CHAPTER VII: Duality

In the chapter on data analysis, we will see many cases where a penalization term is added to the objective function, that is, we minimize $f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ instead of minimizing $f(\boldsymbol{x})$. We will see in this chapter that this approach is closely related to the problem of minimizing $f(\boldsymbol{x})$ under the constraint "$g(\boldsymbol{x}) \leq \alpha$", for some constant α.

1 Lagrangian dual

In this section, we consider a nonlinear optimization problem of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & f_i(\boldsymbol{x}) \leq 0, \quad (\forall i \in [m]) \\ & h_j(\boldsymbol{x}) = 0, \quad (\forall j \in [p]). \end{aligned} \tag{NLP}$$

We denote its optimal value by $p^* = \inf\{f_0(\boldsymbol{x}) |\ \boldsymbol{x} \in \mathbb{R}^n,\ f_i(\boldsymbol{x}) \leq 0,\ \forall i \in [m],\quad h_j(\boldsymbol{x}) = 0, \forall j \in [p]\}$.

Note: We do not assume (yet) that (NLP) is convex.

Definition 1 (Lagrangian)**.** The Lagrangian of Problem (NLP) is the function $\mathcal{L} : \mathbb{R}^{n+m+p} \to \mathbb{R} \cup \{\infty\}$, defined by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f_0(\boldsymbol{x}) + \sum_{i\in[m]} \lambda_i f_i(\boldsymbol{x}) + \sum_{j\in[p]} \mu_j h_j(\boldsymbol{x}).$$

We exclude the trivial case where the domains of the f_i's and the g_j's do not intersect, so for all $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, there exists an $\boldsymbol{x}$ such that $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) < \infty$.

Definition 2 (Langrange dual function)**.** The Lagrange dual function of Problem (NLP) is the function $g : \mathbb{R}^{m+p} \to \mathbb{R} \cup \{-\infty\}$, defined by

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \inf_{\boldsymbol{x}\in\mathbb{R}^n} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

Its domain is naturally defined as $\textbf{dom}\, g = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^m \times \mathbb{R}^p :\ g(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$. Note that g is concave (pointwise minimum of affine functions).

The Lagrange dual function gives us a natural bound on the optimal value of p^*. Again, we insist that this bound is always valid (even if (NLP) is nonconvex):

Theorem 1 (Weak duality)**.** *Let $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$, with $\boldsymbol{\lambda} \geq \mathbf{0}$. Then,*

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*.$$

Proof. We first assume that (NLP) is feasible. Let $\tilde{\boldsymbol{x}}$ be a feasible vector. We have

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\tilde{\boldsymbol{x}}) + \sum_{i\in[m]} \lambda_i f_i(\tilde{\boldsymbol{x}}) + \sum_{j\in[p]} \mu_j h_j(\tilde{\boldsymbol{x}}).$$

Since $\tilde{\boldsymbol{x}}$ is feasible, we have $f_i(\tilde{\boldsymbol{x}}) \leq 0$ and $h_j(\tilde{\boldsymbol{x}}) = 0$ for all $i, j \in [m] \times [p]$. Hence,

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f_0(\tilde{\boldsymbol{x}}) + \sum_{i\in[m]} \underbrace{\lambda_i f_i(\tilde{\boldsymbol{x}})}_{\leq 0} + \sum_{j\in[p]} \underbrace{\mu_j h_j(\tilde{\boldsymbol{x}})}_{=0} \leq f_0(\tilde{\boldsymbol{x}}).$$

Since this inequality must hold for all feasible $\tilde{\boldsymbol{x}}$, we obtain $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*$.

If (NLP) is infeasible, then $p^* = \infty$, while $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) < \infty$, so the statement still holds. □

Theorem 2 (Weak duality (2nd version))**.** *An alternative, equivalent formulation of the weak duality theorem is the following inequality:*

$$\sup_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m_+ \\ \boldsymbol{\mu}\in\mathbb{R}^p}} \inf_{\boldsymbol{x}\in\mathbb{R}^n} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} \sup_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m_+ \\ \boldsymbol{\mu}\in\mathbb{R}^p}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proof. It should be clear that the weak duality theorem can be rewritten as

$$\sup_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m_+ \\ \boldsymbol{\mu}\in\mathbb{R}^p}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*.$$

Replacing the Lagrange dual function $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ by its definition, we obtain the expression at the left-hand side of the inequality sign. Then, we claim that the expression at the right-hand side is p^*. This follows from the fact that the maximization problem with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ is *separable* (it is a sum of independent problems), so it can solved easily:

$$\sup_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m_+ \\ \boldsymbol{\mu}\in\mathbb{R}^p}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^m \sup_{\lambda_i \geq 0} \lambda_i f_i(\boldsymbol{x}) + \sum_{j=1}^p \sup_{\mu_j\in\mathbb{R}} \mu_j h_j(\boldsymbol{x}) = \begin{cases} f_0(\boldsymbol{x}) & \text{if } \boldsymbol{x} \text{ is feasible} \\ +\infty & \text{otherwise.} \end{cases}$$

Then, minimizing the above expression is the same as minimizing $f_0(\boldsymbol{x})$ over the feasible set, so

$$\inf_{\boldsymbol{x}\in\mathbb{R}^n} \sup_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m_+ \\ \boldsymbol{\mu}\in\mathbb{R}^p}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = p^*.$$

□

The weak duality theorem gives us a simple way to compute lower bounds for the optimization problem (NLP). So, we may ask ourselves what is the *best possible bound* which we can obtain in this way. The problem of finding the best possible lower weak-duality bound is, in fact, the optimization problem that appears on the left-hand side of the previous theorem:

Definition 3 (Lagrangian Dual)**.** The *Lagrangian dual* of (NLP) is the optimization problem

$$\begin{aligned} \underset{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^p}{\textbf{maximize}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s.t. \quad & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

We denote its optimal value by d^*, and the weak duality theorem tells us that $d^* \leq p^*$.

If $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \textbf{dom}\, g \cap (\mathbb{R}^m_+ \times \mathbb{R}^p)$, we say that $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a pair of *feasible Lagrange multipliers.*

In general, the duality inequality $d^* \leq p^*$ can be strict. We will see that sometimes, we can guarantee that the primal and the dual optimization problems have the same optimal value:

> **Definition 4** (Strong duality)**.** If $p^* = d^*$, then we say that *strong duality* holds.

In general, it can be quite complicated to compute the Lagrangian dual function. We show in the next example that there is a connection between the Lagrangian function and the Fenchel dual, for linearly constrained problems.

