# 1 Overview

In the previous lectures, we studied unconstrained convex optimization. In this lecture we will begin our discussion about constrained optimization, and in particular introduce Lagrangian multipliers and Lagrangian duality. In general, strong duality does not always hold, but we will introduce *Slater's condition* which guarantees strong duality holds.

# 2 Convex Optimization under Constraints

Consider the following problem of minimizing $x+y$ s.t. $x^2+y^2 \leq 1$. How should we go about solving this problem? In our discussions thus far we dealt with linear optimization under linear constraints (or problems that can be reduced to that, such as piecewise linear functions), or convex optimization where there we no constraints. This problem involves minimizing a linear (and hence convex) function under *convex* constraints. Note that the set of feasible points $(x, y)$ where $x^2 + y^2 \leq 1$ is convex. For this particular problem, the constraint will be active for an optimal solution, so in fact $x^2 + y^2 = 1$ for optimal points.

For the problem above we know that $x + y$ is both convex and concave and therefore, if there hadn't been any constraints the stationary point $(x, y)$ for which $\nabla f(x, y) = 0$ would be an optimum. But we somehow need to understand how we can make sure the constraint $x^2 + y^2 = 1$ is met, or equivalently $x^2 + y^2 - 1 = 0$. One way of doing this is to consider the problem of optimizing the unconstrained function, using $\lambda > 0$:

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

For now, let's find a stationary point $\nabla L(x, y, \lambda) = 0$ and see what we can obtain. In this case:

$$\nabla L = \begin{pmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} 1 + 2\lambda x \\ 1 + 2\lambda y \\ x^2 + y^2 - 1 \end{pmatrix}$$

Solving for $\nabla L(x, y, \lambda) = (0, 0, 0, )$, from the first two equations we get that as long as $\lambda \neq 0$ we have that $x = y = -\frac{1}{2\lambda}$, and substituting $x$ and $y$ into the last equation we get that $\lambda = \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$. Therefore the stationary points are points are $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ and $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$.

## 2.1 Lagrangian Multipliers

The above exercise was an example of using *Lagrange multipliers*.

**Definition.** *Given the optimization problem:*

$$\begin{aligned} &\min\ f(\boldsymbol{x}) & \\ &s.t.\ g_i(\mathbf{x}) \leq 0 & \forall i \in [m] \\ &\quad\ \ h_j(\mathbf{x}) = 0 & \forall j \in [p] \end{aligned}$$

*The **Lagrangian** associated with the optimization problem is:*

$$\min L(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

*The variables $\lambda_1, \ldots, \lambda_m$ and $\nu_1, \ldots, \nu_p$ are called the **Lagrangian multipliers**.*

**Example: least squares.** Consider the following program:

$$\begin{aligned} &\min\ \mathbf{x}^\intercal \mathbf{x} \\ &\text{s.t. } A\mathbf{x} = \mathbf{b} \end{aligned}$$

where $A \in \mathbb{R}^{p \times n}$. There are no inequality constraints, and $p$ linear equality constraints. The Lagrangian here is:

$$L(\mathbf{x}, \nu) = \mathbf{x}^\intercal \mathbf{x} + \nu^\intercal (A^\intercal \mathbf{x} - \mathbf{b}).$$

**Example: linear optimization.** Consider a linear program:

$$\begin{aligned} &\min\ \mathbf{c}^\intercal \mathbf{x} \\ &\text{s.t. } A\mathbf{x} = \mathbf{b} \\ &\quad\ \ \mathbf{x} \geq \mathbf{0} \end{aligned}$$

In this case the Lagrangian is:

$$L(\mathbf{x}, \nu) = \mathbf{c}^\intercal \mathbf{x} - \lambda^\intercal \mathbf{x} + \nu^\intercal (A^\intercal \mathbf{x} - \mathbf{b}) = -\mathbf{b}^\intercal \nu + (\mathbf{c} + A^\intercal \nu - \lambda)^\intercal \mathbf{x}$$

## 2.2 Lagrangian duality

When we discussed linear optimization, we saw that the concept of duality was quite powerful, and eventually enabled us to solve linear optimization problems efficiently. In convex optimization we have a generalization of this idea, called *Lagrangian duality*. For an objective $L$ as the one stated above, the dual function is defined as follows.

