$$= \begin{cases} -\vec{\nu}^\top \vec{y}, & \text{if } \vec{c} + A^\top \vec{\nu} - \vec{\lambda} = \vec{0} \\ -\infty, & \text{otherwise.} \end{cases} \tag{7.73}$$

(It is a good exercise to figure out why this last equality is correct.) Thus, the dual is

$$\begin{aligned} d^\star = \max_{\substack{\vec{\lambda} \in \mathbb{R}^n \\ \vec{\nu} \in \mathbb{R}^m}} \quad & g(\vec{\lambda}, \vec{\nu}) \\ \text{s.t.} \quad & \lambda_i \geq 0, \qquad \forall i \in \{1, \dots, n\}. \end{aligned} \tag{7.74}$$

Expanding out g, this problem is equivalent to

$$\begin{aligned} d^\star = \max_{\substack{\vec{\lambda} \in \mathbb{R}^n \\ \vec{\nu} \in \mathbb{R}^m}} \quad & -\vec{\nu}^\top \vec{y} \\ \text{s.t.} \quad & \lambda_i \geq 0, \qquad \forall i \in \{1, \dots, n\} \\ & \vec{c} + A^\top \vec{\nu} - \vec{\lambda} = \vec{0}. \end{aligned} \tag{7.75}$$

Example 169 (Shadow Prices)**.** In this example, we determine an economic interpretation of Lagrange multipliers.

Suppose we have 200 kilos of merlot grapes and 300 kilos of shiraz grapes. Consider the following possible blends:

- 4 kilos merlot, 1 kilo shiraz for \$20 per bottle.
- 2 kilos merlot, 3 kilos shiraz for \$15 per bottle.

We want to maximize our profits. Suppose we make q_1 bottles of the first blend, and q_2 of the second. Our optimization problem is:

$$\begin{aligned} p^\star = \max_{q_1, q_2 \in \mathbb{R}} \quad & 20q_1 + 15q_2 \\ \text{s.t.} \quad & 4q_1 + 2q_2 \leq 200 \\ & q_1 + 3q_2 \leq 300 \\ & q_1 \geq 0 \\ & q_2 \geq 0. \end{aligned} \tag{7.76}$$

In reality, we can't make fractional bottles of wine, so we can round q_1 and q_2 to the nearest integer if needed. Now consider the following modification. Suppose that we actually want to sell off some of the grapes instead of turning them into wine. We can earn λ_1 dollars per kilo of merlot and λ_2 per kilo of shiraz. This yields a new optimization problem

$$\max_{q_1, q_2 \in \mathbb{R}_+} \{20q_1 + 15q_2 + \lambda_1(200 - 4q_1 - 2q_2) + \lambda_2(300 - q_1 - 3q_2)\} \tag{7.77}$$

$$= \max_{q_1, q_2 \in \mathbb{R}_+} \{(20 - 4\lambda_1 - \lambda_2)q_1 + (15 - 2\lambda_1 - 3\lambda_2)q_2 + 200\lambda_1 + 300\lambda_2\}. \tag{7.78}$$

If the coefficient of q_1 is negative, we shouldn't make any of the first blend. If the coefficient of q_2 is negative, we shouldn't make any of the second blend. If both are positive, we should make as many of each as possible. What happens when $20 - 4\lambda_1 - \lambda_2 = 0$ and $15 - 2\lambda_1 - 3\lambda_2 = 0$? This is an indifference point, i.e. the point at which our profit is the same no matter how many bottles of either wine we make. Under this condition, the minimum profit we could possibly make is

$$\min_{\lambda_1, \lambda_2 \in \mathbb{R}_+} \quad 200\lambda_1 + 300\lambda_2 \tag{7.79}$$

$$\begin{aligned} \text{s.t.} \quad & 20 - 4\lambda_1 - \lambda_2 = 0 \\ & 15 - 2\lambda_1 - 3\lambda_2 = 0. \end{aligned}$$

This problem is the *dual* to our original (primal) problem. The λ_i's are called *shadow prices*, and capture how much we're willing to pay to violate our constraints.

7.4 Karush-Kuhn-Tucker (KKT) Conditions

We now go into some more involved theory which connects strong duality to optimality conditions.

A broadly applicable set of conditions which are sometimes necessary and sometimes sufficient for optimality is called the *Karush-Kuhn-Tucker* (KKT) conditions.

Definition 170 (KKT Conditions)

Let $(\vec{\widetilde{x}}, \vec{\widetilde{\lambda}}, \vec{\widetilde{\nu}}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ be a decision variable and Lagrange multipliers. Suppose that the objective function f_0 and constraint functions $f_1, \ldots, f_m, h_1, \ldots, h_p$ are differentiable. We say that $(\vec{\widetilde{x}}, \vec{\widetilde{\lambda}}, \vec{\widetilde{\nu}})$ fulfills the *KKT conditions* if

- (Primal feasibility.) $\vec{\widetilde{x}}$ is feasible for $\mathcal{P}$, i.e.,

$$f_i(\vec{\widetilde{x}}) \leq 0, \qquad \forall i \in \{1, \ldots, m\} \tag{7.80}$$

$$\text{and} \quad h_j(\vec{\widetilde{x}}) = 0, \qquad \forall j \in \{1, \ldots, p\}. \tag{7.81}$$

- (Dual feasibility.) $(\vec{\widetilde{\lambda}}, \vec{\widetilde{\nu}})$ is feasible for $\mathcal{D}$, i.e.,

$$\widetilde{\lambda}_i \geq 0, \qquad \forall i \in \{1, \ldots, m\}. \tag{7.82}$$

