Local First-Order Optimality Conditions

- Convex optimization problems

Constraint qualifications are not needed for a *sufficient* condition for general *convex optimization problems*.

Suppose that $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions with f and g_i convex and h_j (affinely) linear ($i \in \mathcal{I}, j \in \mathcal{E}$), and consider the following convex optimization problem:

$$(CP) \begin{cases} \min\limits_{\mathbf{x}} \ f(\mathbf{x}), \\ g_i(\mathbf{x}) \le 0, \ \text{for } i \in \mathcal{I} \\ h_j(\mathbf{x}) = 0, \ \text{for } j \in \mathcal{E}. \end{cases}$$

We will show that *for this special kind of problem every KKT point already gives a (global) minimum.*

Local First-Order Optimality Conditions

- Convex optimization problems

Theorem 2. *Suppose* $\mathbf{x}_0 \in \mathcal{F}$ *and there exist* $\boldsymbol{\lambda} \in \mathbb{R}^m_+$ *and* $\boldsymbol{\mu} \in \mathbb{R}^p$ *such that*

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_i \nabla h_j(x_0) = \mathbf{0} \text{ and}$$

$$\lambda_i g_i(\mathbf{x}_0) = 0 \text{ for } i = 1, ..., m,$$

then (CP) attains its global minimum at $\mathbf{x}_0$.

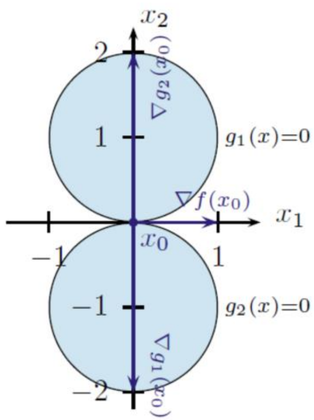
Proof. By convexity, we get for $\mathbf{x} \in \mathcal{F}$:

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &\underset{f \text{ convex}}{\geq} f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\ &= -\sum_{i=1}^{m} \lambda_i g_i'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) - \sum_{j=1}^{p} \mu_j \underbrace{h_j'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)}_{=h_j(\mathbf{x}) - h_j(\mathbf{x}_0) = 0} \\ &\underset{g_i \text{ convex}}{\geq} -\sum_{i=1}^{m} \lambda_i (g_i(\mathbf{x}) - g_i(\mathbf{x}_0)) = -\sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq 0. \end{aligned}$$

Local First-Order Optimality Conditions

- Convex optimization problems

Even if we have convex problems the KKT conditions are not necessary for minimal points.
Example:



$$(P)\begin{cases}\min\limits_{\mathbf{x}} \ f(\mathbf{x}) := x_1,\\ g_1(\mathbf{x}) := x_1^2 + (x_2-1)^2 - 1 \leq 0,\\ g_2(\mathbf{x}) := x_1^2 + (x_2+1)^2 - 1 \leq 0.\end{cases}$$

Obviously, only the point $\mathbf{x}_0 := (0,0)^\top$ is feasible. Hence, $\mathbf{x}_0$ is the (global) minimal point. Since $\nabla f(\mathbf{x}_0) = (1,0)^\top$, $\nabla g_1(\mathbf{x}_0) = (0,-2)^\top$ and $\nabla g_2(\mathbf{x}_0) = (0,2)^\top$, the gradient condition of the KKT conditions is not met.
Note that the Slater condition is not fulfilled.

Local First-Order Optimality Conditions

- Relationship among the conditions

$$
\begin{tikzcd}
\mathbf{x}_0 \text{ is a local minimizer} \arrow[r, shift right, "\text{CQ}"'] \arrow[d, Rightarrow] & \mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset \arrow[l, shift right, "\text{CP}"'] \arrow[r, Leftrightarrow] \arrow[dl, Rightarrow] & \text{KKT conditions} \\
\mathcal{C}_t(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset \arrow[d, Rightarrow] & & \\
\mathcal{C}_{fd}(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset & &
\end{tikzcd}
$$

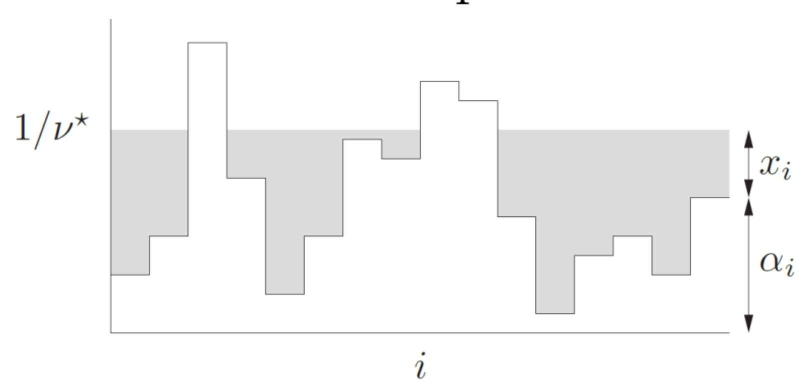
Optimality conditions

- KKT optimality conditions for convex problems

If a convex optimization problem 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 $\mathbf{x}$ is optimal iff there are $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that, together with $\mathbf{x}$, satisfy the KKT conditions.

