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

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

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^*:

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

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,

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

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

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

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

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

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

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

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.

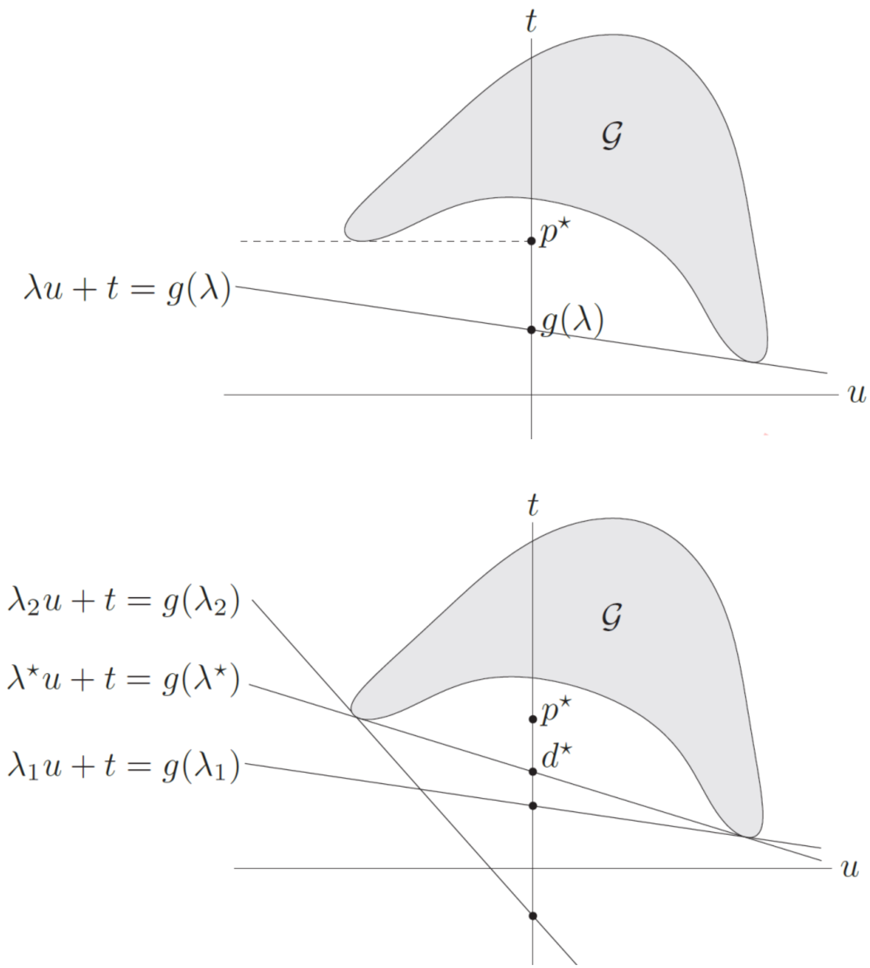
Now suppose $\boldsymbol{\lambda} \geq \mathbf{0}$. Then, obviously, $t \geq (\boldsymbol{\lambda}, \boldsymbol{\nu}, 1)^T(\mathbf{u}, \mathbf{v}, t)$ if $\mathbf{u} \leq \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$. Therefore

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

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,

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

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

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

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

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

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 $\mathcal{A}$.

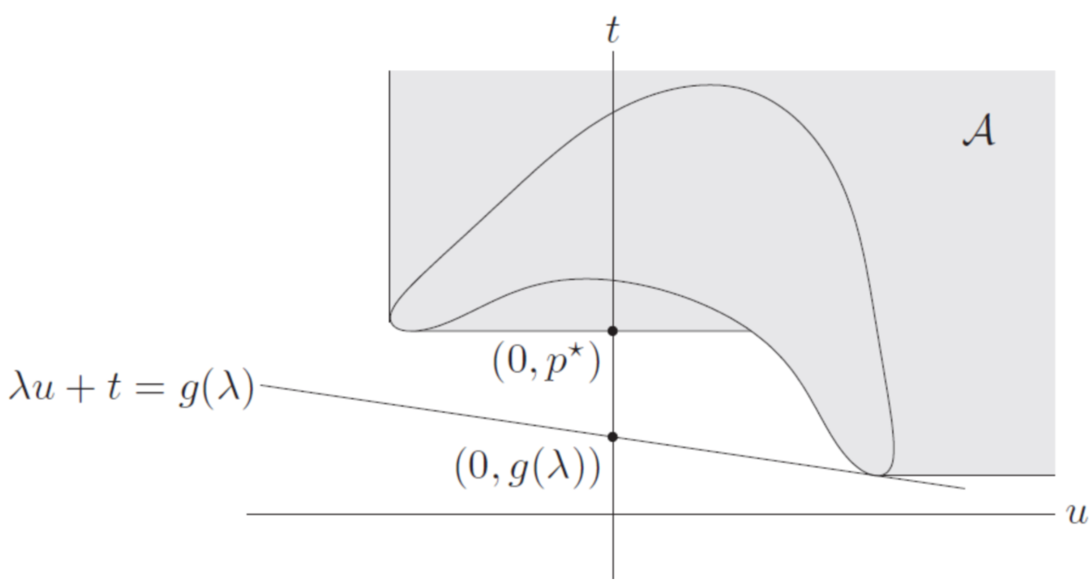
Duality

- 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 $(\mathbf{0}, \mathbf{0}, p^*)$.