- (Complementary slackness.)

$$\widetilde{\lambda}_i f_i(\vec{\widetilde{x}}) = 0, \qquad \forall i \in \{1, \ldots, m\}. \tag{7.83}$$

- (Stationarity, or first-order condition.)

$$\vec{0} = \nabla_{\vec{x}} L(\vec{\widetilde{x}}, \vec{\widetilde{\lambda}}, \vec{\widetilde{\nu}}) \tag{7.84}$$

$$= \nabla f(\vec{\widetilde{x}}) + \sum_{i=1}^{m} \widetilde{\lambda}_i \nabla f_i(\vec{\widetilde{x}}) + \sum_{j=1}^{p} \widetilde{\nu}_j \nabla h_j(\vec{\widetilde{x}}). \tag{7.85}$$

Given an arbitrary optimization problem, the KKT conditions do *not* have to be related to the optimality conditions. But it turns out that in many cases, they are related.

Theorem 171 (If Strong Duality Holds, then KKT Conditions are Necessary for Optimality)

Suppose $\mathcal{P}$ is a primal problem with dual $\mathcal{D}$. Suppose that the objective function f_0 and constraint functions $f_1, \ldots, f_m, h_1, \ldots, h_p$ are differentiable, strong duality holds, and $(\vec{x}^\star, \vec{\lambda}^\star, \vec{\nu}^\star)$ are optimal primal and dual variables. Then $(\vec{x}^\star, \vec{\lambda}^\star, \vec{\nu}^\star)$ fulfill the KKT conditions.

Proof. By assumption, $\vec{x}^\star$ is feasible for $\mathcal{P}$ and $(\vec{\lambda}^\star, \vec{\nu}^\star)$ is feasible for $\mathcal{D}$. This implies that $\lambda_i^\star f_i(\vec{x}^\star) \leq 0$ for all i, and

$\nu_j^\star h_j(\vec{x}^\star) = 0$ for all j. Thus, we have

$$\begin{aligned} d^\star &= g(\vec{\lambda}^\star, \vec{\nu}^\star) && (7.86)\\ &= \min_{\vec{x}\in\mathbb{R}^n} L(\vec{x}, \vec{\lambda}^\star, \vec{\nu}^\star) && (7.87)\\ &\leq L(\vec{x}^\star, \vec{\lambda}^\star, \vec{\nu}^\star) && (7.88)\\ &= f_0(\vec{x}^\star) + \sum_{i=1}^{m} \underbrace{\lambda_i^\star f_i(\vec{x}^\star)}_{\leq 0} + \sum_{j=1}^{p} \underbrace{\nu_j^\star h_j(\vec{x}^\star)}_{=0} && (7.89)\\ &\leq f_0(\vec{x}^\star) && (7.90)\\ &= p^\star. && (7.91) \end{aligned}$$

Because $p^\star = d^\star$, all inequalities in the above chain are actually equalities. This means that $\vec{x}^\star$ minimizes $L(\vec{x}, \vec{\lambda}^\star, \vec{\nu}^\star)$ over $\vec{x} \in \mathbb{R}^n$. By the main theorem of vector calculus, this implies that $\nabla_{\vec{x}} L(\vec{x}^\star, \vec{\lambda}^\star, \vec{\nu}^\star) = \vec{0}$, which is the stationarity condition. It also implies that $\sum_{i=1}^{m} \lambda_i^\star f_i(\vec{x}^\star) = 0$. But each term in the sum is ≤ 0, so they must all be 0. This means that $\lambda_i^\star f_i(\vec{x}^\star) = 0$ for each i. This gives the complementary slackness condition. Thus the KKT conditions hold for $(\vec{x}^\star, \vec{\lambda}^\star, \vec{\nu}^\star)$. □

Theorem 172 (If Convexity Holds, then KKT Conditions are Sufficient for Optimality)
Suppose $\mathcal{P}$ is a primal problem with dual $\mathcal{D}$. Suppose that the objective function f_0 and constraint functions $f_1, \ldots, f_m, h_1, \ldots, h_p$ of $\mathcal{P}$ are differentiable. Suppose that $f_0, f_1, \ldots, f_m$ are convex and $h_1, \ldots, h_p$ are affine. Suppose that $(\tilde{\vec{x}}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$ fulfill the KKT conditions. Then $\mathcal{P}$ is convex, strong duality holds, and $(\tilde{\vec{x}}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$ are optimal primal and dual variables.

Proof. By assumption, $\tilde{\vec{x}}$ is feasible for $\mathcal{P}$ and $(\tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$ is feasible for $\mathcal{D}$. Since the f_i are convex and the h_j are affine, $\mathcal{P}$ is convex, and

$$L(\vec{x}, \vec{\lambda}, \vec{\nu}) = f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x}) \tag{7.92}$$

is convex in $\vec{x}$. Since stationarity holds, we have

$$\nabla_{\vec{x}} L(\tilde{\vec{x}}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}}) = \vec{0}, \tag{7.93}$$

so because the primal problem is convex, we have that $\tilde{\vec{x}}$ minimizes $L(\vec{x}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$ over all $\vec{x} \in \mathbb{R}^n$. Thus

$$\begin{aligned} g(\tilde{\vec{\lambda}}, \tilde{\vec{\nu}}) &= \min_{\vec{x}\in\mathbb{R}^n} L(\vec{x}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}}) && (7.94)\\ &= L(\tilde{\vec{x}}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}}) && (7.95)\\ &= f_0(\tilde{\vec{x}}) + \sum_{i=1}^{m} \underbrace{\tilde{\lambda}_i f_i(\tilde{\vec{x}})}_{=0} + \sum_{j=1}^{p} \underbrace{\tilde{\nu}_i h_i(\tilde{\vec{x}})}_{=0} && (7.96)\\ &= f_0(\tilde{\vec{x}}). && (7.97) \end{aligned}$$

Thus $f_0(\tilde{\vec{x}}) = g(\tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$, so the duality gap is 0 and strong duality holds, with $(\tilde{\vec{x}}, \tilde{\vec{\lambda}}, \tilde{\vec{\nu}})$ being optimal primal and dual variables. □

In this course, most problems will be convex and strong duality will hold; in this case the above two theorems can apply and we get that the KKT conditions are *equivalent* to optimality conditions. This is the source of the intuition that convex problems are easier to optimize.