The KKT conditions play an important role in optimization. In a few special cases it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically. More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.

Examples: Equality constrained convex quadratic minimization, Water-filling

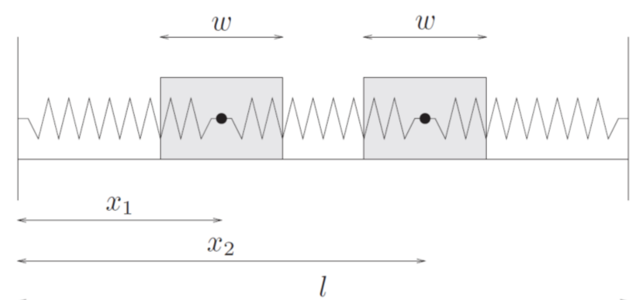


Optimality conditions

- Mechanics interpretation of KKT

The KKT conditions can be given a nice interpretation in mechanics (which indeed, was one of Lagrange's primary motivations). We illustrate the idea with a simple system shown in the figure. The position of the blocks are given by $\mathbf{x} \in \mathbb{R}^2$ where x_1 is the displacement of the (middle of the) left block, and x_2 is the displacement of the right block. The left wall is at position 0, and the right wall is at position l.

The potential energy in the springs is given by



$$f_0(x_1, x_2) = \tfrac{1}{2}k_1 x_1^2 + \tfrac{1}{2}k_2(x_2 - x_1)^2 + \tfrac{1}{2}k_3(l - x_2)^2,$$

where $k_i > 0$ are the stiffness constants of the three springs. The equilibrium position $\mathbf{x}^*$ is the position that minimizes the potential energy subject to

$$w/2 - x_1 \le 0, \quad w + x_1 - x_2 \le 0, \quad w/2 - l + x_2 \le 0. \tag{19}$$

Optimality conditions

- Mechanics interpretation of KKT

These constraints are called *kinematic constraints*, and express the fact that the blocks have width $w > 0$, and cannot penetrate each other or the walls. The equilibrium position is therefore given by the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2) \\ \text{subject to} & w/2 - x_1 \leq 0 \\ & w + x_1 - x_2 \leq 0 \\ & w/2 - l + x_2 \leq 0, \end{array} \tag{20}$$

which is a QP.

Optimality conditions

- Mechanics interpretation of KKT

With λ_1, λ_2, λ_3 as Lagrange multipliers, the KKT conditions for this problem consist of the kinematic constraints (19), the nonnegativity constraints $\lambda_i \geq 0$, the complementary slackness conditions

$$\lambda_1(w/2 - x_1) = 0, \quad \lambda_2(w - x_2 + x_1) = 0, \quad \lambda_3(w/2 - l + x_2) = 0, \tag{21}$$

and the zero gradient condition

$$\left[\begin{array}{c} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{array}\right] + \lambda_1 \left[\begin{array}{c} -1 \\ 0 \end{array}\right] + \lambda_2 \left[\begin{array}{c} 1 \\ -1 \end{array}\right] + \lambda_3 \left[\begin{array}{c} 0 \\ 1 \end{array}\right] = 0. \tag{22}$$

The equation (21) can be interpreted as the force balance equations for the two blocks, provided we interpret the Lagrange multipliers as *contact forces* that act between the walls and blocks.

Optimality conditions

- Mechanics interpretation of KKT

The first equation states that the sum of the forces on the first block is zero: The term $-k_1x_1$ is the force exerted on the left block by the left spring, the term $k_2(x_2 - x_1)$ is the force exerted by the middle spring, λ_1 is the force exerted by the left wall, and $-\lambda_2$ is the force exerted by the right block. The contact forces must point away from the contact surface (as expressed by the constraints $\lambda_1 \geq 0$ and $-\lambda_2 \leq 0$), and are nonzero only when there is contact (as expressed by the first two complementary slackness conditions (21)). In a similar way, the second equation in (22) is the force balance for the second block, and the last condition in (21) states that λ_3 is zero unless the right block touches the wall.

In this example, the potential energy and kinematic constraint functions are convex, and (the refined form of) Slater's constraint qualification holds provided $2w \leq l$, *i.e.*, there is enough room between the walls to fit the two blocks, so we can conclude that the energy formulation of the equilibrium given by (20), gives the same result as the force balance formulation, given by the KKT conditions.

Duality

- The Lagrange dual function - The Lagrangian

We consider an optimization problem in the standard form:

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

We assume its domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\ h_i$ is nonempty, and denote the optimal value of (1) by p^*.

Duality

- The Lagrange dual function - The Lagrangian

We define the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R}$ associated with the problem (1) as

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

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m_+ \times \mathbb{R}^p$. We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint $f_i(x) \leq 0$; similarly we refer to ν_i as the *Lagrange multiplier* associated with the ith equality constraint $h_i(x) = 0$. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called the *dual variables or Lagrange multiplier vectors* associated with the problem (1).

Duality

- The Lagrange dual function - The Lagrange dual function

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 $\mathbf{x}$: for $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, $\boldsymbol{\nu} \in \mathbb{R}^p$,

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

When the Lagrangian is unbounded below in $\mathbf{x}$, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, it is concave, even when the problem (1) is not convex.