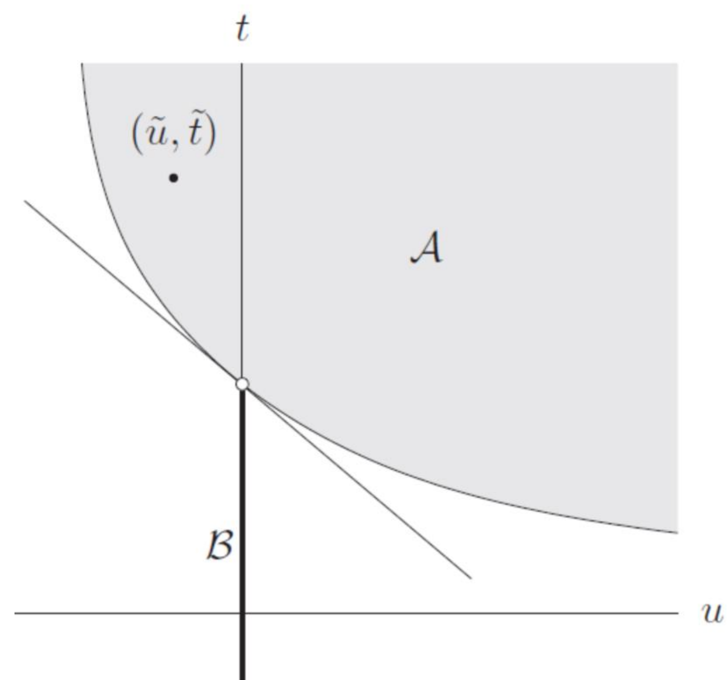


Duality

- 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

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

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.

Duality

- Optimality conditions - Certificate of suboptimality

We do not assume the problem (1) is convex, unless explicitly stated.

If we can find a dual feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, we establish a lower bound on the optimal value of the primal problem: $p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Thus a dual feasible point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ provides a *proof or certificate* that $p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Strong duality means there exist arbitrarily good certificates.

Dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of p^*. Indeed, if bx is primal feasible and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is dual feasible, then

$$f_0(\mathbf{x}) - p^* \leq f_0(\mathbf{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\nu}).$$

In particular, this establishes that $\mathbf{x}$ is ϵ-suboptimal, with $\epsilon = f_0(\mathbf{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. (It also establishes that $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is ϵ-suboptimal for the dual problem.)

Duality

- Optimality conditions - Certificate of suboptimality

If the duality gap of the primal dual feasible pair $\mathbf{x}, (\boldsymbol{\lambda}, \boldsymbol{\nu})$ is zero, i.e., $f_0(\mathbf{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$, then $\mathbf{x}$ is primal optimal and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is dual optimal. We can think of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ as a certificate that proves $\mathbf{x}$ is optimal (and, similarly, we can think of $\mathbf{x}$ as a certificate that proves $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is dual optimal).

These observations can be used in optimization algorithms to provide non-heuristic stopping criteria. Suppose an algorithm produces a sequence of primal feasible $\mathbf{x}^{(k)}$ and dual feasible $(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})$, for $k = 1, 2, \cdots$, and $\epsilon_{\text{abs}} > 0$ is a given required absolute accuracy. Then the stopping criterion ($i.e.$, the condition for terminating the algorithm)

$$f_0(\mathbf{x}^{(k)}) - g(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)}) \leq \epsilon_{\text{abs}}$$

guarantees that when the algorithm terminates, $\mathbf{x}^{(k)}$ is ϵ_{abs}-suboptimal. Indeed, $(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})$ is a certificate that proves it. (Of course strong duality must hold if this method is to work for arbitrarily small tolerances ϵ_{abs}.)

Duality

- Optimality conditions - Certificate of suboptimality

A similar condition can be used to guarantee a given relative accuracy. If

$$\frac{f_0(bx^{(k)}) - g(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})}{|g(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})|} \leq \epsilon_{\mathrm{rel}}$$

holds, or

$$\frac{f_0(\mathbf{x}^{(k)}) - g(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})}{|f_0(\mathbf{x}^{(k)})|} \leq \epsilon_{\mathrm{rel}}$$

holds, then $p^* \neq 0$ and the relative error

$$\frac{f_0(\mathbf{x}^{(k)}) - p^*}{|p^*|}$$

is guaranteed to be less than or equal to ϵ_{rel}.