Corollary 173 (If Convexity and Strong Duality Hold, then KKT Conditions are Necessary and Sufficient for Optimality)**.** *Suppose $\mathcal{P}$ is a primal problem with dual $\mathcal{D}$. Suppose strong duality holds for $\mathcal{P}$ and $\mathcal{D}$. Suppose that the objective function f_0 and constraint functions $f_1, \ldots, f_m, h_1, \ldots, h_p$ for $\mathcal{P}$ are differentiable. Suppose that $f_0, f_1, \ldots, f_m$ are convex and $h_1, \ldots, h_P$ are affine. Let $(\vec{x}, \vec{\lambda}, \vec{\nu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ be primal and dual variables. Then $(\vec{x}, \vec{\lambda}, \vec{\nu})$ are optimal primal and dual variables if and only if they fulfill the KKT conditions.*

The following is a generic sequence of steps that you can try for any convex optimization problem.

Problem Solving Strategy 174 (Solving Convex Optimization Problems Using KKT Conditions)**.** *Suppose you have a problem $\mathcal{P}$ with dual $\mathcal{D}$.*

1. *Show that $\mathcal{P}$ is convex and the objective and constraint functions are differentiable.*

2. *Show Slater's condition holds and/or that strong duality holds for $\mathcal{P}$ and $\mathcal{D}$.*

3. *Compute the KKT conditions for $\mathcal{P}$ and $\mathcal{D}$.*

4. *Solve for the optimal primal and dual variables using the KKT conditions.*

Even for more complicated problems where it is not possible to solve the KKT conditions analytically, many algorithms can be interpreted as solving the KKT conditions under various conditions.

Example 175 (Example 161, with KKT Conditions)**.** In this example, we apply the KKT conditions to the problem in Example 161. Consider the following problem:

$$p^\star = \min_{x \in \mathbb{R}} \quad 5x^2 \tag{7.98}$$

$$\text{s.t.} \quad 3x \leq 5. \tag{7.99}$$

Its Lagrangian is

$$L(x, \lambda) = 5x^2 + \lambda(3x - 5). \tag{7.100}$$

The problem is convex, and there exists a strictly feasible point in the relative interior of the feasible set, e.g., $x = 0$ (notice that this is also global optimum, but it does not have to be; $x = -1$ would have sufficed just as well to be a strictly feasible point in the relative interior of the feasible set). Thus Slater's condition holds, strong duality holds, and the KKT conditions are necessary and sufficient for optimality. Now we solve the system to global optimum using KKT conditions.

Let $(\widetilde{x}, \widetilde{\lambda})$ solve the KKT conditions. Then they must obey:

- Primal feasibility: $3\widetilde{x} \leq 5$.

- Dual feasibility: $\widetilde{\lambda} \geq 0$.

- Complementary slackness: $\widetilde{\lambda}(3\widetilde{x} - 5) = 0$.

- Stationarity: $0 = \nabla_x L(\widetilde{x}, \widetilde{\lambda}) = 10\widetilde{x} + 3\widetilde{\lambda}$.

Now we solve for $\widetilde{\lambda}$ in terms of $\widetilde{x}$. We obtain from stationarity that

$$\widetilde{\lambda} = -\frac{10}{3}\widetilde{x}. \tag{7.101}$$

Thus by complementary slackness we have

$$0 = -\frac{10}{3}\widetilde{x}(3\widetilde{x} - 5), \tag{7.102}$$

so that

$$\widetilde{x} = 0 \qquad \text{or} \qquad \widetilde{x} = \frac{5}{3}. \tag{7.103}$$

Supposing that $\widetilde{x} = 5/3$, then we have $\widetilde{\lambda} = -5/3$, which violates dual feasibility. Thus we must have $\widetilde{x} = 0$, implying $\widetilde{\lambda} = 0$. And this indeed satisfies all of the KKT conditions. In particular, we can tell from inspection that at least the primal variable $\widetilde{x} = 0 = x^\star$ is globally optimal.

Note that even when solving the KKT system correctly, we (momentarily) ended up with an answer that would not satisfy the KKT conditions in the first place. This is one way to solve such systems: simplify the system as much as possible, generate two or several possible solutions $(\widetilde{x}, \widetilde{\lambda})$, and check again to see which one(s) still make the KKT conditions hold.

The following content is optional/out of scope for this semester. Regardless, it may be helpful to read it to gain context, or get a deeper understanding of various results.

7.5 (OPTIONAL) Conic Duality

Below, we introduce how Lagrange duality can be extended to optimization problems with inequality constraints defined via generalized inequalities, as defined below. For more details, please see [1, Section 5.9].