Duality

- The Lagrange dual function - Lower bounds on optimal value

For any $\boldsymbol{\lambda} \geq \mathbf{0}$ and any $\boldsymbol{\nu}$ we have an important relationship:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*. \tag{2}$$

Proof. Suppose $\tilde{\mathbf{x}}$ is a feasible point for the problem (1), *i.e.*, $f_i(\tilde{\mathbf{x}}) \leq 0$ and $h_i(\tilde{\mathbf{x}}) = 0$, and $\boldsymbol{\lambda} \geq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{\mathbf{x}}) \leq 0.$$

Therefore

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

Hence

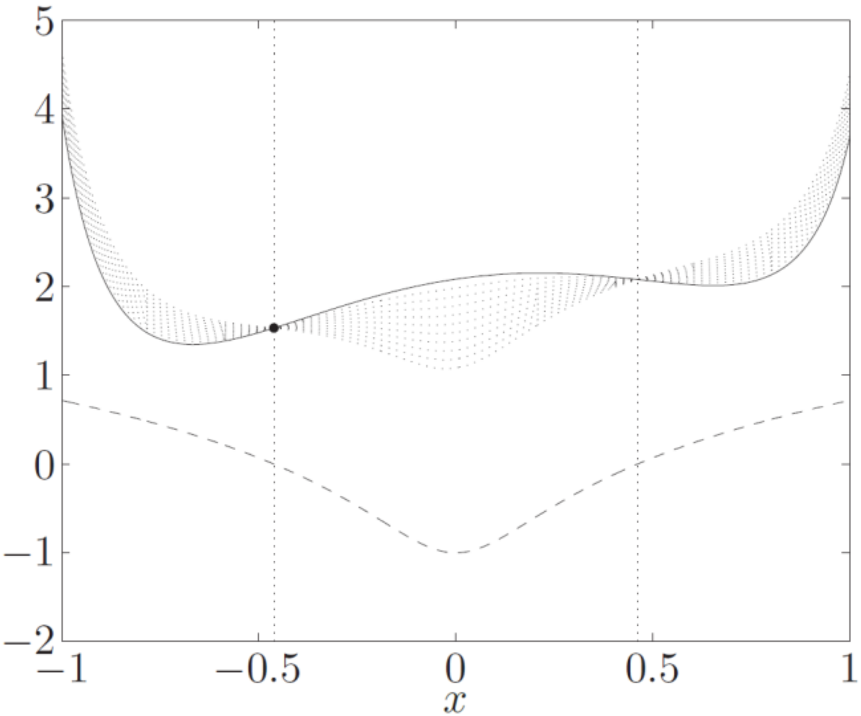
$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\mathbf{x}}).$$

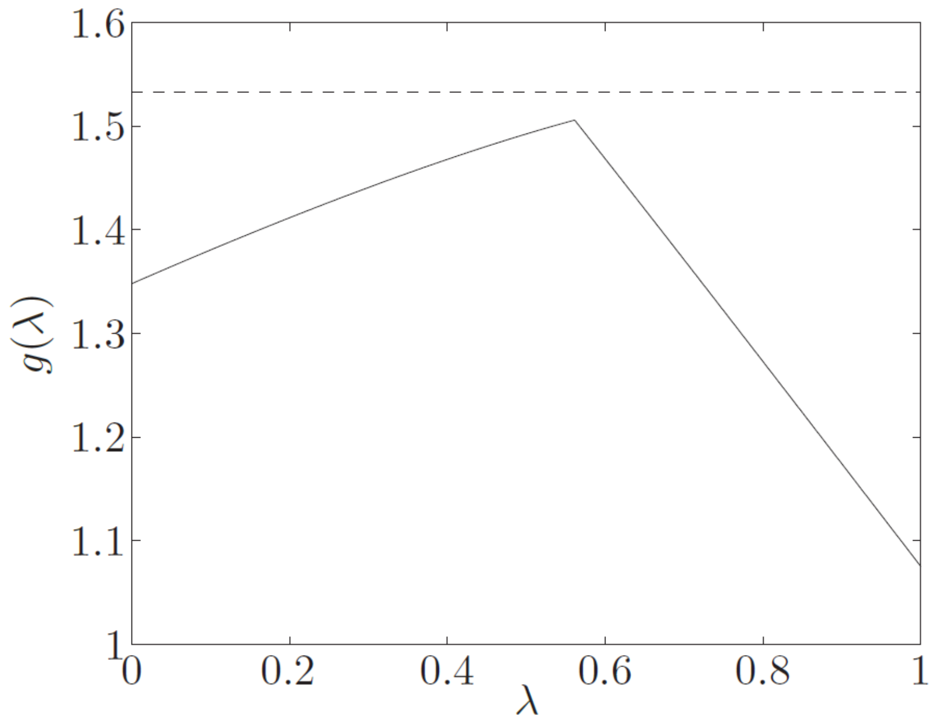
Since $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\mathbf{x}})$ holds for every feasible point $\tilde{\mathbf{x}}$, the inequality (2) follows.

Duality

- The Lagrange dual function - Lower bounds on optimal value

The *domain* of dual function g is $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ such that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$. We refer to a pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \succeq \mathbf{0}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g$ as *dual feasible*.





