 

# Lagrangian Duality

In the previous lecture, we considered constrained minimization problems of the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0},$$

where we denote by $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ is the vector of $m$ inequality constraints, interpreting the inequality element-wise. We showed (KKT conditions) that a point $\mathbf{x}^*$ is a constrained minimizer of the above problem, if it is *regular* and can be associated with a set of Lagrange multipliers $\boldsymbol{\lambda}^* \ge \mathbf{0}$ such that $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$, where

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{g}(\mathbf{x})$$

is the *Lagrangian* associated with the problem.

In what follows, we will have a deeper look at the properties of the Lagrangian, showing dual formulations of constrained problems.

## 1 Min-max inequality

Let us consider some function $f(\mathbf{x}, \boldsymbol{\lambda})$ on $\mathbb{R}^{n+m}$ (for the time being, we are not connecting the discussion to the Lagrangian). Fixing $\boldsymbol{\lambda}$ and minimizing $f$ over $\mathbf{x}$ yields a function of $\boldsymbol{\lambda}$,

$$p(\boldsymbol{\lambda}) = \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}).$$

For a fixed $\boldsymbol{\lambda}$, $p(\boldsymbol{\lambda})$ has the smallest value than $f(\mathbf{x}, \boldsymbol{\lambda})$ for any $\mathbf{x}$; consequently, the following inequality is straightforward for every $\mathbf{x}$:

$$f(\mathbf{x}, \boldsymbol{\lambda}) \ge p(\boldsymbol{\lambda}).$$

Note that the left-hand-side has two independent variables ($\mathbf{x}$ and $\boldsymbol{\lambda}$), while the right-hand-side only one ($\boldsymbol{\lambda}$).

Let us now fix $\mathbf{x}$ and define

$$q(\mathbf{x}) = \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}).$$

Due to the former inequality, we have

$$q(\mathbf{x}) = \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}) \ge \max_{\boldsymbol{\lambda}} p(\boldsymbol{\lambda}).$$

Note that now the left-hand-side has only one independent variable ($\mathbf{x}$), while the right-hand-side is a constant.

Since the inequality holds for every $\mathbf{x}$, it also holds for one minimizing the left-hand-size:

$$\min_{\mathbf{x}} q(\mathbf{x}) \ge \max_{\boldsymbol{\lambda}} p(\boldsymbol{\lambda}).$$

Substituting $p$ and $q$ explicitly yields

**Theorem 1** (Min-max inequality)**.**

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}) \geq \max_{\boldsymbol{\lambda}} \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}).$$

The theorem can be also formulated in terms of inf replacing min and sup replacing max.

# 2 Min-max and game theory

This result can be applied in game theory (from which it actually comes). Let there be two players, Eric and Laura playing a game in which players alternatingly make a move; denote the move performed by Eric as $x \in X$, and that of Laura as $\lambda \in \Lambda$. After both players made their moves, Laura receives from Eric the amount described by the "payoff" function $f(x, \lambda)$ (if the amount is negative, Laura has to reach out for the wallet).

When Laura plays *second*, Eric has already made his move $x$, so Laura's goal is to maximize her payoff by choosing

$$\lambda = \arg\max_{\lambda} f(x, \lambda),$$

and her payoff is

$$q(x) = \max_{\lambda} f(x, \lambda)$$

(note that the payoff depends on Eric's move). Eric is of course aware of this strategy, so he tries to make a move that will minimize Laura's best payoff:

$$x = \arg\min_{x} q(x) = \min_{x} \max_{\lambda} f(x, \lambda).$$

When Laura plays *first*, she knows that Eric will make a move after her to minimize her payoff (hence, maximizing his own payoff, since it's a zero-sum game)

$$x = \arg\min_{x} f(x, \lambda),$$

and her payoff will be

$$p(\lambda) = \min_{x} f(x, \lambda)$$

(note that Laura's payoff depends on her move). So Laura will try to maximize it by choosing

$$\lambda = \arg\max_{\lambda} p(\lambda) = \max_{\lambda} \min_{x} f(x, \lambda).$$

The min-max inequality tells us that

$$\min_{x} \max_{\lambda} f(x, \lambda) \geq \max_{\lambda} \min_{x} f(x, \lambda),$$

which in our terms means that whoever plays second is always in an advantageous situation (or, more accurately, is never in a disadvantageous situation).

# 3 Saddle points

Cases where the min-max inequality holds with equality (sometimes referred to as strong min-max) have an important relation to saddle points.