**Definition.** *Given a Lagrangian $L(\mathbf{x}, \lambda, \nu)$ of some optimization problem over domain $\mathcal{D}$, the **Lagrangian dual** is the function:*

$$F(\lambda, \nu) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \nu).$$

**Example: Lagrangian duality of least squares.** The Lagrangian dual function is:

$$F(\nu) = \inf_{\mathbf{x}} L(\mathbf{x}, \nu) = \inf_{\mathbf{x}} \{\mathbf{x}^\intercal \mathbf{x} + \nu^\intercal (A\mathbf{x} - \mathbf{b})\}$$

Here the Lagrangian is convex in $\mathbf{x}$, and we can therefore find its infimum via a stationary point:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \nu) = 2\mathbf{x} + \nu^\intercal A$$

the stationary point is:

$$\begin{aligned} & \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \nu) = \mathbf{0} \\ \iff & 2\mathbf{x} + \nu^\intercal A = \mathbf{0} \\ \iff & \mathbf{x} = -\frac{1}{2}\nu^\intercal A \end{aligned}$$

The Lagrangian dual is:

$$F(\nu) = \inf_{\mathbf{x}} L(\mathbf{x}, \nu) = L\left(-\frac{1}{2}\nu^\intercal A, \nu\right) = \frac{1}{4}\nu^\intercal A A^\intercal \nu - \frac{1}{2}\nu^\intercal A A^\intercal \nu - \nu^\intercal \mathbf{b} = -\frac{1}{4}\nu^\intercal A A^\intercal \nu - \nu^\intercal \mathbf{b}$$

**Example: Largrangian duality of linear optimization.** The Lagrangian dual function is:

$$F(\lambda, \nu) = \inf_{\mathbf{x}} L(\mathbf{x}, \lambda, \nu) = \inf_{\mathbf{x}} \{-\mathbf{b}^\intercal \nu + (\mathbf{c} + A^\intercal \nu - \lambda)^\intercal \mathbf{x}\} = -\mathbf{b}^\intercal \nu + \inf_{\mathbf{x}} \{(\mathbf{c} + A^\intercal \nu - \lambda)^\intercal \mathbf{x}\}$$

Since the function $(\mathbf{c} + A^\intercal \nu - \lambda)^\intercal \mathbf{x}$ is linear, it is bounded from below only when it is identically zero. Thus:

$$F(\lambda, \nu) = \begin{cases} -\mathbf{b}^\intercal \nu & \mathbf{c} + A^\intercal \nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

If we replace $\nu$ with $-\mathbf{y}$ we get:

$$F(\lambda, \nu) = \begin{cases} \mathbf{b}^\intercal \mathbf{y} & A^\intercal \mathbf{y} \leq \mathbf{c} \\ -\infty & \text{otherwise} \end{cases}$$

Maximizing the Lagrangian dual function thus translates to the following optimization problem:

$$\begin{aligned} & \max \mathbf{b}^\intercal \mathbf{y} \\ & \text{ s.t. } A^\intercal \mathbf{y} \leq \mathbf{c} \end{aligned}$$

which is the dual problem of the linear program.

# 3 The Dual Optimization Problem

**Definition.** *The constrained optimization problem:*

$$\begin{aligned} \min\ & f(\boldsymbol{x}) & \\ s.t.\ & g_i(\mathbf{x}) \leq 0 & \forall i \in [m] \\ & h_j(\mathbf{x}) = 0 & \forall j \in [p] \end{aligned}$$

*will henceforth be referred to as the* ***primal optimization problem****.*

**Definition.** *Given a primal optimization problem, the* ***dual optimization problem*** *is:*

$$\max F(\lambda, \nu)$$
$$s.t. \lambda \geq 0$$

*where $F(\lambda, \nu)$ is the* Lagrangian dual *function associated with the function $f$ above.*

# 4 Weak Duality

From here on we will frequently use $\alpha^\star$ to denote the value of the optimal solution for the primal problem, and $\beta^\star$ to denote the optimal solution for the dual problem.

**Theorem 1.** *Let $\alpha^\star$ be the value of the optimal solution for the primal problem, and $\beta^\star$ be the value of the optimal solution for the dual problem. Then: $\beta^\star \leq \alpha^\star$.*

*Proof.* Let $\mathcal{D}$ denote the feasible region of the primal problem.

$$F(\lambda, \nu) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \nu) = \inf_{\mathbf{x} \in \mathcal{D}} \Big( f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \Big)$$

For any feasible point $\mathbf{x} \in \mathcal{D}$ we have that $g_i(\mathbf{x}) \leq 0$, for all $i \in [m]$, and $h_i(\mathbf{x}) = 0$, for all $i \in [p]$. Since $\lambda_i$ are all nonnegative, we have that:

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \leq f(\mathbf{x})$$

Since this holds for all $\lambda, \nu$, in particular it holds for the dual optimal values $(\lambda^\star, \nu^\star)$, i.e. the values for which $F(\lambda^\star, \nu^\star) = \max_{\lambda,\nu} F(\lambda, \nu) = \beta^\star$. Thus:

$$\beta^\star = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \lambda^\star, \nu^\star) \leq \inf_{\mathbf{x} \in D} f(\mathbf{x}) = \alpha^\star.$$ □

# 5 Slater's Condition and Strong Duality

In linear optimization we proved that we always have strong duality. That is, when the functions and constraints are linear we know that $\alpha^\star = \beta^\star$. For general optimization problems this is not necessarily the case. In this lecture and the lecture following next we will discuss conditions under which strong duality holds. The condition we will discuss today is called *Slater's condition.*

**Definition.** *For a primal optimization problem, we say that it respects* ***Slater's condition*** *if the objective function $f$ is convex, the constraint functions $\{g_i\}_{i=1}^m$ are convex, the constraint functions $\{h_j\}_{j=1}^p$ are linear, and there exists a point $\bar{\mathbf{x}}$ in the interior of the region, i.e.*

1. *$\bar{\mathbf{x}} \in D$ and $x$ is not in the boundary of $D$ (a small ball around $\bar{\mathbf{x}}$ is also in $D$).*
2. *$g_i(\bar{\mathbf{x}}) < 0$ for all $i \in [m]$,*
3. *$h_j(\bar{\mathbf{x}}) = 0$ for all $j \in [p]$.*

**Theorem 2.** *Suppose that Slater's condition hold and the region has a non-empty interior. Then, we have strong duality.*

Before proving the theorem, we will prove the following lemma.

**Lemma 3.** *Let $\alpha^\star$ be the value of the optimal solution for the primal problem and define the sets:*

- $\mathcal{A} = \{(\mathbf{u}, \mathbf{v}, t) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \ : \ \exists \mathbf{x} \in \mathcal{D}, g_i(\mathbf{x}) \le u_i \ \forall i \in [m], h_j(\mathbf{x}) = v_j \ \forall j \in [p], f(\mathbf{x}) \le t\}$
- $\mathcal{B} = \{(\mathbf{0}, \mathbf{0}, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \ : s < \alpha^\star \}$

*Then the sets are convex and $\mathcal{A} \cap \mathcal{B} = \emptyset$.*

*Proof.* Verifying that these sets are convex follows from their definition. To see that $\mathcal{A}$ and $\mathcal{B}$ do not intersect, assume for purpose of contradiction that there exists a point in $\mathcal{A} \cap \mathcal{B}$. Since the point is in $\mathcal{A}$ there exists a point $\mathbf{x}'$ s.t. the $g_i(\mathbf{x}') \le u_i$ and $h_j(\mathbf{x}') \le v_j$ for all $i \in [m]$ and $j \in [p]$. Since the point is also in $\mathcal{B}$ it must be the case that $u_i, v_j = 0$ for all $i \in [m]$ and $j \in [p]$. Thus, $\mathbf{x}'$ is feasible. The fact that the point is in $\mathcal{B}$ implies that $f(\mathbf{x}') < \alpha^\star$. But since $\alpha^\star$ is the minimal value obtained by a feasible point, this is a contradiction. $\square$

*Proof of strong duality under Slater's condition.* Without loss of generality assume that the rank of $A$ is $p$ and that the primal objective is finite (otherwise $\alpha^\star = -\infty$ and then by weak duality $\beta^\star = -\infty$). Since $\mathcal{A}$ and $\mathcal{B}$ are convex and do not intersect we can apply the separating hyperplane theorem. In this case this implies that there exists a point $(\tilde{\lambda}, \tilde{\nu}, \tilde{\mu}) \neq 0$ and value $\alpha$ s.t.:

- $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{A} \implies \tilde{\lambda}^\intercal \mathbf{u} + \tilde{\nu}^\intercal \mathbf{v} + \tilde{\mu} t \ge \alpha$
- $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{B} \implies \tilde{\lambda}^\intercal \mathbf{u} + \tilde{\nu}^\intercal \mathbf{v} + \tilde{\mu} t \le \alpha$

