$

Note, that we still have the only variable $x \in \mathbb{R}^n$, while treating $u \in \mathbb{R}^m, v \in \mathbb{R}^p$ as parameters. It is obvious, that $\text{Per}(u, v) \to \text{P}$ if $u = 0, v = 0$. We will denote the optimal value of Per as $p^*(u, v)$, while the optimal value of the original problem P is just p^*. One can immediately say, that $p^*(u, v) = p^*$.

Speaking of the value of some i-th constraint we can say, that

- $u_i = 0$ leaves the original problem
- $u_i > 0$ means that we have relaxed the inequality
- $u_i < 0$ means that we have tightened the constraint

One can even show, that when P is convex optimization problem, $p^*(u, v)$ is a convex function.

Suppose, that strong duality holds for the orriginal problem and suppose, that x is any feasible point for the perturbed problem:

$$\begin{aligned} p^*(0,0) = p^* = d^* = g(\lambda^*, \nu^*) \leq \\ \leq L(x, \lambda^*, \nu^*) = \\ = f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \leq \\ \leq f_0(x) + \sum_{i=1}^m \lambda_i^* u_i + \sum_{i=1}^p \nu_i^* v_i \end{aligned}$$

Which means

$$f_0(x) \geq p^*(0,0) - {\lambda^*}^T u - {\nu^*}^T v$$

And taking the optimal x for the perturbed problem, we have:

$$p^*(u,v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v \tag{4}$$

In scenarios where strong duality holds, we can draw several insights about the sensitivity of optimal solutions in relation to the Lagrange multipliers. These insights are derived from the inequality expressed in equation above:

1. **Impact of Tightening a Constraint (Large $\lambda_i^\star$):**
   When the ith constraint's Lagrange multiplier, $\lambda_i^\star$, holds a substantial value, and if this constraint is tightened (choosing $u_i < 0$), there is a guarantee that the optimal value, denoted by $p^\star(u,v)$, will significantly increase.

2. **Effect of Adjusting Constraints with Large Positive or Negative $\nu_i^\star$:**

   - If $\nu_i^\star$ is large and positive and $v_i < 0$ is chosen, or
   - If $\nu_i^\star$ is large and negative and $v_i > 0$ is selected,
     then in either scenario, the optimal value $p^\star(u,v)$ is expected to increase greatly.

3. **Consequences of Loosening a Constraint (Small $\lambda_i^\star$):**
   If the Lagrange multiplier $\lambda_i^\star$ for the ith constraint is relatively small, and the constraint is loosened (choosing $u_i > 0$), it is anticipated that the optimal value $p^\star(u,v)$ will not significantly decrease.

4. **Outcomes of Tiny Adjustments in Constraints with Small $\nu_i^\star$:**

   - When $\nu_i^\star$ is small and positive, and $v_i > 0$ is chosen, or
   - When $\nu_i^\star$ is small and negative, and $v_i < 0$ is opted for,
     in both cases, the optimal value $p^\star(u,v)$ will not significantly decrease.

These interpretations provide a framework for understanding how changes in constraints, reflected through their corresponding Lagrange multipliers, impact the optimal solution in problems where strong duality holds.

5.3 Local sensitivity

Suppose now that $p^*(u,v)$ is differentiable at $u=0, v=0$.

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i} \tag{5}$$

To show this result we consider the directional derivative of $p^*(u,v)$ along the direction of some i-th basis vector e_i:

$$\lim_{t\to 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} = \frac{\partial p^*(0,0)}{\partial u_i}$$

From the inequality Equation 4 and taking the limit $t \to 0$ with $t > 0$ we have

$$\frac{p^*(te_i,0) - p^*}{t} \geq -\lambda_i^* \to \frac{\partial p^*(0,0)}{\partial u_i} \geq -\lambda_i^*$$

For the negative $t < 0$ we have:

$$\frac{p^*(te_i,0) - p^*}{t} \leq -\lambda_i^* \to \frac{\partial p^*(0,0)}{\partial u_i} \leq -\lambda_i^*$$

The same idea can be used to establish the fact about v_i.

The local sensitivity result Equation 5 provides a way to understand the impact of constraints on the optimal solution x^* of an optimization problem. If a constraint $f_i(x^*)$ is negative at x^*, it's not affecting the optimal solution, meaning small changes to this constraint won't alter the optimal value. In this case, the corresponding optimal Lagrange multiplier will be zero, as per the principle of complementary slackness.

However, if $f_i(x^*) = 0$, meaning the constraint is precisely met at the optimum, then the situation is different. The value of the i-th optimal Lagrange multiplier, λ_i^*, gives us insight into how 'sensitive' or 'active' this constraint is. A small λ_i^* indicates that slight adjustments to the constraint won't significantly affect the optimal value. Conversely, a large λ_i^* implies that even minor changes to the constraint can have a significant impact on the optimal solution.

5.4 Shadow prices or tax interpretation

Consider an enterprise where x represents its operational strategy and $f_0(x)$ is the operating cost. Therefore, $-f_0(x)$ denotes the profit in dollars. Each constraint $f_i(x) \leq 0$ signifies a resource or regulatory limit. The goal is to maximize profit while adhering to these limits, which is equivalent to solving:

$$
\begin{split}
& f_0(x) \to \min\limits_{x \in \mathbb{R}^n}\\
\text{s.t. } & f_i(x) \leq 0, \; i = 1,\ldots,m
\end{split}
$$

The optimal profit here is $-p^*$.

Now, imagine a scenario where exceeding limits is allowed, but at a cost. This cost is linear to the extent of violation, quantified by f_i. The charge for breaching the i^{th} constraint is $\lambda_i f_i(x)$. If $f_i(x) < 0$, meaning the constraint is not fully utilized, $\lambda_i f_i(x)$ represents income for the firm. Here, λ_i is the cost (in dollars) per unit of violation for $f_i(x)$.