Definition 176 (Generalized Inequalities using Convex Cones)
Let $K \subseteq \mathbb{R}^n$ be a proper cone. Let $\operatorname{int}(K)$ be the interior of K. The *generalized inequality* induced by K is a relation $\succeq_K$ defined on pairs of vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ with the following properties:

(a) If $\vec{u} - \vec{v} \in K$, we write $\vec{u} \succeq_K \vec{v}$ and $\vec{v} \preceq_K \vec{u}$.

(b) If $\vec{u} - \vec{v} \in \operatorname{int}(K)$, we write $\vec{u} \succ_K \vec{v}$ and $\vec{v} \prec_K \vec{u}$.

Substituting $\vec{0}$ into the above notation, we get the following notations:

(c) If $\vec{u} \in K$, we write $\vec{u} \succeq_K \vec{0}$.

(d) If $-\vec{v} \in K$, we write $\vec{v} \preceq_K \vec{0}$.

(e) If $\vec{u} \in \operatorname{int}(K)$, we write $\vec{u} \succ_K \vec{0}$.

(f) If $-\vec{v} \in \operatorname{int}(K)$, we write $\vec{v} \prec_K \vec{0}$.

It is possible that $\vec{u} - \vec{v} \notin K$, in which case there is no generalized inequality with respect to K between them.

Example 177. While the above may look scary, we have already seen generalized inequalities in this course. The set of positive definite matrices $\mathbb{S}^n_+$ is a proper cone in the set of symmetric matrices $\mathbb{S}^n$. Thus when we write that $A \succeq 0$ for a symmetric matrix A, we are really using the generalized inequality with respect to the cone $K = \mathbb{S}^n_+$. Correspondingly, we can write $A \succ B$ to mean $A - B$ is positive definite.

Generalized inequalities can also help simplify or generalize several concepts introduced earlier in the course. Suppose for example that we have a familiar convex optimization problem:

$$\min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{7.104}$$
$$\text{s.t.} \quad f_i(\vec{x}) \leq 0, \qquad \forall i \in \{1,\dots,m\}. \tag{7.105}$$

(One can also add equality constraints to this problem without issue.) Now consider the set $K = \mathbb{R}^m_+ = \{\vec{x} \in \mathbb{R}^m : x_i \geq 0 \; \forall i \in \{1,\dots,m\}\}$, i.e., the *non-negative orthant*. Indeed, K is a proper cone (proof is an exercise). If we collect all constraints into a single vector-valued constraint function $\vec{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ given by

$$\vec{f}(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix} \tag{7.106}$$

then the constraint $\vec{f}(\vec{x}) \preceq_K \vec{0}$ means that $-\vec{f}(\vec{x}) \in K$, that is, each $-f_i(\vec{x}) \geq 0$, or equivalently $f_i(\vec{x}) \leq 0$. These forms of the m constraints are absolutely equivalent. Thus, our original familiar convex optimization problem is equivalent to the following problem:

$$\min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{7.107}$$
$$\text{s.t.} \quad \vec{f}(\vec{x}) \preceq_K \vec{0}. \tag{7.108}$$

We will work with generalizations of this problem shortly.

Observe that, for any proper cone $K \subseteq \mathbb{R}^n$ and given $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^n$, if we have $\vec{v}_1 \succeq_K \vec{v}_2$ and $\vec{v}_2 \succeq_K \vec{v}_3$, then $\vec{v}_1 \succeq_K \vec{v}_3$.

Now we use this generalized inequality system to define more general convex optimization problems. Let $f_0\colon \mathbb{R}^n \to \mathbb{R}$ be the objective function. For each $i \in \{1,\dots,m\}$, let $\vec{f}_i\colon \mathbb{R}^n \to \mathbb{R}^{d_i}$ be a *vector-valued* inequality constraint function (we will see shortly how this works), and let $K_i \subseteq \mathbb{R}^{d_i}$ be a proper cone (this could be called the *constraint cone*). For convenience, let $d \doteq \sum_{i=1}^m d_i$. Finally, for each $j \in \{1,\dots,p\}$, let $h_j\colon \mathbb{R}^n \to \mathbb{R}$ be a (usual scalar-valued) equality constrained function.

Consider the following primal optimization problem over $\mathbb{R}^n$ with generalized inequality constraints:

$$\text{problem } \mathcal{P}: \qquad p^\star = \min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{7.109}$$
$$\text{s.t.} \quad \vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0}, \qquad \forall i \in \{1,\dots,m\}$$
$$h_j(\vec{x}) = 0, \qquad \forall j \in \{1,\dots,p\}.$$

We aim to derive a theory of generalized Lagrangian duality for this system.

As before, let $\mathbb{1}[\cdot]$ be an indicator function that returns 0 when the input condition is true, and $+\infty$ otherwise. Then, as before, we can look at the unconstrained variant:

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} \left[f_0(\vec{x}) + \sum_{i=1}^m \mathbb{1}\left[\vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0}\right] + \sum_{j=1}^p \mathbb{1}\left[h_j(\vec{x}) = 0\right] \right]. \tag{7.110}$$

Recall that we showed, en route to the Lagrangian duality presented in Section 7.1, that

$$\mathbb{1}[f_i(\vec{x}) \leq 0] = \max_{\lambda_i \in \mathbb{R}_+} \lambda_i f_i(\vec{x}), \qquad \mathbb{1}[h_j(\vec{x}) = 0] = \max_{\nu_j \in \mathbb{R}} \nu_j h_j(\vec{x}). \tag{7.111}$$