Example:
Lagrangian dual of a problem with linear constraints.
We will see that the dual of an optimization problem with linear constraints involves the Fenchel conjugate of the objective function. Consider the following optmization problem:

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\ & F\boldsymbol{x} = \boldsymbol{g}. \end{aligned}$$

The Lagrangian dual function reads

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x}\in\mathbb{R}^n} \; f(\boldsymbol{x}) + \boldsymbol{\lambda}^T(A\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\mu}^T(F\boldsymbol{x} - \boldsymbol{g}) \\ &= -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{g}^T\boldsymbol{\mu} + \inf_{\boldsymbol{x}\in\mathbb{R}^n} \; f(\boldsymbol{x}) + \boldsymbol{x}^T(A^T\boldsymbol{\lambda} + F^T\boldsymbol{\mu}) \end{aligned}$$ #1

Recall the definition of the Fenchel conjugate of f: $f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x}} \; \boldsymbol{x}^T\boldsymbol{y} - f(\boldsymbol{x}) = -\inf_{\boldsymbol{x}} \; f(\boldsymbol{x}) - \boldsymbol{x}^T\boldsymbol{y}$. Then, we have:

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{g}^T\boldsymbol{\mu} - f^*(-A^T\boldsymbol{\lambda} - F^T\boldsymbol{\mu}).$$

Since f^* is convex, the dual is the convex optimization problem

$$\begin{aligned} \underset{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^p}{\textbf{maximize}} \quad & -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{g}^T\boldsymbol{\mu} - f^*(-A^T\boldsymbol{\lambda} - F^T\boldsymbol{\mu}) \\ s.t. \quad & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

There are, in fact, some hidden constraints in the formulation of the dual we derived above, because we did not write explictly the domain of the Fenchel conjugate. For certain functions f, the constraints $-A^T\boldsymbol{\lambda} - F^T\boldsymbol{\mu} \in \operatorname{\mathbf{dom}} f^*$ can be difficult to handle.

2 Dual of conic programming problems

Things become much easier with conic programming problems: We will see that the dual of a conic programming problem is another conic programming problem, which we can write explictly. For this we first need to define the Lagrangian dual of a problem with generalized conic ($\preceq_K$) inequalities. Consider the conic programming problem

$$\begin{aligned} p^* = \inf_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ s.t. \quad & A_0\,\boldsymbol{x} = \boldsymbol{b}_0 \\ & A_i\,\boldsymbol{x} \succeq_{K_i} \boldsymbol{b}_i, \qquad (i = 1, \ldots, m), \end{aligned} \tag{CP}$$

where the cones $K_1, \dots, K_m$ are proper. The Lagrangian of (CP) is defined in a simmilar manner as for (NLP). For each constraint we introduce a Lagrange multiplier $\boldsymbol{y}_i \in \mathbb{R}^{n_i}$, where n_i is the dimension of $\boldsymbol{b}_i$ (for $i = 0, \dots, m$). We denote by $\boldsymbol{y} = [\boldsymbol{y}_0^T, \dots, \boldsymbol{y}_m^T]$ the big vector of multipliers. Then,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{c}^T\boldsymbol{x} + \sum_{i=0}^{m} \boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\boldsymbol{x}).$$

The Lagrange dual function is defined as before:

$$g(\boldsymbol{y}) := \inf_{\boldsymbol{x}\in\mathbb{R}^n} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}).$$

Then, we obtain the weak duality theorem by replacing the requirements that multipliers of $\leq$-inequalities must be nonnegative, by $\boldsymbol{y}_i \succeq_{K_i^*} \boldsymbol{0}$.

Theorem 3 (Weak duality for Conic Programming)**.** *For all vectors of Lagrange multipliers* $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \dots, \boldsymbol{y}_m)$*, with* $\boldsymbol{y}_1 \succeq_{K_1^*} \boldsymbol{0}, \dots, \boldsymbol{y}_m \succeq_{K_m^*} \boldsymbol{0}$*,*

$$g(\boldsymbol{y}) \leq p^*.$$

In other words,

$$d^* := \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0}\\ \forall i\in[m], \boldsymbol{y}_i \succeq_{K_i^*}\boldsymbol{0}}} \ \inf_{\boldsymbol{x}\in\mathbb{R}^n} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} \ \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0}\\ \forall i\in[m], \boldsymbol{y}_i \succeq_{K_i^*}\boldsymbol{0}}} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) = p^*.$$

The program on the left-hand side in the above equation is called the Lagrangian dual *of* (CP)*.*

Remark: With a slight abuse of notation, we could treat linear equalities as other constraints. To do this, we could rewrite $A_0\boldsymbol{x} = \boldsymbol{b}_0$ as $A_0\boldsymbol{x} \succeq_{K_0} \boldsymbol{b}_0$, where $K_0 = \{\boldsymbol{0}\} \subseteq \mathbb{R}^{n_0}$ is the trivial cone. This is an abuse of notation, because K_0 is not proper. However, we would still obtain the correct form for the dual, because the dual cone of K_0 is the whole space $\mathbb{R}^{n_0}$.

Proof. Let $\tilde{\boldsymbol{x}}$ be feasible for (CP), $\boldsymbol{y}_0 \in \mathbb{R}^{n_0}$ and $\boldsymbol{y}_i \in K_i^*$ (for all $i \in [m]$). Then,

$$g(\boldsymbol{y}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) \leq \mathcal{L}(\tilde{\boldsymbol{x}}, \boldsymbol{y}) = \boldsymbol{c}^T\tilde{\boldsymbol{x}} + \sum_{i=0}^{m} \boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\tilde{\boldsymbol{x}}).$$

In the above sum, we claim that all terms are ≤ 0. The term of index 0 is even $= 0$, because $\boldsymbol{b}_0 = A_0\tilde{\boldsymbol{x}}$. The other terms are scalar products of $\boldsymbol{b}_i - A_i\tilde{\boldsymbol{x}} \in -K_i$ and $\boldsymbol{y}_i \in K_i^*$, hence nonpositive by definition of the dual cone.