Duality

- The Lagrange dual function - Linear approximation interpretation

The Lagrangian and lower bound property can be given a simple interpretation, based on a linear approximation of the indicator functions of the sets $\{0\}$ and $-\mathbb{R}_+$.

We first rewrite the original problem (1) as an unconstrained problem,

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)), \tag{3}$$

where $I_- : \mathbb{R} \to \mathbb{R}$ is the indicator function for the nonpositive reals,

$$I_-(u) = \begin{cases} 0, & u \leq 0, \\ \infty, & u > 0, \end{cases}$$

and similarly, I_0 is the indicator function of $\{0\}$.

Duality

- The Lagrange dual function - Linear approximation interpretation

Now suppose in the formulation (3) we replace the function $I_-(u)$ with the linear function $\lambda_i u$, where $\lambda_i \geq 0$, and the function $I_0(u)$ with $\nu_i u$. The objective becomes the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, and the dual function value $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is the optimal value of the problem

$$\min_{\mathbf{x}} \quad L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}). \tag{4}$$

In this formulation, we use a linear or "soft" displeasure function in place of I_- and I_0. For an inequality constraint, our displeasure is zero when $f_i(\mathbf{x}) = 0$, and is positive when $f_i(x) > 0$ (assuming $\lambda_i > 0$); our displeasure grows as the constraint becomes "more violated". Unlike the original formulation, in which any nonpositive value of $f_i(\mathbf{x})$ is acceptable, in the soft formulation we actually derive pleasure from constraints that have margin, *i.e.*, from $f_i(\mathbf{x}) < 0$.

Duality

- The Lagrange dual function - Linear approximation interpretation

Clearly the approximation of the indicator function $I_-(u)$ with a linear function $\lambda_i u$ is rather poor. But the linear function is at least an *underestimator* of the indicator function. Since $\lambda_i u \leq I_-(u)$ and $\nu_i u \leq I_0(u)$ for all u, we see immediately that the dual function yields a lower bound on the optimal value of the original problem.

Examples: Least-squares solution of linear equations, Standard form LP, Two-way partitioning problem,

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

The conjugate function and Lagrange dual function are closely related. To see one simple connection, consider the problem

$$\begin{aligned} &\min\ f(\mathbf{x}), \\ &s.t.\ \mathbf{x} = \mathbf{0}. \end{aligned}$$

This problem has Lagrangian $L(\mathbf{x}, \boldsymbol{\nu}) = f(\mathbf{x}) + \boldsymbol{\nu}^T\mathbf{x}$, and dual function

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x}} \left(f(\mathbf{x}) + \boldsymbol{\nu}^T\mathbf{x} \right) = -\sup_{\mathbf{x}} \left((-\boldsymbol{\nu})^T\mathbf{x} - f(\mathbf{x}) \right) = -f^*(-\boldsymbol{\nu}).$$

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

More generally (and more usefully), consider an optimization problem with linear inequality and equality constraints,

$$\begin{aligned} \min_{\mathbf{x}} \ & f_0(\mathbf{x}), \\ s.t. \ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \\ & \mathbf{C}\mathbf{x} = \mathbf{d}. \end{aligned}$$

Using the conjugate of f_0 we can write its dual function as

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\nu}^T(\mathbf{C}\mathbf{x} - \mathbf{d}) \right) \\ &= -\mathbf{b}^T\boldsymbol{\lambda} - \mathbf{d}^T\boldsymbol{\nu} + \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + (\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\nu})^T\mathbf{x} \right) \\ &= -\mathbf{b}^T\boldsymbol{\lambda} - \mathbf{d}^T\boldsymbol{\nu} - f_0^*(-\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{C}^T\boldsymbol{\nu}). \end{aligned}$$

The domain of g follows from the domain of f_0^*:

$$\mathbf{dom}\ g = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) |\ \boldsymbol{\lambda} \geq \mathbf{0}, -\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{C}^T\boldsymbol{\nu} \in \mathbf{dom}\ f_0^*\}.$$

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

Examples: Equality constrained norm minimization, Entropy maximization, Minimum volume covering ellipsoid.

Duality

- The Lagrange Dual Problem

For each pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq \mathbf{0}$, the Lagrange dual function gives us a lower bound on the optimal value p^* of the optimization problem (1). Thus we have a lower bound that depends on some parameters $\boldsymbol{\lambda}, \boldsymbol{\nu}$. A natural question is: What is the *best* lower bound that can be obtained from the Lagrange dual function?

This leads to the optimization problem

convex no matter the primal problem (1) is convex or not

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \; & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ s.t. \; & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{5}$$

This problem is called the *Lagrange dual problem* associated with the problem (1). In this context the original problem (1) is sometimes called the *primal problem*. The term *dual feasible*, to describe a pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq \mathbf{0}$ and $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$, now makes sense. It means that $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is feasible for the dual problem (5). We refer to $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ as *dual optimal* or *optimal Lagrange multipliers* if they are optimal for the problem (5).

Duality

- The Lagrange Dual Problem - Making dual constraints explicit

It is not uncommon for the domain of the dual function,

$$\mathbf{dom}\ g = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) |\ \boldsymbol{\lambda} \geq \mathbf{0}, g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty\},$$

to have dimension smaller than $m + p$. In many cases we can identify the affine hull of **dom** g, and describe it as a set of linear equality constraints. Roughly speaking, this means we can identify the equality constraints that are 'hidden' or 'implicit' in the objective g of the dual problem (5). In this case we can form an equivalent problem, in which these equality constraints are given explicitly as constraints.