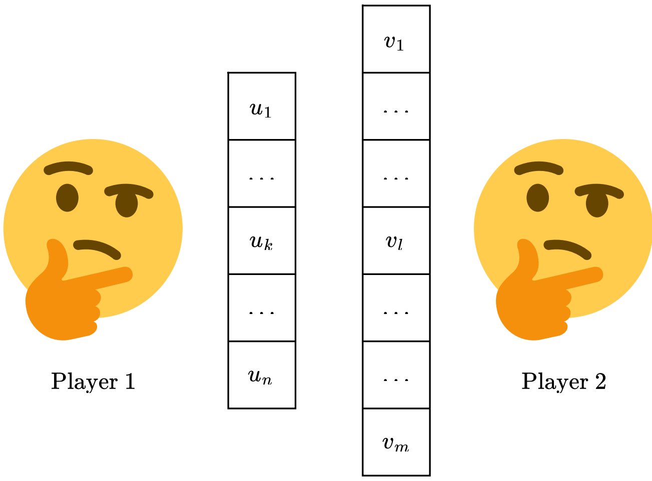
For instance, if $f_1(x) \leq 0$ limits warehouse space, the firm can rent out extra space at λ_1 dollars per square meter or rent out unused space for the same rate.

The firm's total cost, considering operational and constraint costs, is $L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$. The firm aims to minimize $L(x, \lambda)$, resulting in an optimal cost $g(\lambda)$. The dual function $g(\lambda)$ represents the best possible cost for the firm based on the prices of constraints λ, and the optimal dual value d^* is this cost under the most unfavorable price conditions.

Weak duality implies that the cost in this flexible scenario (where the firm can trade constraint violations) is always less than or equal to the cost in the strict original scenario. This is because any optimal operation x^* from the original scenario will cost less in the flexible scenario, as the firm can earn from underused constraints.

If strong duality holds and the dual optimum is reached, the optimal λ^* represents prices where the firm gains no extra advantage from trading constraint violations. These optimal λ^* values are often termed 'shadow prices' for the original problem, indicating the hypothetical cost of constraint flexibility.

5.5 Mixed strategies for matrix games



The scheme of a mixed strategy matrix game

In zero-sum matrix games, players 1 and 2 choose actions from sets $\{1,\ldots,n\}$ and $\{1,\ldots,m\}$, respectively. The outcome is a payment from player 1 to player 2, determined by a payoff matrix $P \in \mathbb{R}^{n \times m}$. Each player aims to use mixed strategies, choosing actions according to a probability distribution: player 1 uses probabilities u_k for each action i, and player 2 uses v_l.

The expected payoff from player 1 to player 2 is given by $\sum_{k=1}^n \sum_{l=1}^m u_k v_l P_{kl} = u^T P v$. Player 1 seeks to minimize this expected payoff, while player 2 aims to maximize it.

5.5.1 Player 1's Perspective

Assuming player 2 knows player 1's strategy u, player 2 will choose v to maximize $u^T P v$. The worst-case expected payoff is thus:

$$\max_{v \geq 0, 1^T v = 1} u^T P v = \max_{i=1,\ldots,m} (P^T u)_i$$

Player 1's optimal strategy minimizes this worst-case payoff, leading to the optimization problem:

$$\begin{aligned} &\min \max_{i=1,\ldots,m} (P^T u)_i \\ &\text{s.t. } u \geq 0 \\ &1^T u = 1 \end{aligned} \tag{6}$$

This forms a convex optimization problem with the optimal value denoted as p_1^*.

5.5.2 Player 2's Perspective

Conversely, if player 1 knows player 2's strategy v, the goal is to minimize $u^T P v$. This leads to:

$$\min_{u \geq 0, 1^T u = 1} u^T P v = \min_{i=1,\ldots,n} (P v)_i$$

Player 2 then maximizes this to get the largest guaranteed payoff, solving the optimization problem:

$$\begin{aligned} &\max \min_{i=1,\ldots,n} (P v)_i \\ &\text{s.t. } v \geq 0 \\ &1^T v = 1 \end{aligned} \tag{7}$$

The optimal value here is p_2^*.

5.5.3 Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

5.5.4 Formulating and Solving the Lagrange Dual

We approach problem Equation 6 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

5.5.5 Constructing the Lagrangian

We introduce multipliers for the constraints: λ for $P^T u \leq t\mathbf{1}$, μ for $u \geq 0$, and ν for $1^T u = 1$. The Lagrangian is then formed as:

$$L = t + \lambda^T (P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - 1^T u) = \nu + (1 - 1^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u$$

5.5.6 Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T \lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

5.5.7 Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T \lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu \mathbf{1} = \mu$.

Upon eliminating μ, we obtain the Lagrange dual of Equation 6:

$$
\begin{aligned}
&\max \nu \\
&\text{s.t. } \lambda \geq 0 \\
&\lambda \geq 0 \\
&P\lambda \geq \nu \mathbf{1}
\end{aligned}
$$

5.5.8 Conclusion

This formulation shows that the Lagrange dual problem is equivalent to problem Equation 7. Given the feasibility of these linear programs, strong duality holds, meaning the optimal values of Equation 6 and Equation 7 are equal.

6 References

- Convex Optimization — Boyd & Vandenberghe @ Stanford
- Course Notes for EE227C. Lecture 13
- Course Notes for EE227C. Lecture 14
- Optimality conditions
- Seminar 7 @ CMC MSU
- Seminar 8 @ CMC MSU
- Convex Optimization @ Berkeley - 10th lecture