This already shows that $d^* \leq p^*$. It remains to show that p^* is equal to the $\inf - \sup$ problem on the right-hand side. This follows from

$$\sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0}\\ \forall i\in[m], \boldsymbol{y}_i \succeq_{K_i^*}\boldsymbol{0}}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \boldsymbol{c}^T\boldsymbol{x} & \text{if } \boldsymbol{x} \text{ is feasible} \\ +\infty & \text{otherwise.} \end{cases}$$

This again, is a consequence of the definition of dual cones. First, if $\boldsymbol{x}$ is feasible, we know that each term $\boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\boldsymbol{x})$ is ≤ 0, and equality is attained by taking $\boldsymbol{y}_i = \boldsymbol{0}$. Otherwise, if $\boldsymbol{x}$ violates the ith inequality, then $(\boldsymbol{b}_i - A_i\boldsymbol{x}) \notin -K_i$, so there exists $\boldsymbol{z}_i \in K_i^*$ such that $\boldsymbol{z}_i^T(\boldsymbol{b}_i - A_i\boldsymbol{x}) > 0$. Then, the term $\boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\boldsymbol{x})$ can be made arbitrarily large, by setting $\boldsymbol{y}_i = t\,\boldsymbol{z}_i$ for $t \to \infty$. ☐

Remark: If $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$ satisfies the assumptions of Theorem (3), i.e.,

$$\boldsymbol{y}_0 \in \mathbb{R}^{n_0}, \quad \boldsymbol{y}_i \succeq_{K_i^*} \boldsymbol{0}, \ \forall i \in [m],$$

we say that $\boldsymbol{y}$ is a *feasible vector of Lagrange multipliers*.

Theorem 4. *The Lagrangian dual of* (CP) *is the following conic programming problem:*

$$\begin{aligned} d^* = \sup_{\boldsymbol{y}_0,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m} \quad & \sum_{i=0}^m \boldsymbol{b}_i^T \boldsymbol{y}_i \\ s.t. \quad & \sum_{i=0}^m A_i^T \boldsymbol{y}_i = \boldsymbol{c} \\ & \boldsymbol{y}_i \succeq_{K_i^*} \boldsymbol{0}, \qquad (i = 1,\ldots,m). \end{aligned} \tag{CD}$$

Proof. We first write the primal problem as the inf − sup of the Lagrangian:

$$p^* = \inf_{\boldsymbol{x}\in\mathbb{R}^n} \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0} \\ \forall i\in[m], \boldsymbol{y}_i\succeq_{K_i^*}\boldsymbol{0}}} \boldsymbol{c}^T\tilde{\boldsymbol{x}} + \sum_{i=0}^m \boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\tilde{\boldsymbol{x}}).$$

Then, by definition, the dual of (CP) is obtained by switching the order of inf and sup:

$$\begin{aligned} d^* &= \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0} \\ \forall i\in[m], \boldsymbol{y}_i\succeq_{K_i^*}\boldsymbol{0}}} \inf_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{c}^T\tilde{\boldsymbol{x}} + \sum_{i=0}^m \boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\tilde{\boldsymbol{x}}) \\ &= \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0} \\ \forall i\in[m], \boldsymbol{y}_i\succeq_{K_i^*}\boldsymbol{0}}} \inf_{\boldsymbol{x}\in\mathbb{R}^n} \sum_{i=0}^m \boldsymbol{y}_i^T\boldsymbol{b}_i + \boldsymbol{x}^T(\boldsymbol{c} - \sum_{i=0}^m A_i^T\boldsymbol{y}_i) \\ &= \sup_{\substack{\boldsymbol{y}_0\in\mathbb{R}^{n_0} \\ \forall i\in[m], \boldsymbol{y}_i\succeq_{K_i^*}\boldsymbol{0}}} \begin{cases} \sum_{i=0}^m \boldsymbol{y}_i^T\boldsymbol{b}_i & \text{if } \boldsymbol{c} - \sum_{i=0}^m A_i^T\boldsymbol{y}_i = \boldsymbol{0} \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

because $\inf_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{x}^T\boldsymbol{u}$ is finite (and equal to 0) iff $\boldsymbol{u} = \boldsymbol{0}$. We have shown that the dual problem is to find the feasible lagrange multipliers that maximize $\sum_{i=0}^m \boldsymbol{y}_i^T\boldsymbol{b}_i$, under the additional that $\sum_{i=0}^m A_i^T\boldsymbol{y}_i = \boldsymbol{c}$, which is the conic programming problem (CD). ☐

3 Strong Duality

3.1 Constraint qualification

We will see that when an optimization problem is convex, then usually strong duality holds. Here, "usually" indicates that some pathological cases exist. Fortunately, we will see that there exists some sufficient conditions, called *constraints qualifications*, which guarantee strong duality in most cases:

Definition 5. We say that the constraints of a convex optimization problem are *qualified* if they satisfy a condition which ensures that strong duality holds, i.e., $p^* = d^*$.

There exist many types of constraint qualifications. In this lecture, we will focus on *Slater's condition*.

3.2 Slater's condition for conic programming

We consider a pair of optimization problems of the form

$$\begin{aligned} p^* = \inf_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ s.t. \quad & A_0\, \boldsymbol{x} = \boldsymbol{b}_0 \\ & A_i\, \boldsymbol{x} \succeq_{K_i} \boldsymbol{b}_i, \qquad (i = 1, \ldots, m), \end{aligned} \tag{CP}$$

and

$$\begin{aligned} d^* = \sup_{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m} \quad & \sum_{i=0}^{m} \boldsymbol{b}_i^T\, \boldsymbol{y}_i \\ s.t. \quad & \sum_{i=0}^{m} A_i^T\, \boldsymbol{y}_i = \boldsymbol{c} \\ & \boldsymbol{y}_i \succeq_{K_i^*} \boldsymbol{0}, \qquad (i = 1, \ldots, m), \end{aligned} \tag{CD}$$

where for all i, K_i is a proper cone. Denote by $I_N \subseteq [m]$ the set of indices such that K_i is not polyhedral.

Definition 6 (Essentially strict feasibility)**.** Problem (CP) is called *(essentially) strictly feasible* if there is a feasible vector $\boldsymbol{x}$ that satisfies all (nonlinear) conic inequalities strictly:

$$\exists \boldsymbol{x} \in \mathbb{R}^n : \; A_0\, \boldsymbol{x} = \boldsymbol{b}_0, \quad A_i\, \boldsymbol{x} \succ_{K_i} \boldsymbol{b}_i \quad (\forall i \in [m]).$$
$$(\text{or } A_i\, \boldsymbol{x} \succeq_{K_i} \boldsymbol{b}_i, \forall i \notin I_N \text{ and } A_i\, \boldsymbol{x} \succ_{K_i} \boldsymbol{b}_i, \forall i \in I_N).$$

Problem (CD) is called *(essentially) strictly feasible* if there is a dual feasible vector $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$ which satisfies all (nonlinear) conic inequalities strictly:

$$\exists (\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m) : \; \sum_{i=0}^{m} A_i^T\, \boldsymbol{y}_i = \boldsymbol{c}, \quad \boldsymbol{y}_i \succ_{K_i^*} \boldsymbol{0}, \qquad (\forall i \in [m]).$$
$$(\text{or } \boldsymbol{y}_i \succeq_{K_i} \boldsymbol{0}, \forall i \notin I_N \text{ and } \boldsymbol{y}_i \succ_{K_i} \boldsymbol{0}, \forall i \in I_N).$$