The latter equality will help us here again, but the former will not, and we have to derive an analogue. Indeed, we claim that

$$\mathbb{1}\left[\vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0}\right] = \max_{\substack{\vec{\lambda}_i \in \mathbb{R}^{d_i} \\ \vec{\lambda}_i \succeq_{K_i^\star} \vec{0}}} \vec{\lambda}_i^\top \vec{f}_i(\vec{x}), \tag{7.112}$$

where $K_i^\star$ denotes the dual cone of K_i in $\mathbb{R}^{d_i}$. Why is this true? It turns out to be for a similar reason as the more special case above. If $\vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0}$, then $-\vec{f}_i(\vec{x}) \in K_i$. By definition, for $\vec{\lambda}_i \in K_i^\star$, we must have

$$0 \leq \vec{\lambda}_i^\top(-\vec{f}_i(\vec{x})) = -\vec{\lambda}_i^\top \vec{f}_i(\vec{x}) \implies \vec{\lambda}_i^\top \vec{f}_i(\vec{x}) \leq 0. \tag{7.113}$$

Equality in this case is obtained by selecting $\vec{\lambda}_i = \vec{0}$, which we are assured is legal because, by definition of a proper cone, $\vec{0} \in K_i$. Thus we have shown

$$\vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0} \implies \max_{\substack{\vec{\lambda}_i \in \mathbb{R}^{d_i} \\ \vec{\lambda}_i \succeq_{K_i^\star} \vec{0}}} \vec{\lambda}_i^\top \vec{f}_i(\vec{x}) = 0. \tag{7.114}$$

In the other case, suppose $\vec{f}_i(\vec{x}) \not\preceq_{K_i} \vec{0}$. Then we have $-\vec{f}_i(\vec{x}) \notin K_i$. Since K_i is a proper cone, it is closed and convex, so $K_i = (K_i^\star)^\star$. We thus have $-\vec{f}_i(\vec{x}) \notin (K_i^\star)^\star$. Namely, there exists some $\vec{\lambda}_i \in K_i^\star$ such that $\vec{\lambda}_i^\top(-\vec{f}_i(\vec{x})) < 0$, or equivalently $\vec{\lambda}_i^\top \vec{f}_i(\vec{x}) > 0$. Since $K_i^\star$ is a cone, it is closed under positive scalar multiplication; thus, for any $\alpha_i > 0$, we have $\alpha_i \vec{\lambda}_i \in K_i^\star$, so

$$(\alpha_i \vec{\lambda}_i)^\top \vec{f}_i(\vec{x}) = \alpha_i(\vec{\lambda}_i^\top \vec{f}_i(\vec{x})) > 0 \tag{7.115}$$

Taking $\alpha_i \to \infty$ obtains that the maximum over $\vec{\lambda}_i$ is $+\infty$ in this case. Namely, we have shown

$$\vec{f}_i(\vec{x}) \not\preceq_{K_i} \vec{0} \implies \max_{\substack{\vec{\lambda}_i \in \mathbb{R}^{d_i} \\ \vec{\lambda}_i \succeq_{K_i^\star} \vec{0}}} \vec{\lambda}_i^\top \vec{f}_i(\vec{x}) = \infty. \tag{7.116}$$

Thus, defining $\vec{\lambda} = (\vec{\lambda}_1, \ldots, \vec{\lambda}_m) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m} = \mathbb{R}^d$ for convenience, we have

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \max_{\substack{\vec{\lambda} \in \mathbb{R}^d \\ \vec{\nu} \in \mathbb{R}^p \\ \vec{\lambda}_i \preceq_{K_i} \vec{0}\ \forall i \in \{1, \ldots, m\}}} \left[f_0(\vec{x}) + \sum_{i=1}^{m} \vec{\lambda}_i^\top \vec{f}_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x}) \right]. \tag{7.117}$$

The above discussion inspires the following definition of a generalized Lagrangian $L\colon \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R}^p$ for the given primal optimization problem with generalized inequalities.[2]

$$L(\vec{x}, \vec{\lambda}, \vec{\nu}) \doteq f_0(\vec{x}) + \sum_{i=1}^{m} \vec{\lambda}_i^\top \vec{f}_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x}). \tag{7.118}$$

As before, we define the dual function $g\colon \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$ by

$$g(\vec{\lambda}, \vec{\nu}) \doteq \min_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\lambda}, \vec{\nu}). \tag{7.119}$$

Thus, we can write the dual problem of (7.109) as

$$\text{problem } \mathcal{D}\colon \qquad d^\star = \max_{\substack{\vec{\lambda} \in \mathbb{R}^d \\ \vec{\nu} \in \mathbb{R}^p}} \quad g(\vec{\lambda}, \vec{\nu}) \tag{7.120}$$

[2]If the optimization problem under study were over a general inner product space, then the expressions $\vec{\lambda}_i^\top \vec{f}_i(\vec{x})$ should be replaced with $\left\langle \vec{\lambda}_i, \vec{f}_i(\vec{x}) \right\rangle$. This occurs primarily in semidefinite programming, where $\vec{f}_i$ could produce a symmetric matrix as an output. In this case, $\vec{\lambda}_i$ would similarly be a symmetric matrix, and the inner product would be the Frobenius inner product.

$$\text{s.t.} \quad \vec{\lambda}_i \succeq_{K_i^\star} \vec{0}, \qquad \forall i \in \{1, \ldots, m\}.$$

Note that weak duality, i.e., $d^\star \leq p^\star$, always holds, by the minimax inequality.

An analog of Slater's condition holds for optimization problems with generalized inequalities induced by proper cones. We first require the following definition, which extends the notion of the convexity of a function to generalized convexity, where the inequalities in the definition of convexity are turned into general inequalities induced by proper cones.