Examples: Lagrange dual of standard form LP, Lagrange dual of inequality form LP

Note the interesting symmetry between the standard and inequality form LPs and their duals: The dual of a standard form LP is an LP with only inequality constraints, and vice versa.

Duality

- The Lagrange Dual Problem - Weak duality

The optimal value of the Lagrange dual problem, which we denote d^*, is, by definition, the best lower bound on p^* that can be obtained from the Lagrange dual function. In particular, we have the simple but important inequality

$$d^* \leq p^*, \tag{6}$$

which holds even if the original problem is not convex. This property is called *weak duality*.

The weak duality inequality (6) holds when d^* and p^* are infinite. For example, if the primal problem is unbounded below, so that $p^* = -\infty$, we must have $d^* = -\infty$, *i.e.*, the Lagrange dual problem is infeasible. Conversely, if the dual problem is unbounded above, so that $d^* = \infty$, we must have $p^* = \infty$, *i.e.*, the primal problem is infeasible.

Duality

- The Lagrange Dual Problem - Weak duality

We refer to the difference $p^* - d^*$ as the *optimal duality gap* of the original problem, since it gives the gap between the optimal value of the primal problem and the best (*i.e.*, greatest) lower bound on it that can be obtained from the Lagrange dual function. The optimal duality gap is always nonnegative.

The bound (6) can sometimes be used to find a lower bound on the optimal value of a problem that is difficult to solve, since the dual problem is always convex, and in many cases can be solved efficiently, to find d^*.

Example: Dual of the two-way partitioning problem

Duality

- The Lagrange Dual Problem – Strong duality & Slatter's CQ

If the equality

$$d^* = p^* \tag{7}$$

holds, *i.e.*, the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight. Strong duality does not, in general, hold. But if the primal problem (1) is convex, *i.e.*, of the form

$$\begin{aligned} \min_{\mathbf{x}} \;& f_0(\mathbf{x}) \\ s.t. \;& f_i(\mathbf{x}) \le 0, \quad i = 1, \cdots, m, \\ & \mathbf{Ax} = \mathbf{b}, \end{aligned} \tag{8}$$

with $f_0, \cdots, f_m$ convex, we usually (but not always) have strong duality. There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called *constraint qualifications*.

Duality

- The Lagrange Dual Problem – Strong duality & Slatter's CQ

One simple constraint qualification is *Slater's condition*: There exists an $\mathbf{x} \in \mathbf{relint}\ \mathcal{D}$ such that

$$f_i(\mathbf{x}) < 0 \quad i = 1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{9}$$

Such a point is sometimes called *strictly feasible*, since the inequality constraints hold with strict inequalities. Slater's theorem states that strong duality holds, if Slater's condition holds (and the problem is convex).

Slater's condition can be refined when some of the inequality constraint functions f_i are affine. If the first k constraint functions $f_1, \cdots, f_k$ are affine, then strong duality holds provided the following weaker condition holds: There exists an $\mathbf{x} \in \mathbf{relint}\ \mathcal{D}$ with

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots k, \quad f_i(\mathbf{x}) < 0, \quad i = k+1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{10}$$

In other words, the affine inequalities do not need to hold with strict inequality. Note that the refined Slater condition (10) reduces to feasibility when the constraints are all linear equalities and inequalities, and $\mathbf{dom}\ f_0$ is open.

Duality

- The Lagrange Dual Problem – Strong duality & Slatter's CQ

Slater's condition (and the refinement (10)) not only implies strong duality for convex problems. It also implies that the dual optimal value is attained when $d^* > -\infty$, *i.e.*, there exists a dual feasible $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ with $g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = d^* = p^*$. We will prove that strong duality obtains, when the primal problem is convex and Slater's condition holds, in Section 7.1.3.2.

Examples: Least-squares solution of linear equations, Lagrange dual of LP, Lagrange dual of QCQP, Entropy maximization, Minimum volume covering ellipsoid, A nonconvex quadratic problem with strong duality

Duality

- Geometric interpretation - Weak and strong duality

We can give a simple geometric interpretation of the dual function in terms of the set

$$\mathcal{G} = \{(f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}), h_1(\mathbf{x}), \cdots, h_p(\mathbf{x}), f_0(\mathbf{x})) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} |\ \mathbf{x} \in \mathcal{D}\}, \quad (11)$$

which is the set of values taken on by the constraint and objective functions. The optimal value p^* of (1) is easily expressed in terms of $\mathcal{G}$ as

$$p^* = \inf\{t|\ (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G},\ \mathbf{u} \leq \mathbf{0},\ \mathbf{v} = \mathbf{0}\}.$$

To evaluate the dual function at $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, we minimize the affine function

$$(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T (\mathbf{u}, \mathbf{v}, t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} \nu_i v_i + t$$

over $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}$, *i.e.*, we have

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf\{(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T (\mathbf{u}, \mathbf{v}, t)|\ (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}\}.$$

Duality

- Geometric interpretation - Weak and strong duality

In particular, we see that if the infimum is finite, then the inequality

$$(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

defines a supporting hyperplane to $\mathcal{G}$. This is sometimes referred to as a *nonvertical* supporting hyperplane, because the last component of the normal vector is nonzero.