Theorem 5 (Strong duality Theorem for conic programming)**.** Consider a pair of primal and dual conic programs (CP) and (CD). Then,

1. *[Weak duality]:* $d^* \leq p^*$.
2. *[Symmetry]:* The duality is symmetric: (CD) is a conic program, and the dual of (CD) is (CP) (or, to be more precise, the dual of (CD) is equivalent to (CP)).
3. *[Strong duality]:* If one of the programs (CP),(CD) is essentially strictly feasible and bounded, then the other is solvable (dual attainement) and $p^* = d^*$ (no duality gap). In particular, if both problems are essentially strictly feasible, then there exists a pair $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ of primal-dual optimal solutions.
4. *[Optimality conditions]:* Let $\boldsymbol{x}^*$ be feasible for (CP) and $\boldsymbol{y}^*$ be feasible for (CD). If strong duality holds (in particular, if one of the two problems is e.s.f.), then the following statements are equivalent
   (i) *[optimality]*: $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a pair of primal-dual optimal solutions
   (ii) *[no duality gap]*: $\boldsymbol{c}^T \boldsymbol{x}^* = \boldsymbol{b}^T \boldsymbol{y}^*$ (here we use $\boldsymbol{b}^T \boldsymbol{y}^*$ as a shorthand for $\sum_{i=0}^m \boldsymbol{b}_i^T\, \boldsymbol{y}_i^*$)

(iii) *[complementary slackness]*: $\boldsymbol{y}_i^{*T}(A_i\,\boldsymbol{x}^* - \boldsymbol{b}_i) = 0, \quad (i = 1, \ldots, m).$

3.3 Proof of the strong duality theorem for conic programming

Proof. **Weak Duality**: We already proved this.

Symmetry: Let us compute the dual problem of (CD). To this end, we first express d^* as a $\sup - \inf$ problem, and then we will switch the order of sup and inf:

$$d^* = \sup_{\boldsymbol{y}_i \in \mathbb{R}^{n_i},\ \forall i \in [m]} \quad \sum_{i=0}^{m} \boldsymbol{y}_i^T \boldsymbol{b}_i + \inf_{\boldsymbol{x}\in\mathbb{R}^n} \ \boldsymbol{x}^T(\boldsymbol{c} - \sum_{i=0}^{m} A_i^T \boldsymbol{y}_i) + \sum_{i=1}^{m} \inf_{\boldsymbol{z}_i \in K_i} \boldsymbol{z}_i^T \boldsymbol{y}_i.$$

So, the dual of (CD) is

$$\begin{aligned}
&\inf_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}_i \succeq_{K_i} 0} \ \sup_{\boldsymbol{y}_i\in\mathbb{R}^{n_i}} \quad \sum_{i=0}^{m} \boldsymbol{y}_i^T \boldsymbol{b}_i + \boldsymbol{x}^T(\boldsymbol{c} - \sum_{i=0}^{m} A_i^T \boldsymbol{y}_i) + \sum_{i=1}^{m} \boldsymbol{z}_i^T \boldsymbol{y}_i \\
&= \inf_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}_i \succeq_{K_i} 0} \boldsymbol{c}^T \boldsymbol{x} + \sup_{\forall i,\ \boldsymbol{y}_i\in\mathbb{R}^{n_i}} \quad \boldsymbol{y}_0^T(\boldsymbol{b}_0 - A_0 \boldsymbol{x}) + \sum_{i=1}^{m} \boldsymbol{y}_i^T(\boldsymbol{b}_i - A_i\boldsymbol{x} + \boldsymbol{z}_i)
\end{aligned}$$

The supremum above has finite value (and is equal to 0) iff $A_0\boldsymbol{x} = \boldsymbol{b}_0$ and $\boldsymbol{z}_i = A_i\boldsymbol{x} - \boldsymbol{b}_i, \forall i \in [m]$. So the dual of (CD) is

$$\begin{aligned}
\underset{\substack{\boldsymbol{x}\in\mathbb{R}^n,\\ \boldsymbol{z}_i\in\mathbb{R}^{n_i}, \forall i\in[m]}}{\textbf{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\
s.t. \quad & A_0\boldsymbol{x} = \boldsymbol{b}_0 \\
& \boldsymbol{z}_i = A_i\boldsymbol{x} - \boldsymbol{b}_i, \quad (\forall i \in [m]) \\
& \boldsymbol{z}_i \succeq \boldsymbol{0}, \quad (\forall i \in [m]).
\end{aligned}$$

Finally, we observe that $\boldsymbol{z}_i$ plays the role of a slack variable, and the above problem is equivalent to (CP).

Strong Duality: We will only handle the case of strict feasibility. For a detailed proof of the refined result with *essentially strict feasibility*, cf. [1, Section 7.1].

We have already seen in a previous lecture that every conic programming problem was equivalent to a conic programming problem in standard form

$$p^* = \inf\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n,\ A\boldsymbol{x} \succeq_K \boldsymbol{b}\} \tag{SCP}$$

where $K \subset \mathbb{R}^m$ is a proper cone. The associated dual standard form is

$$d^* = \sup\{\boldsymbol{b}^T\boldsymbol{y} \mid \boldsymbol{y} \in \mathbb{R}^m,\ A^T\boldsymbol{y} = \boldsymbol{c},\ \boldsymbol{y} \succeq_{K^*} \boldsymbol{0}\}. \tag{SCD}$$

By symmetry, we can assume (without loss of generality) that the strict feasibility condition is satisfied by the primal problem (SCP), that is,

$$\exists \boldsymbol{x}_0 \in \mathbb{R}^n : A\boldsymbol{x}_0 \succ_K \boldsymbol{b}.$$

We already now that $d^* \leq p^*$ by weak duality, and we want to show that the optimal value is attained in the dual problem, and $d^* \geq p^*$, that is

$$\exists \boldsymbol{y}^* \succeq_{K^*} \boldsymbol{0} : \quad A^T\boldsymbol{y}^* = \boldsymbol{c} \ \text{ and } \ \boldsymbol{b}^T\boldsymbol{y}^* \geq p^*. \tag{1}$$

We start with the case $\boldsymbol{c} \neq \boldsymbol{0}$, and we define

$$M = \{A\boldsymbol{x} - \boldsymbol{b} : \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{c}^T\boldsymbol{x} \leq p^*\}.$$

We observe that $M \neq \emptyset$ (because (SCP) is bounded, so $p^* > -\infty$ and $\{\boldsymbol{x} : \boldsymbol{c}^T\boldsymbol{x} \leq p^*\}$ is a halfspace). Moreover, we claim that $M \cap \textbf{int}\, K = \emptyset$. Otherwise, there would be an $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $A\bar{\boldsymbol{x}} \succ_K \boldsymbol{b}$ and $\boldsymbol{c}^T\bar{\boldsymbol{x}} \leq p^*$, so $A\bar{\boldsymbol{x}} \succeq_K \boldsymbol{b}$

holds in a neighborhood of $\bar{\boldsymbol{x}}$, and we could find an $\boldsymbol{x}^*$ such that $\boldsymbol{c}^T\boldsymbol{x}^* < \boldsymbol{c}^T\bar{\boldsymbol{x}} \leq p^*$ in this neighborhood, which contradicts that p^* is the optimal value of (SCP).