Definition 178 (Generalized Convexity)

Let $\vec{f} \colon \mathbb{R}^n \to \mathbb{R}^d$, and let $K \subseteq \mathbb{R}^d$ be a proper cone.

(a) We say that $\vec{f}$ is K*-convex* if for any $\alpha \in [0, 1]$ and $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$\vec{f}(\alpha\vec{x} + (1-\alpha)\vec{y}) \preceq_K \alpha\vec{f}(\vec{x}) + (1-\alpha)\vec{f}(\vec{y}). \tag{7.121}$$

(b) We say that $\vec{f}$ is *strictly* K*-convex* if for any $\alpha \in (0, 1)$ and $\vec{x}, \vec{y} \in \mathbb{R}^n$ with $\vec{x} \neq \vec{y}$, we have

$$\vec{f}(\alpha\vec{x} + (1-\alpha)\vec{y}) \prec_K \alpha\vec{f}(\vec{x}) + (1-\alpha)\vec{f}(\vec{y}). \tag{7.122}$$

Theorem 179 (Generalized Slater's Condition)

Consider a primal problem $\mathcal{P}$:

$$\begin{aligned} \text{problem } \mathcal{P}: \qquad p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & f_0(\vec{x}) \\ \text{s.t.} \quad & \vec{f}_i(\vec{x}) \preceq_{K_i} \vec{0}, \qquad \forall i \in \{1, \ldots, m\} \\ & h_j(\vec{x}) = 0, \qquad \forall j \in \{1, \ldots, p\}, \end{aligned} \tag{7.109}$$

where

- the function $f_0 \colon \mathbb{R}^n \to \mathbb{R}$ is convex.;
- for each $i \in \{1, \ldots, m\}$, the function $\vec{f}_i \colon \mathbb{R}^n \to \mathbb{R}^{d_i}$ is K_i-convex, where $K_i \subseteq \mathbb{R}^{d_i}$ is a proper cone;
- for each $j \in \{1, \ldots, p\}$, the function $h_j \colon \mathbb{R}^n \to \mathbb{R}$ is affine.

If there exists any point $\vec{x} \in \operatorname{relint}(\Omega)$ which is strictly feasible, i.e., such that all of the following holds:

- for each $i \in \{1, \ldots, m\}$, we have $\vec{f}_i(\vec{x}) \prec_{K_i} \vec{0}$;
- and for each $j \in \{1, \ldots, p\}$, we have $h_j(\vec{x}) = 0$;

then strong duality holds for $\mathcal{P}$ and its dual $\mathcal{D}$, i.e., the duality gap is 0. [a]

[a] Analogous conditions hold for the general case where the primal optimization problem is defined over a vector space equipped with an inner product and a norm, as in the case of semidefinite programming.

Chapter 8

Types of Optimization Problems

Relevant sections of the textbooks:

- [1] Chapter 4.
- [2] Chapters 9, 10, 12.

8.1 Linear Programs

In this chapter we introduce a *taxonomy* of common optimization problems that can be efficiently solved through a variety of ways. We use the notation $\vec{u} \geq \vec{v}$ to denote $u_i \geq v_i$ for all i.

A linear program is one with an affine objective function and affine constraint functions (both inequality and equality constraints). It is the most restrictive (e.g. smallest) class in the taxonomy we present. It has a *standard form* denoted below.

Definition 180 (Linear Program)

A *linear program* (LP) is an optimization problem with an affine objective and affine constraints. A *standard form linear program* is an optimization problem of the following form:[a]

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & A\vec{x} = \vec{y} \\ & \vec{x} \geq \vec{0}. \end{aligned} \tag{8.1}$$

[a]Constant terms have been omitted from the objective function.

There are many equivalent forms of a linear program; in particular, the following proposition (whose proof we leave as an exercise) can be shown using slack variables.

Proposition 181

Any linear program is equivalent to a standard form linear program.

Putting linear programs in standard form is important because the standard form is commonly accepted by optimization algorithms and implementations. Usually if you provide a linear program that isn't in standard form to a solver, they will convert it to standard form first before running their algorithms. Conversions to-and-from standard form may increase the number of variables in the problem and the eventual algorithmic complexity of solving it. One example of this conversion is done below.

Example 182. Suppose we have the linear program

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^2} \quad & 2x_1 + 4x_2 && (8.2)\\ \text{s.t.} \quad & x_1 + x_2 \geq 3 && (8.3)\\ & 3x_1 + 2x_2 = 14 && (8.4)\\ & x_1 \geq 0. && (8.5) \end{aligned}$$

To convert this to standard form, first we note that

$$x_1 + x_2 \geq 3 \iff -x_1 - x_2 \leq -3 \tag{8.6}$$

which obtains the following system:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^2} \quad & 2x_1 + 4x_2 && (8.7)\\ \text{s.t.} \quad & -x_1 - x_2 \leq -3 && (8.8)\\ & 3x_1 + 2x_2 = 14 && (8.9)\\ & x_1 \geq 0. && (8.10) \end{aligned}$$

Adding a slack variable $x_3 \geq 0$ to the first constraint yields equality:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^3} \quad & 2x_1 + 4x_2 && (8.11)\\ \text{s.t.} \quad & -x_1 - x_2 + x_3 = -3 && (8.12)\\ & 3x_1 + 2x_2 = 14 && (8.13)\\ & x_1 \geq 0 && (8.14)\\ & x_3 \geq 0. && (8.15) \end{aligned}$$