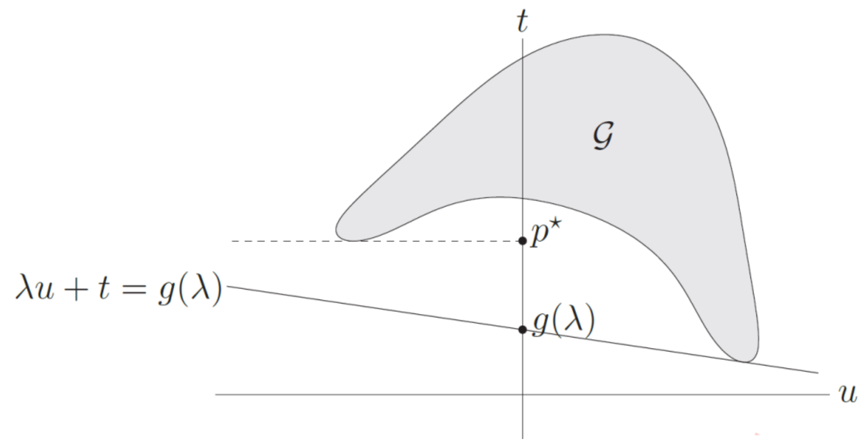
Now suppose $\boldsymbol{\lambda} \geq \mathbf{0}$. Then, obviously, $t \geq (\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t)$ if $\mathbf{u} \leq \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$. Therefore

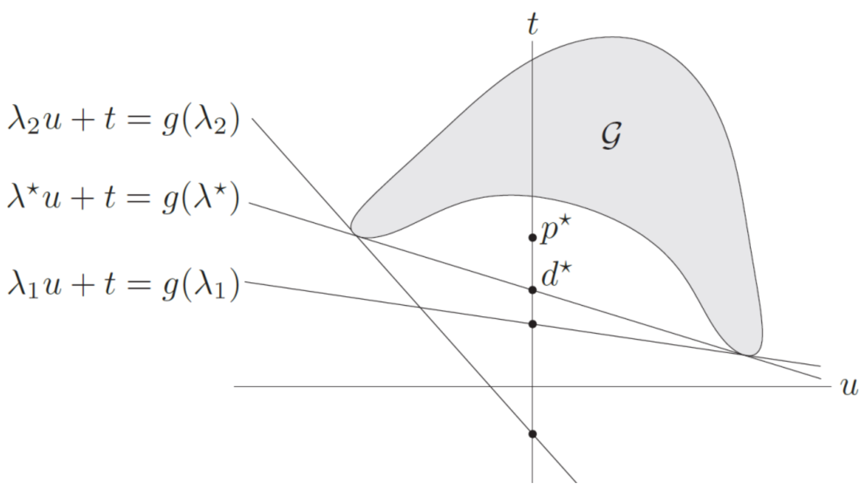
$$\begin{aligned} p^* &= \inf\{t \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G},\ \mathbf{u} \leq \mathbf{0},\ \mathbf{v} = \mathbf{0}\} \\ &\geq \inf\{(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G},\ \mathbf{u} \leq \mathbf{0},\ \mathbf{v} = \mathbf{0}\} \\ &\geq \inf\{(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}\} \\ &= g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \end{aligned}$$

i.e., we have weak duality.

Duality

- Geometric interpretation - Weak and strong duality





Duality

- Geometric interpretation - Weak and strong duality

Epigraph variation

We describe a variation on the geometric interpretation of duality in terms of $\mathcal{G}$, which explains why strong duality obtains for (most) convex problems. We define the set $A \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$ as

$$\mathcal{A} = \mathcal{G} + \left(\mathbb{R}_+^m \times \{0\} \times \mathbb{R}_+\right), \tag{12}$$

or, more explicitly,

$$\begin{aligned} \mathcal{A} \quad = \quad & \{(\mathbf{u}, \mathbf{v}, t) |\ \exists \mathbf{x} \in \mathcal{D},\ f_i(\mathbf{x}) \le u_i,\ i = 1, \cdots, m, \\ & h_i(\mathbf{x}) = v_i,\ i = 1, \cdots, p,\ f_0(\mathbf{x}) \le t\}, \end{aligned}$$

We can think of $\mathcal{A}$ as a sort of epigraph form of $\mathcal{G}$, since $\mathcal{A}$ includes all the points in $\mathcal{G}$, as well as points that are 'worse', *i.e.*, those with larger objective or inequality constraint function values.

Duality

- Geometric interpretation - Weak and strong duality

We can express the optimal value in terms of $\mathcal{A}$ as

$$p^* = \inf\{t \mid (\mathbf{0}, \mathbf{0}, t) \in \mathcal{A}\}.$$

To evaluate the dual function at a point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq \mathbf{0}$, we can minimize the affine function $(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t)$ over $\mathcal{A}$: If $\boldsymbol{\lambda} \geq \mathbf{0}$, then

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf\{(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{A}\}.$$

If the infimum is finite, then

$$(\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

defines a nonvertical supporting hyperplane to $\mathcal{A}$.