Since $\alpha$ is a lower bound on $\tilde{\lambda}^\intercal \mathbf{u} + \tilde{\nu}^\intercal \mathbf{v} + \tilde{\mu} t$ for points $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{A}$ we can conclude that $\tilde{\lambda} \ge 0$ and $\tilde{\mu} \ge 0$ as otherwise we have that as $(\mathbf{u}', \mathbf{v}, t') \in \mathcal{A}$ for any $\mathbf{u}' \ge \mathbf{u}$ and $t' \ge t$, that $\tilde{\lambda}^\intercal \mathbf{u} + \tilde{\nu}^\intercal \mathbf{v} + \tilde{\mu} t$ can be made arbitrarily small (goes to $-\infty$), and this contradicts the lower bound of $\alpha$. Since the points in $\mathcal{B}$ are those for which $\mathbf{u} = \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$, the fact that $\alpha$ is an upper bound on $\tilde{\lambda}^\intercal \mathbf{u} + \tilde{\nu}^\intercal \mathbf{v}$ for points $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{B}$ implies that $\tilde{\mu} t \le \alpha$ for all $t < \alpha^\star$, and thus $\tilde{\mu} \alpha^\star \le \alpha$. Together we get that for any $\mathbf{x}$ in our domain:

$$\sum_{i=1}^{m} \tilde{\lambda}_i g_i(\mathbf{x}) + \tilde{\nu}^\intercal (A\mathbf{x} - \mathbf{b}) + \tilde{\mu} f(\mathbf{x}) \ge \alpha \ge \tilde{\mu} \alpha^\star$$

If $\tilde{\mu} > 0$ we can simply divide by $\tilde{\mu}$ and get that:

$$L(\mathbf{x}, \tilde{\lambda}/\tilde{\mu}, \tilde{\nu}/\tilde{\mu}) = \sum_{i=1}^{m} \frac{\tilde{\lambda}_i}{\tilde{\mu}} g_i(\mathbf{x}) + \left(\frac{\tilde{\nu}}{\tilde{\mu}}\right)^\intercal (A\mathbf{x} - \mathbf{b}) + f(\mathbf{x}) \ge \alpha^\star$$

In particular this also holds for the point that minimizes $L(\cdot, \tilde{\lambda}/\tilde{\mu}, \tilde{\nu}/\tilde{\mu})$:

$$\bar{\mathbf{x}} \in \arg\inf_{\mathbf{x}} L(\mathbf{x}, \tilde{\lambda}/\tilde{\mu}, \tilde{\nu}/\tilde{\mu}) = F(\tilde{\lambda}/\tilde{\mu}, \tilde{\nu}/\tilde{\mu}) \le \beta^\star$$

Let $\lambda = \tilde{\lambda}/\tilde{\mu}$ and $\nu = \tilde{\nu}/\tilde{\mu}$. By weak duality we have $F(\lambda, \nu) \leq \alpha^\star$, so in fact $F(\lambda, \nu) = \alpha^\star$. This shows that strong duality holds, and that the dual optimum is attained, in the case when $\tilde{\mu} > 0$.

In the case that $\tilde{\mu} = 0$, the inequality derived from the departing hyperplane theorem implies:

$$\sum_{i=1}^{m} \tilde{\lambda}_i g_i(\mathbf{x}) + \tilde{\nu}^\intercal (A\mathbf{x} - \mathbf{b}) \geq 0 \tag{1}$$

Consider the point $\tilde{\mathbf{x}}$ which satisfies Slater's condition. In this case since this point respects $A\tilde{\mathbf{x}} = \mathbf{b}$ we have that:

$$\sum_{i=1}^{m} \tilde{\lambda}_i g_i(\tilde{\mathbf{x}}) \geq 0$$

Since $g_i(\tilde{\mathbf{x}}) \leq 0$ and $\tilde{\lambda}_i \geq 0$ for all $i \in [m]$, this implies that $\tilde{\lambda} = 0$. From (1) this now implies that for all $\mathbf{x}$ we have:

$$\tilde{\nu}^\intercal (A\mathbf{x} - b) \geq 0$$

By the separating hyperplane theorem we know that $(\tilde{\lambda}, \tilde{\nu}, \tilde{\mu}) \neq 0$, and since $\tilde{\lambda} = 0$ and $\tilde{\mu} = 0$ this implies that $\tilde{\nu} \neq 0$. The point $\tilde{\mathbf{x}}$ satisfies $\nu^\intercal (A\tilde{\mathbf{x}} - \mathbf{b}) = 0$, and since $\tilde{\mathbf{x}}$ is in the interior of $D$, we know that there are points $\mathbf{x}$ for which $\tilde{\nu}^\intercal (A\mathbf{x} - b) < 0$ unless $A^\intercal \nu = 0$. But this contradicts our assumption that $A$ is of rank $p$. □

## 6 Discussion and Further Reading

This lecture is based on Chapter 5 from [1]. For more examples, applications, and interpretations of duality see Chapter 5 in [1].

## References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, New York, NY, USA, 2004.