This shows that we can invoke the separating hyperplane theorem:

$$\exists \boldsymbol{z} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\},\ \exists u \in \mathbb{R} : \begin{cases} \boldsymbol{z}^T\boldsymbol{y} \leq u, & \forall \boldsymbol{y} \in M \\ \boldsymbol{z}^T\boldsymbol{y} \geq u, & \forall \boldsymbol{y} \in \mathbf{int}\, K. \end{cases}$$

We can assume $u \geq 0$ w.l.o.g. Indeed, $\mathbf{int}\, K$ is a cone, so if $\exists \bar{\boldsymbol{y}} \in \mathbf{int}\, K : \boldsymbol{z}^T\bar{\boldsymbol{y}} < 0$, then we would have $\inf_{\boldsymbol{y}\in \mathbf{int}\, K} \boldsymbol{z}^T\boldsymbol{y} = -\infty$ by taking $\boldsymbol{y} = t\,\bar{\boldsymbol{y}}$ for $t \to \infty$. On the other hand, we must have $u \leq 0$, which can be seen by taking $\boldsymbol{y} \to \boldsymbol{0}$. So, we can assume that $u = 0$ and

$$\forall \boldsymbol{y} \in \mathbf{int}\, K, \quad \boldsymbol{z}^T\boldsymbol{y} \geq 0 \implies \forall \boldsymbol{y} \in K, \quad \boldsymbol{z}^T\boldsymbol{y} \geq 0 \implies \boldsymbol{z} \in K^*.$$

For $\boldsymbol{x}$ such that $\boldsymbol{c}^T\boldsymbol{x} \leq p^*$, we have $\boldsymbol{y} = A\boldsymbol{x} - \boldsymbol{b} \in M$, and the separating hyperplane theorem tells us that $\boldsymbol{z}^T(A\boldsymbol{x} - \boldsymbol{b}) \leq u = 0$, that is,

$$\boldsymbol{c}^T\boldsymbol{x} \leq p^* \implies \boldsymbol{z}^T A\boldsymbol{x} \leq \boldsymbol{z}^T\boldsymbol{b}. \tag{2}$$

In particular, the linear form $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, A^T\boldsymbol{z}\rangle$ is bounded on the halfspace $\{\boldsymbol{x} : \boldsymbol{c}^T\boldsymbol{x} \leq p^*\}$. Clearly, this is only possible if $A^T\boldsymbol{z}$ is a nonnegative multiple of $\boldsymbol{c}$: $\exists \mu \geq 0 : A^T\boldsymbol{z} = \mu\boldsymbol{c}$.

Now, the vector $\boldsymbol{y}^* = \boldsymbol{z}/\mu$ is going to be our candidate to prove the statement (1). But first, we must show that $\mu \neq 0$.

Assume $\mu = 0$. Then, we have $A^T\boldsymbol{z} = \boldsymbol{0}$. Then, Eq.(2) shows that $\boldsymbol{b}^T\boldsymbol{z} \geq 0$. Now, we use our strictly feasible vector $\boldsymbol{x}_0$. The conic inequalities $A\boldsymbol{x}_0 \succ_K \boldsymbol{b}$ and $\boldsymbol{z} \succeq_{K^*} \boldsymbol{0}, \boldsymbol{z} \neq \boldsymbol{0}$ imply $\langle A\boldsymbol{x}_0 - \boldsymbol{b}, \boldsymbol{z}\rangle > 0$. But then, we have

$$\underbrace{\boldsymbol{x}_0^T A^T\boldsymbol{z}}_{=0} > \boldsymbol{b}^T\boldsymbol{z} \geq 0,$$

a contradiction.

Finally, we write $\boldsymbol{y}^* = \frac{1}{\mu}\boldsymbol{z}$. We have $\boldsymbol{y}^* \succeq_{K^*} \boldsymbol{0}$ and $A^T\boldsymbol{y}^* = \boldsymbol{c}$. For any $\boldsymbol{x}$ such that $\boldsymbol{c}^T\boldsymbol{x} \leq p^*$, we also have, by Eq. (2), $\boldsymbol{x}^T A^T\boldsymbol{z} \leq \boldsymbol{b}^T\boldsymbol{z} \implies \frac{1}{\mu}\boldsymbol{x}^T A^T\boldsymbol{z} = \boldsymbol{x}^T A^T\boldsymbol{y}^* = \boldsymbol{x}^T\boldsymbol{c} \leq \boldsymbol{b}^T\boldsymbol{y}^*$. Hence, $p^* \leq \boldsymbol{b}^T\boldsymbol{y}^*$. This concludes the proof of (1) for the case $\boldsymbol{c} \neq \boldsymbol{0}$.

If $\boldsymbol{c} = \boldsymbol{0}$, (SCP) is a feasibility problem, so $p^* = 0$ because the problem is feasible. Then, $\boldsymbol{y}^* = \boldsymbol{0}$ is dual optimal, since $\boldsymbol{b}^T\boldsymbol{y}^* = 0$ and we know that $d^* \leq 0$ by weak duality.