The only reason this is not in standard form is that x_2 is unconstrained. We can represent any real number as the difference of non-negative numbers; one such construction is $x_2 = x_2^+ - x_2^-$ where $x_2^+ = \max\{x_2, 0\}$ and $x_2^- = -\min\{x_2, 0\}$, but there are many others. Thus we can replace x_2 by $x_4 - x_5$ and add the constraints that $x_4 \geq 0$ and $x_5 \geq 0$, obtaining the problem

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^5} \quad & 2x_1 + 4x_4 - 4x_5 && (8.16)\\ \text{s.t.} \quad & -x_1 + x_3 - x_4 + x_5 = -3 && (8.17)\\ & 3x_1 + 2x_4 - 2x_5 = 14 && (8.18)\\ & x_1 \geq 0 && (8.19)\\ & x_3 \geq 0 && (8.20)\\ & x_4 \geq 0 && (8.21)\\ & x_5 \geq 0. && (8.22) \end{aligned}$$

Finally, we notice that x_2 is neither in the objective nor the constraints and so can be eliminated.

$$\begin{aligned}
\min_{\vec{x}\in\mathbb{R}^4} \quad & 2x_1 + 4x_3 - 4x_4 && (8.23)\\
\text{s.t.} \quad & -x_1 + x_2 - x_3 + x_4 = -3 && (8.24)\\
& 3x_1 + 2x_3 - 2x_4 = 14 && (8.25)\\
& x_1 \geq 0 && (8.26)\\
& x_2 \geq 0 && (8.27)\\
& x_3 \geq 0 && (8.28)\\
& x_4 \geq 0. && (8.29)
\end{aligned}$$

Linear programs are convex problems, and in particular the feasible set of a linear program is a convex set.

Proposition 183

Any linear program is a convex optimization problem.

Proposition 184

The dual of the standard form linear program

$$\begin{aligned}
p^\star = \min_{\vec{x}\in\mathbb{R}^n} \quad & \vec{c}^\top \vec{x} && (8.30)\\
\text{s.t.} \quad & A\vec{x} = \vec{y}\\
& \vec{x} \geq \vec{0}
\end{aligned}$$

is

$$\begin{aligned}
d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}^n\\ \vec{\nu}\in\mathbb{R}^m}} \quad & -\vec{y}^\top \vec{\nu} && (8.31)\\
\text{s.t.} \quad & \vec{c} - \vec{\lambda} + A^\top \vec{\nu} = \vec{0}\\
& \vec{\lambda} \geq \vec{0}.
\end{aligned}$$

Proof. The Lagrangian is

$$\begin{aligned}
L(\vec{x}, \vec{\lambda}, \vec{\nu}) &= \vec{c}^\top \vec{x} - \vec{\lambda}^\top \vec{x} + \vec{\nu}^\top (A\vec{x} - \vec{y}) && (8.32)\\
&= (\vec{c} - \vec{\lambda} + A^\top \vec{\nu})^\top \vec{x} - \vec{\nu}^\top \vec{y}. && (8.33)
\end{aligned}$$

The dual function is

$$\begin{aligned}
g(\vec{\lambda}, \vec{\nu}) &= \min_{\vec{x}\in\mathbb{R}^n} L(\vec{x}, \vec{\lambda}, \vec{\nu}) && (8.34)\\
&= \min_{\vec{x}\in\mathbb{R}^n} \left\{ (\vec{c} - \vec{\lambda} + A^\top \vec{\nu})^\top \vec{x} - \vec{\nu}^\top \vec{y} \right\} && (8.35)\\
&= \min_{\vec{x}\in\mathbb{R}^n} \left\{ (\vec{c} - \vec{\lambda} + A^\top \vec{\nu})^\top \vec{x} \right\} - \vec{\nu}^\top \vec{y} && (8.36)\\
&= \begin{cases} -\vec{\nu}^\top \vec{y}, & \text{if } \vec{c} - \vec{\lambda} + A^\top \vec{\nu} = \vec{0} \\ -\infty, & \text{otherwise.} \end{cases} && (8.37)
\end{aligned}$$

The dual problem is

$$d^\star = \max_{\substack{\vec{\lambda} \in \mathbb{R}^n \\ \vec{\nu} \in \mathbb{R}^m}} \quad g(\vec{\lambda}, \vec{\nu}) \tag{8.38}$$

$$\text{s.t.} \quad \vec{\lambda} \geq \vec{0}. \tag{8.39}$$

Expanding this out and adding the domain constraint, we get

$$d^\star = \max_{\substack{\vec{\lambda} \in \mathbb{R}^n \\ \vec{\nu} \in \mathbb{R}^m}} \quad -\vec{\nu}^\top \vec{y} \tag{8.40}$$

$$\text{s.t.} \quad \vec{c} - \vec{\lambda} + A^\top \vec{\nu} = \vec{0} \tag{8.41}$$

$$\vec{\lambda} \geq \vec{0} \tag{8.42}$$

as desired. □

One very relevant question is how to solve linear programs efficiently. There are several methods available to us — for example, any constrained convex optimization solver, such as projected gradient descent, will solve the problem, given an appropriate learning rate and efficient projection method. The algorithm most used in practice is the interior point method; we will learn the basics of interior point methods later in this course.[1] But there is one algorithm which was used for many years in practice, that truly exploits the structure of linear programs to efficiently solve them. It is called the *simplex algorithm*, which was invented by George Dantzig in 1947.

The core idea behind the the simplex algorithm is that *at least one optimal point of a linear program is a "vertex" of its feasible set, so long as this feasible set is bounded*. We need to prove this idea, so far stated informally, and to do this we will first need to characterize the feasible set of a linear program.

Definition 185 (Polyhedron, Polygon)
A *polyhedron* is an intersection of a finite number of half-spaces. A *polygon* is a bounded polyhedron.