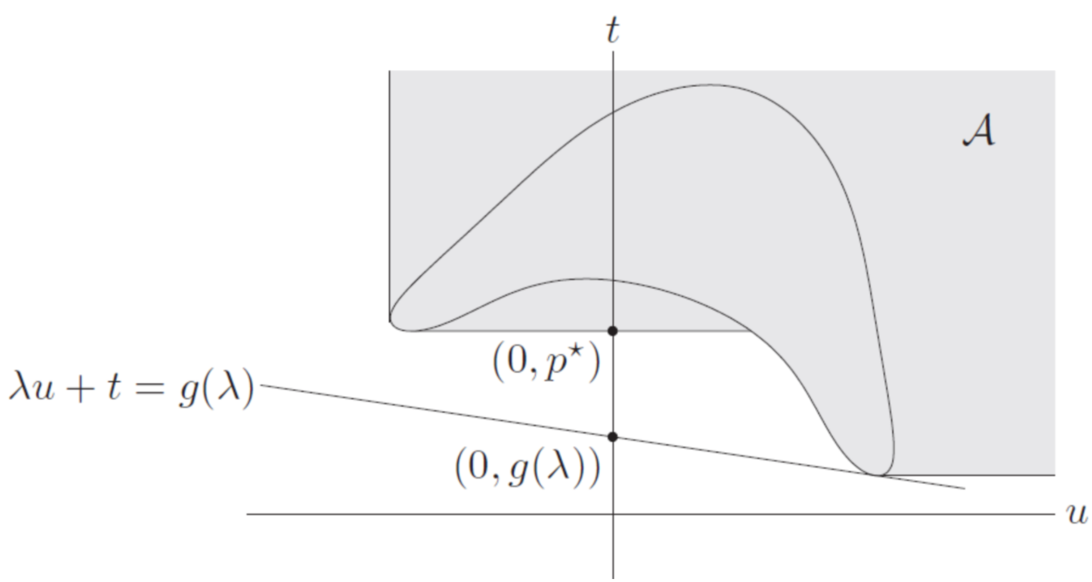
Duality

- Geometric interpretation - Weak and strong duality

In particular, since $(\mathbf{0}, \mathbf{0}, p^*) \in \partial \mathcal{A}$, we have

$$p^* = (\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T (\mathbf{0}, \mathbf{0}, p^*) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{13}$$

the weak duality lower bound. Strong duality holds iff we have equality in (13) for some dual feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, i.e., there exists a nonvertical supporting hyperplane to $\mathcal{A}$ at its boundary point $(\mathbf{0}, \mathbf{0}, p^*)$.

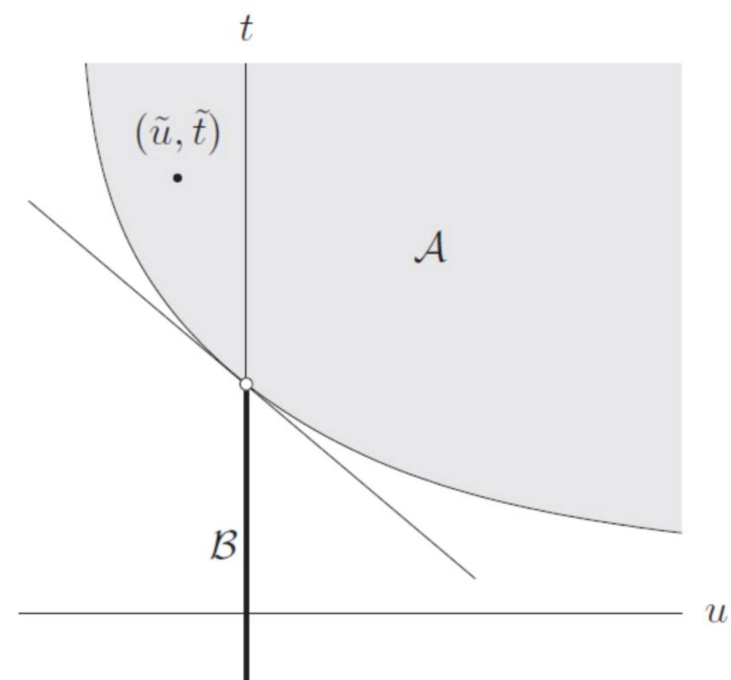


Duality

- Geometric interpretation - Proof of strong duality under CQ

In this section we prove that Slater's constraint qualification guarantees strong duality (and that the dual optimum is attained) for a convex problem. We consider the primal problem (8), with $f_0, \cdots, f_m$ convex, and assume Slater's condition holds: There exists $\tilde{\mathbf{x}} \in \textbf{relint}\ \mathcal{D}$ with $f_i(\tilde{\mathbf{x}}) < 0$, $i = 1, \cdots, m$, and $A\tilde{\mathbf{x}} = \mathbf{b}$. In order to simplify the proof, we make two additional assumptions: first that $\mathcal{D}$ has nonempty interior (hence, $\textbf{relint}\ \mathcal{D} = \textbf{int}\ \mathcal{D}$) and second, that $\operatorname{rank} \mathbf{A} = p$. We assume that p^* is finite.