Optimality Conditions: If strong duality holds, then $\boldsymbol{x}^*, \boldsymbol{y}^*$ are optimal iff $p^* = \boldsymbol{c}^T\boldsymbol{x}^* = d^* = \boldsymbol{b}^T\boldsymbol{y}^*$. In other words, the duality gap $\boldsymbol{c}^T\boldsymbol{x}^* - \boldsymbol{b}^T\boldsymbol{y}^*$ is 0. But, since $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ are feasible, the duality gap is

$$\boldsymbol{c}^T\boldsymbol{x}^* - \boldsymbol{b}^T\boldsymbol{y}^* = \boldsymbol{x}^* A^T\boldsymbol{y}^* - \boldsymbol{b}^T\boldsymbol{y}^* = \boldsymbol{y}^{*T}(A^T\boldsymbol{x}^* - \boldsymbol{b}),$$

so optimality is equivalent to the condition of complementary slackness. □

3.4 The strong duality theorem for semidefinite programming

The above form of the duality theorem assumes that both the primal variables $\boldsymbol{x}$ and the dual variables $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$ are *vectors* of $\mathbb{R}^n$ and $\mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$, respectively. You may wonder how this is consistent with the special case of semidefinite programming, in which variables can be matrices. In fact, the above theorem *implicitly* assumes that every variable has been *vectorized* when needed (for semidefinite programming, you can see $\boldsymbol{x}$ as the vector of all n^2 elements of the matrix X, and the cone K is the cone of all vectorized semidefinite matrices). A more elegant formulation, which avoids the use of vectorization, is as follows. We first need to define the notion of *adjoint operators*:

> **Definition 7** (Adjoint operator)**.** The *adjoint* of a linear operator $A : \mathcal{X} \mapsto \mathcal{Y}$ is the unique linear operator $A^* : \mathcal{Y} \mapsto \mathcal{X}$ such that
>
> $$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X} \times \mathcal{Y},\ \langle A(\boldsymbol{x}), \boldsymbol{y}\rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, A^*(\boldsymbol{y})\rangle_{\mathcal{X}}.$$

Example:
Consider the operator $A : \mathbb{R}^n \to \mathbb{R}^m$, where the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with the usual scalar product. Then, the adjoint of the operator $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is $A^* : \boldsymbol{y} \mapsto A^T\boldsymbol{y}$, because

$$(A\boldsymbol{x})^T\boldsymbol{y} = \boldsymbol{x}^T(A^T\boldsymbol{y}).$$

Consider the operator $F : \mathbb{R}^n \to \mathbb{S}^m$, $\boldsymbol{x} \mapsto \sum_{i=1}^n x_i M_i$, for some matrices $M_i \in \mathbb{S}^m$, and where $\mathbb{R}^n$ and $\mathbb{S}^m$ are equipped with their usual scalar procuct. Then, the adjoint of F is $F^* : Y \mapsto \begin{bmatrix} \langle M_1, Y\rangle \\ \vdots \\ \langle M_n, Y\rangle \end{bmatrix}$, because #2

$$\langle F(\boldsymbol{x}), Y\rangle = \sum_{i=1}^n x_i \langle M_i, Y\rangle = \boldsymbol{x}^T \begin{bmatrix} \langle M_1, Y\rangle \\ \vdots \\ \langle M_n, Y\rangle \end{bmatrix}.$$

Now, the primal variable $\boldsymbol{x}$ lives in a $n-$dimensional Euclidean vector space $\mathcal{X}$ equipped with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{X}}$, and the dual variable $\boldsymbol{y}$ lives in the $\sum_{i=0}^m n_i-$dimensional Euclidean vector space $\big(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}}\big)$.
Then, the primal and dual conic problems become:

$$\begin{aligned} p^* = \inf_{\boldsymbol{x}\in\mathcal{X}} \quad & \langle \boldsymbol{c}, \boldsymbol{x}\rangle_{\mathcal{X}} \\ s.t. \quad & A(\boldsymbol{x}) - \boldsymbol{b} \in \{0\}^{n_0} \times K_1 \times \cdots \times K_m \end{aligned} \tag{CP'}$$

and

$$\begin{aligned} d^* = \sup_{\boldsymbol{y}} \quad & \langle \boldsymbol{b}, \boldsymbol{y}\rangle_{\mathcal{Y}} \\ s.t. \quad & A^*(\boldsymbol{y}) = \boldsymbol{c} \\ & \boldsymbol{y} \in \mathbb{R}^{n_0} \times K_1^* \times \cdots \times K_m^* \end{aligned} \tag{CD'}$$

and the duality theorem translates naturally for the above pair of problems in Euclidean vector spaces, by replacing all scalar products $\boldsymbol{u}^T\boldsymbol{v}$ by an inner product $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ in the appropriate space.

For convenience, we next give the duality theorem for the case of semidefinite programs in standard form:

Theorem 6 (Strong duality theorem for SDPs)**.** *Consider the pair of primal and dual SDPs*

$$\begin{aligned} p^* = \inf_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ s.t. \quad & \sum_{i=1}^n x_i M_i \succeq C \end{aligned} \tag{SP}$$

and

$$\begin{aligned} d^* = \sup_{Y\in\mathbb{S}^m} \quad & \langle C, Y\rangle \\ s.t. \quad & \langle M_i, Y\rangle = c_i, \qquad (\forall i \in [n]) \\ & Y \succeq 0, \end{aligned} \tag{SD}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $C, M_i \in \mathbb{S}^m$ $(\forall i \in [n])$. Then,

1. [Weak duality]: $d^* \leq p^*$.

2. [Symmetry]: *The duality is symmetric: (SD) is an SDP, and the dual of (SD) is (SP) (or, to be more precise, the dual of (SD) is equivalent to (SP)).*

3. [Strong duality]: *If one of the programs (SP),(SD) is strictly feasible and bounded, then the other is solvable (dual attainement) and $p^* = d^*$ (no duality gap). In particular, if both problems are strictly feasible, then there exists a pair $(\boldsymbol{x}^*, Y^*)$ of primal-dual optimal solutions.*

4. [Optimality conditions]: *Let $\boldsymbol{x}^*$ be feasible for (SP) and Y^* be feasible for (SD). If strong duality holds (in particular, if one of the two problems is s.f.), then the following statements are equivalent*

   (i) [optimality]*: $(\boldsymbol{x}^*, Y^*)$ is a pair of primal-dual optimal solutions*

   (ii) [no duality gap]*: $\boldsymbol{c}^T\boldsymbol{x}^* = \langle C, Y^*\rangle$*

   (iii) [complementary slackness]*: $Y^* \left(\sum_{i=1}^n x_i^* M_i - C\right) = 0$.*

Proof. This theorem is the direct translation of Theorem 5 for the case of semidefinite programming, except for one thing. The complementary slackness condition should be $\langle Y^*, \sum_{i=1}^n x_i^* M_i - C\rangle = 0$. However, since Y^* and $\boldsymbol{x}^*$ are feasible, this is the scalar product of two symmetric matrices, and we claim that for all $U \succeq 0, V \succeq 0$,

$$\langle U, V\rangle = 0 \iff UV = 0.$$

The implication $\Longleftarrow$ is trivial. For $\Longrightarrow$, take some decompositions $U = HH^T$ and $V = KK^T$. Then,

$$\langle U, V\rangle = \operatorname{trace} HH^T KK^T = \operatorname{trace} K^T HH^T K = \operatorname{trace}(H^T K)^T (H^T K) = \|H^T K\|_F^2.$$

So, $\langle U, V\rangle = 0 \implies H^T K = 0 \implies UV = HH^T KK^T = 0.$ □

3.5 Importance of the strict feasibility

We now discuss the importance of the strict feasibility for strong duality to hold. In fact, the strong duality theorem contains two assumptions: one of the two problems must be strictly feasible, and it must be bounded.