**Definition** (Saddle point). *A point* $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^n \times \mathbb{R}^m$ *is called a* saddle point *of* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *if for every* $\mathbf{x} \in \mathbb{R}^n$ *and* $\boldsymbol{\lambda}) \in \mathbb{R}^m$*,*

$$f(\mathbf{x}^*, \boldsymbol{\lambda}) \leq f(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq f(\mathbf{x}, \boldsymbol{\lambda}^*).$$

A saddle point is a stationary point of a function that is neither a minimum nor a maximum. For smooth functions, this happens if the gradient vanishes but there are directions of both positive and negative curvature (the Hessian has positive and negative eigenvalues). As a simple example in $\mathbb{R} \times \mathbb{R}$, consider $f(x, \lambda) = x^2 - \lambda^2$.

**Theorem 2** (Strong min-max). *If* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *satisfied the saddle-point condition*

$$f(\mathbf{x}^*, \boldsymbol{\lambda}) \leq f(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq f(\mathbf{x}, \boldsymbol{\lambda}^*),$$

*then*

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}} \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}).$$

To prove this result, let us consider the left-hand-side inequality in the saddle-point condition, $f(\mathbf{x}^*, \boldsymbol{\lambda}) \leq f(\mathbf{x}^*, \boldsymbol{\lambda}^*)$:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}) \leq \max_{\boldsymbol{\lambda}} f(\mathbf{x}^*, \boldsymbol{\lambda}) = f(\mathbf{x}^*, \boldsymbol{\lambda}^*).$$

Similarly, by the right-hand-side inequality $f(\mathbf{x}, \boldsymbol{\lambda}^*) \geq f(\mathbf{x}^*, \boldsymbol{\lambda}^*)$,

$$\max_{\boldsymbol{\lambda}} \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}) \geq \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*, \boldsymbol{\lambda}^*).$$

Combining the two results yields

$$\max_{\boldsymbol{\lambda}} \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\lambda}) \geq \min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} f(\mathbf{x}, \boldsymbol{\lambda}).$$

Note that the inequality is reversed compared to the min-max inequality, from where we get the equality.

# 4 Weak duality

Instead of an abstract $f(\mathbf{x}, \boldsymbol{\lambda})$, let us now consider explicitly the Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{g}(\mathbf{x}).$$

Let us fix $\mathbf{x}$ and maximize $L(\mathbf{x}, \boldsymbol{\lambda})$ over all $\boldsymbol{\lambda} \geq \mathbf{0}$. First observe that the first term is constant in $\boldsymbol{\lambda}$. Regarding the second term, if $\mathbf{x}$ is feasible, all constraints hold, $g_i(\mathbf{x}) \leq 0$. Hence, the second term is non-positive, and the highest value it can achieve is 0. If $\mathbf{x}$ is infeasible, than for at least of $i$, $g_i(\mathbf{x}) > 0$ and we can make the second term infinitely big by choosing $\lambda_i \to \infty$. In other words,

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} f(\mathbf{x}) & : \ \mathbf{x} \text{ is feasible} \\ \infty & : \ \text{otherwise.} \end{cases}$$

Note that this is exactly the ideal penalty aggregate, $\max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}) = F_\infty(\mathbf{x})$, so we can rewrite our constrained minimization problem as

$$f^* = \min_{\mathbf{x}} F_\infty(\mathbf{x}) = \min_{\mathbf{x}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}).$$

Invoking the min-max inequality,

$$f^* = \min_{\mathbf{x}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}) \geq \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \eta(\boldsymbol{\lambda}) = \eta^*,$$

where

$$\eta(\boldsymbol{\lambda}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda})$$

is usually called the *dual function* (we can define this function of $\boldsymbol{\lambda} \geq \mathbf{0}$).

The maximization

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} \eta(\boldsymbol{\lambda})$$

is referred to as the *dual problem*. The dual problem is always concave (the analogy of convex for maximization) even if the original problem is not convex.

Following this notation, the original objective $f(\mathbf{x})$ is called the *primal function*, and the original problem the *primal problem*. The difference $f^* - \eta^*$ is called the *duality gap* and is always non-negative, as the maximum $\eta^*$ of the dual function gives the lower bound to the minimum of the primal $f^*$. This result is usually referred to as *weak duality*.

# 5 Strong duality

For particular problems, the strong min-max inequality holds and the duality gap becomes exactly zero. Such cases are referred to as strong duality:

**Definition.** *Strong duality holds iff the duality gap is zero.*

If strong duality holds, the primal and the dual solutions are equivalent.