See Lecture Note for the proof.



Duality

- Saddle-point interpretation - Max-min characterization

It is possible to express the primal and the dual optimization problems in a form that is more symmetric. To simplify the discussion we assume there are no equality constraints; the results are easily extended to cover them.

First note that

$$\begin{aligned}\sup_{\boldsymbol{\lambda}\geq\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda}) &= \sup_{\boldsymbol{\lambda}\geq\mathbf{0}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) \right) \\ &= \begin{cases} f_0(\mathbf{x}) & f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m \\ \infty & \text{otherwise,} \end{cases}\end{aligned}$$

Indeed, suppose $\mathbf{x}$ is not feasible, and $f_i(\mathbf{x}) > 0$ for some i. Then $\sup_{\boldsymbol{\lambda}\geq\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda}) = \infty$, as can be seen by choosing $\lambda_j = 0$, $j \neq i$, and $\lambda_i \to \infty$. On the other hand, if $f_i(\mathbf{x}) \leq 0$, $i = 1, \cdots, m$, then the optimal choice of $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda} = \mathbf{0}$ and $\sup_{\boldsymbol{\lambda}\geq\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda}) = f_0(\mathbf{x})$. This means that we can have

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

Duality

- Saddle-point interpretation - Max-min characterization

By the definition of the dual function, we also have

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

Thus, weak duality can be expressed as the inequality

$$\sup_{\boldsymbol{\lambda} \geq \mathbf{0}} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \leq \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}) \tag{14}$$

and strong duality as the equality

$$\sup_{\boldsymbol{\lambda} \geq \mathbf{0}} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\lambda}).$$

Strong duality means that the order of the minimization over $\mathbf{x}$ and the maximization over $\boldsymbol{\lambda} \geq \mathbf{0}$ can be switched without affecting the result.

Duality

- Saddle-point interpretation - Max-min characterization

In fact, the inequality (14) does not depend on any properties of L: We have

$$\sup_{\mathbf{z}\in\mathbf{Z}} \inf_{\mathbf{w}\in\mathcal{W}} f(\mathbf{w},\mathbf{z}) \leq \inf_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{z}\in\mathcal{Z}} f(\mathbf{w},\mathbf{z}) \tag{15}$$

for any $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ (and any $\mathcal{W} \subseteq \mathbb{R}^n$ and $\mathcal{Z} \subseteq \mathbb{R}^m$). This general inequality is called the *max-min inequality*. When equality holds, *i.e.*,

$$\sup_{\mathbf{z}\in\mathcal{Z}} \inf_{\mathbf{w}\in\mathcal{W}} f(\mathbf{w},\mathbf{z}) \leq \inf_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{z}\in\mathcal{Z}} f(\mathbf{w},\mathbf{z}) \tag{16}$$

we say that f (and $\mathcal{W}$ and $\mathcal{Z}$) satisfy the *strong max-min property* or the *saddle point* property. Of course the strong max-min property holds only in special cases, for example, when $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the Lagrangian of a problem for which strong duality obtains, $\mathcal{W} = \mathbb{R}^n$, and $\mathcal{Z} = \mathbb{R}^m_+$.

Duality

- Saddle-point interpretation

We refer to a pair $\tilde{\mathbf{w}} \in \mathcal{W}$, $\tilde{\mathbf{z}} \in \mathcal{Z}$ as a *saddle-point* for f (and $\mathcal{W}$ and $\mathcal{Z}$) if

$$f(\tilde{\mathbf{w}}, \mathbf{z}) \le f(\tilde{\mathbf{w}}, \tilde{\mathbf{z}}) \le f(\mathbf{w}, \tilde{\mathbf{z}})$$

for all $\mathbf{w} \in \mathcal{W}$ and $\mathbf{z} \in \mathcal{Z}$. In other words, $\tilde{\mathbf{w}}$ minimizes $f(\mathbf{w}, \tilde{\mathbf{z}})$ (over $\mathbf{w} \in \mathcal{W}$) and $\tilde{\mathbf{z}}$ maximizes $f(\tilde{\mathbf{w}}, \mathbf{z})$ (over $\mathbf{z} \in \mathcal{Z}$):

$$f(\tilde{\mathbf{w}}, \tilde{\mathbf{z}}) = \inf_{\mathbf{w} \in \mathcal{W}} f(\mathbf{w}, \tilde{\mathbf{z}}), \quad f(\tilde{\mathbf{w}}, \tilde{\mathbf{z}}) = \sup_{\mathbf{z} \in \mathcal{Z}} f(\tilde{\mathbf{w}}, \mathbf{z}).$$

This implies that the strong max-min property (16) holds, and that the common value is $f(\tilde{\mathbf{w}}, \tilde{\mathbf{z}})$.

Returning to our discussion of Lagrange duality, we see that if $\mathbf{x}^*$ and $\boldsymbol{\lambda}^*$ are primal and dual optimal points for a problem in which strong duality obtains, they form a saddle-point for the Lagrangian. The converse is also true: If $(\mathbf{x}, \boldsymbol{\lambda})$ is a saddle-point of the Lagrangian, then $\mathbf{x}$ is primal optimal, $\boldsymbol{\lambda}$ is dual optimal, and the optimal duality gap is zero.