The boundedness condition is not so critical, because e.g. if the primal is unbouded, then $p^* = -\infty$, and by weak duality $d^* = -\infty$, too. So in fact, strong duality holds, and the only thing we loose is the property of "dual attainment". Indeed, $d^* = -\infty$ means that the dual problem is infeasible.

The condition of strict feasibility, however, is very important. This is a very different situation as for the case of *linear programming*, where, as you may know, strong duality only fails when both problems are infeasible, i.e., $d^* = -\infty$ and $p^* = \infty$. This is consistent with the condition of *essentially strict feasibility*: since in a linear program all constraints are linear, every LP satisfies the Slater's condition. So the only case where strong duality does not hold in linear programming is when problems are infeasible:

- If the two problems are feasible, then by weak duality they are also bounded, so strong duality holds.
- if exactly one of the two problems is feasible, then this problem is unbounded and the other one is infeasible, so $p^* = d^* = \pm\infty$.
- If the two problems are infeasible, then strong duality fails: $d^* = -\infty$ and $p^* = \infty$.

When there are nonlinear constraints, there are more pathological cases. For example, consider the SOCP

$$p^* = \inf\{y |\, (x, y) \in \mathbb{R}^2, \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| \leq x\} \tag{3}$$

The constraint is equivalent to $x \geq 0$ and $x^2 + y^2 \leq x^2 \implies y = 0$, so $p^* = 0$.

Now, les us form the dual problem. In conic form, the primal problem reads

$$p^* = \inf\{y|\,(x,y)\in\mathbb{R}^2, \begin{bmatrix} x \\ y \\ x \end{bmatrix} \succeq_{\mathbb{L}^3_+} \mathbf{0}\}. \tag{4}$$

So, since $\mathbb{L}^3_+$ is self-dual, we can substitute in the dual form (SCD) with $\boldsymbol{b}=\mathbf{0}, \boldsymbol{c}=[0,1]^T$, and $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^T$. We find

$$d^* = \sup\{0|\, \boldsymbol{z}\in\mathbb{R}^3, z_1+z_3=0, z_2=1,\ \boldsymbol{z}\succeq_{\mathbb{L}^3_+}\mathbf{0},\}. \tag{5}$$

This problem is infeasible, because the conic constraint implies $z_1^2+1 \leq (-z_1)^2 \iff 1\leq 0$. Hence, $d^*=-\infty$.

This example illustrates two weaknesses of the duality theorem for conic programming:

- First, strong duality can fail. Here, the primal is feasible and solvable, but the dual is not even feasible. This is because the primal problem is not strictly feasible (the inequality $x^2+y^2\leq x^2$ cannot be satisfied *strictly*).
- Second, the problem is equivalent to the LP $\inf\{y|\ (x,y)\in\mathbb{R}^2, x\geq 0,\ y=0\}$, for which strong duality clearly holds. Hence, the formulation of the problem matters!

In the exercises, we will see that there are even examples where both the primal and the dual are feasible, so p^* and d^* are finite, but $d^*<p^*$.

3.6 Slater's condition for convex nonlinear programming

Definition 8. Consider the problem (NLP), and denote by

$$\mathcal{D} := (\bigcap_{i=0}^m \mathbf{dom}\, f_i) \cap (\bigcap_{j=1}^p \mathbf{dom}\, h_j)$$

its domain. We say that (NLP) satisfies Slater's condition if it is *strictly feasible*:

$$\exists \boldsymbol{x}\in\mathbf{int}\,\mathcal{D} : f_0(\boldsymbol{x})<\infty, \qquad f_i(\boldsymbol{x})<0, \forall i\in[m], \qquad h_j(\boldsymbol{x})=0, \forall j\in[p].$$

More generally, we also say that that (NLP) satisfies Slater's condition if it is *essentially strictly feasible*, that is, there exists a feasible $\boldsymbol{x}\in\mathbf{relint}\,\mathcal{D}$ that satisfies all *non-affine inequalities* strictly ($f_i(\boldsymbol{x})<0$).

Theorem 7 (Strong Duality under Slater's condition)**.** *Let* (NLP) *be a convex optimization problem that satisfies Slater's condition. Then,*

- *there is no duality gap, i.e.,* $p^*=d^*$*;*
- *moreover, if* (NLP) *is bounded from below, then the optimal value can be attained in the dual problem:*

$$\exists\boldsymbol{\lambda}\in\mathbb{R}^m_+, \exists\boldsymbol{\mu}\in\mathbb{R}^p:\ g(\boldsymbol{\lambda},\boldsymbol{\mu}) = d^* = p^*.$$

We are not going to prove this result, as the proof uses similar ideas as the proof of Theorem (5) for conic programming.

3.7 Karush Kuhn Tucker optimality conditions

We now derive necessary and sufficient optimality conditions for convex nonlinear programming.

Theorem 8 (Karush-Kuhn-Tucker)**.** *Let* (NLP) *be a convex optimization problem, in which the functions* $f_0, \ldots, f_m$ *are differentiable, and assume that* strong duality holds *(in particular, this is the case when the constraints satisfy Slater's condition). [Note that since the problem is convex, the functions* h_j *must be affine, hence they are always differentiable.] Then,* $(\boldsymbol{x}, (\boldsymbol{\lambda}, \boldsymbol{\mu}))$ *is a pair of primal and dual optimal solutions if and only if the following equations (often abbreviated as KKT conditions for Karush-Kuhn-Tucker), hold:*

$$\begin{aligned}
\textit{[Stationarity]:} \quad & \nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j \nabla h_j(\boldsymbol{x}) = \boldsymbol{0} \\
\textit{[Primal feasibility]:} \quad & f_i(\boldsymbol{x}) \leq 0,\ \forall i \in [m], \qquad h_j(\boldsymbol{x}) = 0,\ \forall j \in [p]; \\
\textit{[Dual feasibility]:} \quad & \boldsymbol{\lambda} \geq \boldsymbol{0}; \\
\textit{[Complementary slackness]:} \quad & \lambda_i f_i(\boldsymbol{x}) = 0,\ \forall i \in [m].
\end{aligned}$$

Remark: For *non-convex optimization problems*, when a constraint qualification holds (such as Slater's), it can also be shown that the KKT conditions are *necessary* conditions for local optimality: If $\boldsymbol{x}$ is a local optimum, then there exist Lagrange multipliers $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ such that the KKT conditions hold.