**Theorem 3** (Strong duality)**.** *Strong duality holds if $f$ and $g_i$ are convex and the KKT conditions are satisfied.*

To prove this result, recall first that if $f$ and all the $g_i$'s are convex, the Lagrangian is also convex in $\mathbf{x}$ being a non-negative linear combination of convex functions. By the KKT condition, at the solution point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$; since $L$ is convex in $\mathbf{x}$, this first-order condition is sufficient. Hence,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*),$$

from where for every $\mathbf{x}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}^*) \geq L(\mathbf{x}^*, \boldsymbol{\lambda}^*).$$

Note that the second term $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{g}(\mathbf{x})$ in the Lagrangian is non-negative at the (feasible) solution point $\mathbf{x}^*$. Hence, for every $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{g}(\mathbf{x}^*) \leq 0$. From KKT condition (complementary slackness), $\boldsymbol{\lambda}^{*\mathrm{T}}\mathbf{g}(\mathbf{x}^*) = 0$. This implies

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*).$$

Combining the two inequalities leads to

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*)$$

meaning that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of the Lagrangian. Therefore, by the saddle-point theorem,

$$f^* = \min_{\mathbf{x}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \eta(\boldsymbol{\lambda}) = \eta^*.$$

As a final remark, recall that the KKT conditions required a *regular* solution; this condition was satisfied by linear independence of the constraint gradients at the solution point. This means that we cannot know if strong duality holds before finding the solution. The following very useful result circumvents this drawback:

**Theorem 4** (Slater)**.** *Strong duality holds for a convex problem (i.e., $f$ and $\mathbf{g}$ are convex) if there exists a point $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x}) < \mathbf{0}$ (i.e., all constraints are inactive).*

Linear inequality as well as equality constraints can be excluded from Slater's condition.

# 6 Examples of dual problems

**Quadratic program.** As our first example, consider the particular form of a quadratic problem with linear inequality constraints (such problems are usually known as (*linearly constrained*) *quadratic programs* or QPs for short):

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{0},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

The Lagrangian of the problem is

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b})$$

with $\boldsymbol{\lambda} \in \mathbb{R}^m_+$. Let us fix $\boldsymbol{\lambda}$ and express the minimizer of the Lagrangian

$$\mathbf{x}^*_{\boldsymbol{\lambda}} = \arg\min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}).$$

Since the Lagrangian is convex in $\mathbf{x}$, it is sufficient to demand

$$0 = \nabla_{\mathbf{x}} L(\mathbf{x}^*_{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) = \mathbf{x}^*_{\boldsymbol{\lambda}} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda},$$

from where

$$\mathbf{x}^*_{\boldsymbol{\lambda}} = -\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}.$$

Substituting into $L$, yields

$$\begin{aligned} \eta(\boldsymbol{\lambda}) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = L(\mathbf{x}^*_{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) \\ &= \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} \\ &= -\frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}. \end{aligned}$$

The dual problem is, therefore,

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} -\frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}.$$

**Linear program.**

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad -(\mathbf{A}\mathbf{x} - \mathbf{b}) \leq \mathbf{0}.$$

As before, we write the Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} - \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

and equate its gradient w.r.t. $\mathbf{x}$ to zero:

$$0 = \nabla_{\mathbf{x}} L(\mathbf{x}^*_{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) = \mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}.$$

Note that unlike the case of QP, now the result does not depend on $\mathbf{x}$ at all – there is no unique minimizer! When $\mathbf{c} = \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}$, any $\mathbf{x}$ minimizes the Lagrangian; substituting this condition into the Lagrangian yields the minimum

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}.$$

If the condition does not hold, in the above function the linear term in $\mathbf{x}$ is not canceled. As the result, we can choose a descent direction of the function and make $\mathbf{x}$ in that direction arbitrarily big. In other words, $L(\mathbf{x}, \boldsymbol{\lambda})$ is unbounded, and we can summarize the two cases as

$$\eta(\boldsymbol{\lambda}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} & : \quad \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{c} \\ -\infty & : \quad \text{otherwise.} \end{cases}$$

Note, however, that this is exactly the ideal penalty aggregate for the maximization of $\eta(\boldsymbol{\lambda})$ s.t. the constraint $\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{c}$, so we can formulate the dual problem as

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} \quad \text{s.t.} \quad \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{c}.$$