Duality

- Optimality conditions - Complementary slackness

Suppose that the primal and dual optimal values are attained and equal (so, in particular, strong duality holds). Let $\mathbf{x}^*$ be a primal optimal and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be a dual optimal point. This means that

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\ &\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \leq f_0(\mathbf{x}^*). \end{aligned}$$

The first line states that the optimal duality gap is zero, and the second line is the definition of the dual function. The third line follows since the infimum of the Lagrangian over $\mathbf{x}$ is less than or equal to its value at $\mathbf{x} = \mathbf{x}^*$. The last inequality follows from $\lambda_i^* \geq 0$, $f_i(\mathbf{x}^*) \leq 0$, $i = 1, \cdots, m$, and $h_i(\mathbf{x}^*) = 0$, $i = 1, \cdots, p$. We conclude that the two inequalities in this chain hold with equality.

Duality

- Optimality conditions - Complementary slackness

We can draw several interesting conclusions from this. For example, since the inequality in the third line is an equality, we conclude that $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ over $\mathbf{x}$. (The Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu})^*$ can have other minimizers; $\mathbf{x}^*$ is simply a minimizer.)

Another important conclusion is that

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0.$$

Since each term in this sum is nonpositive, we conclude that

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m. \tag{17}$$

Duality

- Optimality conditions - Complementary slackness

This condition is known as *complementary slackness*; it holds for any primal optimal $\mathbf{x}^*$ and any dual optimal $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ (when strong duality holds). We can express the complementary slackness condition as

$$\lambda_i^* > 0 \Longrightarrow f_i(\mathbf{x}^*) = 0,$$

or, equivalently,

$$f_i(\mathbf{x}^*) < 0 \Longrightarrow \lambda_i^* = 0.$$

Roughly speaking, this means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

Duality

- Optimality conditions - KKT optimality conditions

We now assume that the functions $f_0, \cdots, f_m$, $h_1, \cdots, h_p$ are differentiable (and therefore have open domains), but we make no assumptions yet about convexity.

KKT conditions for nonconvex problems

Let $\mathbf{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be any primal and dual optimal points with zero duality gap. Since $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ over $\mathbf{x}$, it follows that its gradient must vanish at $\mathbf{x}^*$, *i.e.*,

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{n} \nu_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0}.$$

Duality

- Optimality conditions - KKT optimality conditions

Thus we have

$$\begin{aligned} f_i(\mathbf{x}^*) &\le 0, \quad i = 1, \cdots, m \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, p \\ \lambda_i^* &\ge 0, \quad i = 1, \cdots, m \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, m \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(\mathbf{x}^*) &= \mathbf{0}, \end{aligned} \tag{18}$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

To summarize, for *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions (18).

Duality

- Optimality conditions - KKT optimality conditions

KKT conditions for convex problems

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. In other words, if f_i are convex and h_i are affine, and $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, $\tilde{\boldsymbol{\nu}}$ are any points that satisfy the KKT conditions

$$
\begin{aligned}
f_i(\tilde{\mathbf{x}}) &\leq 0, \quad i = 1, \cdots, m \\
h_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \cdots, p \\
\tilde{\lambda}_i &\geq 0, \quad i = 1, \cdots, m \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \cdots, m \\
\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) &= \mathbf{0},
\end{aligned}
$$

then $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are primal and dual optimal, with zero duality gap.

Duality

- Optimality conditions - KKT optimality conditions

Proof. Note that the first two conditions state that $\tilde{\mathbf{x}}$ is primal feasible. Since $\tilde{\lambda}_i \geq 0$, $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is convex in $\mathbf{x}$; the last KKT condition states that its gradient with respect to $\mathbf{x}$ vanishes at $\mathbf{x} = \tilde{\mathbf{x}}$, so it follows that $\tilde{\mathbf{x}}$ minimizes $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ over $\mathbf{x}$. From this we conclude that

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\ &= f_0(\tilde{\mathbf{x}}), \end{aligned}$$

where in the last line we use $h_i(\tilde{\mathbf{x}}) = 0$ and $\tilde{\boldsymbol{\lambda}}_i f_i(\tilde{\mathbf{x}}) = 0$. This shows that $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, and therefore are primal and dual optimal. In summary, for any *convex* optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

Duality

- **Optimality conditions** - Solving the primal problem via the dual

If strong duality holds and a dual optimal solution $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ exists, then any primal optimal point is also a minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$. This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution.

More precisely, suppose we have strong duality and an optimal $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is known. Suppose that the minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, *i.e.*, the solution of

$$\text{minimize} \quad f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}), \tag{19}$$

is unique. (For a convex problem this occurs, for example, if $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is a strictly convex function of $\mathbf{x}$.) Then if the solution of (19) is primal feasible, it must be primal optimal; if it is not primal feasible, then no primal optimal point can exist, *i.e.*, we can conclude that the primal optimum is not attained. This observation is interesting when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited.

Examples.

Duality

- Solving the primal problem via the dual – Three types of reformulation

Simple equivalent reformulations of a problem can lead to very different dual problems. We consider the following types of reformulations:

- Introducing new variables and associated equality constraints.
- Replacing the objective with an increasing function of the original objective.
- Making explicit constraints implicit, i.e., incorporating them into the domain of the objective.

Examples.