Proof. $\Longrightarrow$: Let $(\boldsymbol{x}^*, (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*))$ be a pair of primal and dual optimal solutions. Then, it is clear that the conditions of primal and dual feasibility hold. Then, the Lagrange dual function at $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$ is

$$\begin{aligned}
d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \inf_{\boldsymbol{x}} f_0(\boldsymbol{x}) + \sum_{i \in [m]} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{j \in [p]} \mu_j^* h_j(\boldsymbol{x}) \\
&\leq f_0(\boldsymbol{x}^*) + \sum_{i \in [m]} \underbrace{\lambda_i^* f_i(\boldsymbol{x}^*)}_{\leq 0} + \sum_{j \in [p]} \underbrace{\mu_j^* h_j(\boldsymbol{x}^*)}_{=0} \\
&\leq f_0(\boldsymbol{x}^*) = p^*.
\end{aligned}$$

Since strong duality holds, $p^* = d^*$, and the above inequalities must be equalities. It follows that, for all $i \in [m]$, $\lambda_i^* f_i(\boldsymbol{x}^*) = 0$ (this is the condition of complementary slackness), and $\boldsymbol{x}^*$ minimizes the Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ over $\mathbb{R}^n$, so its gradient vanishes at $\boldsymbol{x}^*$ (this is the condition of stationarity).

$\Longleftarrow$: Let $(\boldsymbol{x}^*, (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*))$ satisfy the KKT conditions. The stationarity condition tells us that $\boldsymbol{x}^*$ minimizes the Lagrangian $\boldsymbol{x} \mapsto \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ over $\mathbb{R}^n$ (because this is a convex function), that is, $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Moreover,

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f_0(\boldsymbol{x}^*) + \sum_{i \in [m]} \underbrace{\lambda_i^* f_i(\boldsymbol{x}^*)}_{=0} + \sum_{j \in [p]} \underbrace{\mu_j^* h_j(\boldsymbol{x}^*)}_{=0} = f_0(\boldsymbol{x}^*),$$

where the "= 0" follow from primal feasibility and complementary slackness. We have shown that $f_0(\boldsymbol{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. There is no duality gap, so $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ are optimal. □

4 Sensitivity analysis

In this section, we consider a perturbed version of (NLP):

$$\begin{aligned}
p^*(\boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f_0(\boldsymbol{x}) \\
s.t. \quad & f_i(\boldsymbol{x}) \leq u_i,\ \forall i \in [m] \\
& h_j(\boldsymbol{x}) \leq v_j,\ \forall j \in [p]
\end{aligned} \tag{$P_{\boldsymbol{u},\boldsymbol{v}}$}$$

When we perform an analysis of sensitivity of Problem (NLP), we ask ourselves how the optimal value p^* varies when we perturb the right-hand sides $\boldsymbol{u}$ and $\boldsymbol{v}$ of the constraints.

We first have a global sensitivity result:

Theorem 9 (Global sensitivity)**.** *Let* (NLP) *be a convex optimization problem for which strong duality holds. Denote by* $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *a pair of optimal Lagrange multiplers. Then,*

$$p^*(\boldsymbol{u}, \boldsymbol{v}) \geq p^*(\mathbf{0}, \mathbf{0}) - \boldsymbol{\lambda}^{*T}\boldsymbol{u} - \boldsymbol{\mu}^{*T}\boldsymbol{v}$$

Proof. By strong duality, $p^*(\mathbf{0}, \mathbf{0}) = p^* = d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

Now, let $\boldsymbol{x}$ be any vector which is feasible for the perturbed problem $(P_{\boldsymbol{u},\boldsymbol{v}})$. It holds

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq f_0(\boldsymbol{x}) + \sum_{i\in[m]} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{j\in[p]} \mu_j^* h_j(\boldsymbol{x})$$

Since $\boldsymbol{\lambda}^* \geq \mathbf{0}$ by dual feasibility, and $f_i(\boldsymbol{x}) \leq u_i, h_j(\boldsymbol{x}) = v_j$, we get

$$p^*(\mathbf{0}, \mathbf{0}) = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq f_0(\boldsymbol{x}) + \boldsymbol{\lambda}^{*T}\boldsymbol{u} + \boldsymbol{\mu}^{*T}\boldsymbol{v}.$$

The result of the theorem follows by taking the infimum over all $\boldsymbol{x}$ in the feasible set of $(P_{\boldsymbol{u},\boldsymbol{v}})$. □

We also mention a local sensitivity result, which helps for the interpretation of the optimal dual values:

Theorem 10 (Local sensitivity)**.** *Let* (NLP) *be a convex optimization problem for which strong duality holds. Denote by* $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *a pair of optimal Lagrange multiplers, and assume that* $p^*(\boldsymbol{u}, \boldsymbol{v})$*, seen as a function of* $\boldsymbol{u}$ *and* $\boldsymbol{v}$*, is differentiable at* $\boldsymbol{u} = \mathbf{0}, \boldsymbol{v} = \mathbf{0}$*. Then, for all* $(i, j) \in [m] \times [p]$,

$$\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*, \qquad \frac{\partial p^*(0,0)}{\partial v_j} = -\mu_j.$$

Proof. By definition,

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t\to 0} \frac{p^*(t\boldsymbol{e}_i, \mathbf{0}) - p^*}{t}.$$

From the global sensitivity result, we know that for all t, $p^*(t\boldsymbol{e}_i, \mathbf{0}) \geq p^* - t\lambda_i^*$. That is, $\frac{p^*(t\boldsymbol{e}_i,\mathbf{0})-p^*}{t} \geq -\lambda_i^*$ for $t > 0$, and $\frac{p^*(t\boldsymbol{e}_i,\mathbf{0})-p^*}{t} \leq -\lambda_i^*$ for $t < 0$. If $p^*(\cdot, \cdot)$ is differentiable, then the limit exists and must be equal to $-\lambda_i$.

The proof for the derivative relative to v_j is similar.

□

References

[1] Nemirovski, A. *Introduction to Linear Optimization*. Lecture notes ISYE 6661, Georgia Tech, 2012. `http://www2.isye.gatech.edu/~nemirovs/OPTI_LectureNotes2016.pdf`.