From the definition of a linear program, we see that its feasible region must be a polyhedron. For the standard form linear program, we can check explicitly that the feasible set is the intersection of the three classes of half-spaces $\{\vec{x} \in \mathbb{R}^n \mid \vec{a}_i^\top \vec{x} \leq y_i\}$, $\{\vec{x} \in \mathbb{R}^n \mid \vec{a}_i^\top \vec{x} \geq y_i\}$, and $\{\vec{x} \in \mathbb{R}^n \mid x_j \geq 0\}$, where $\vec{a}_i^\top$ are the rows of A, y_i are the entries of $\vec{y}$, and x_j are the entries of $\vec{x}$. This feasible region can be unbounded or bounded; if it is bounded, it will be a polygon.

Definition 186 (Extreme Point, Vertex)
Let $K \subseteq \mathbb{R}^n$ be a set. We say that $\vec{x} \in K$ is an *extreme point* of K if there *does not* exist $\vec{y}, \vec{z} \in K \setminus \{\vec{x}\}$ and $\theta \in [0, 1]$ such that $\vec{x} = \theta\vec{y} + (1 - \theta)\vec{z}$.

An extreme point of a polyhedron is called a *vertex* of the polyhedron.

If the feasible set of a linear program is an unbounded polyhedron, then there are examples where the optimal value is not achieved at a vertex, as demonstrated in the following example.

Example 187. Consider the following linear program over variables $\vec{x} \in \mathbb{R}^2$:

$$\min_{x_1, x_2 \in \mathbb{R}} \quad \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8.43}$$

[1] Talk to your TA Tarun if you want to learn more about interior point methods!

$$\begin{aligned} \text{s.t.} \quad & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 = 1. \end{aligned}$$

Define $\vec{x}_\alpha \doteq \begin{bmatrix} 1 \\ \alpha \end{bmatrix}$. One can check that $\vec{x}_\alpha$ is feasible for all $\alpha \geq 0$, and that the objective value at $\vec{x}_\alpha$ is $-\alpha$. By sending $\alpha \to \infty$, we get $p^\star = -\infty$, and the optimum is not achieved at a vertex (or indeed by any $\vec{x} \in \mathbb{R}^2$).

With this understanding, we now seek to prove the main idea we had earlier. There are several proofs, but the cleanest uses the following intuitive fact:

Proposition 188
A polygon has finitely many vertices and is the convex hull of its vertices.

Unfortunately, the proof is (surprisingly?) quite complicated, so we omit it. A complete proof is in **ziegler2012lectures**, for example.

Theorem 189 (Main Theorem of Linear Programming)
Consider the standard form linear program:

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & A\vec{x} = \vec{y} \\ & \vec{x} \geq \vec{0}. \end{aligned} \tag{8.44}$$

Suppose that the feasible set $\Omega \doteq \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{y}, \vec{x} \geq \vec{0}\}$ is bounded. Then the optimal value is achieved at a vertex.

Namely, one can find an optimal point which is a vertex. There may be optimal points that are not vertices. The simplest example is to set the objective as $\vec{c} = \vec{0}$, so every feasible point is optimal with objective value 0, but there are other examples which are a bit more complicated to set up.

Proof. Since the feasible set Ω is a bounded polyhedron, it is a polygon, and so it is the convex hull of its vertices, say $\vec{v}_1, \ldots, \vec{v}_k$. Thus any $\vec{x} \in \Omega$ can be written as a convex combination of the vertices $\vec{v}_i$, namely,

$$\vec{x} = \sum_{i=1}^{k} \alpha_i \vec{v}_i \tag{8.45}$$

where each $\alpha_i \geq 0$ and $\sum_{i=1}^{k} \alpha_i = 1$. It then follows that

$$p^\star = \min_{\alpha_1, \ldots, \alpha_k \in \mathbb{R}} \quad \sum_{i=1}^{k} \alpha_i (\vec{c}^\top \vec{v}_i) \tag{8.46}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \qquad \forall i \in \{1, \ldots, k\} \tag{8.47}$$

$$\sum_{i=1}^{k} \alpha_i = 1. \tag{8.48}$$

Now we have

$$\sum_{i=1}^{k} \alpha_i(\vec{c}^\top \vec{v}_i) \geq \sum_{i=1}^{k} \alpha_i \left(\min_{j \in \{1,\dots,k\}} \vec{c}^\top \vec{v}_j \right) \tag{8.49}$$

$$= \left(\min_{j \in \{1,\dots,k\}} \vec{c}^\top \vec{v}_j \right) \underbrace{\left(\sum_{i=1}^{k} \alpha_i \right)}_{=1} \tag{8.50}$$

$$= \min_{j \in \{1,\dots,k\}} \vec{c}^\top \vec{v}_j. \tag{8.51}$$

Let $i^\star \in \{1, \dots, k\}$ be an index such that $\vec{c}^\top \vec{v}_{i^\star}$ achieves the above minimum, i.e., $\vec{c}^\top \vec{v}_{i^\star} = \min_{j \in \{1,\dots,k\}} \vec{c}^\top \vec{v}_j$. Then the above lower bound is achieved when $\alpha_{i^\star} = 1$ and $\alpha_i = 0$ for $i \neq i^\star$, for example. Thus $\vec{v}_{i^\star}$ is an optimal point for the original linear program, concluding the proof. □

This theorem says is that to solve a linear program, we only need to check the vertices of the constraint polyhedron. This reduces an optimization problem over $\mathbb{R}^n$ to an optimization problem over the finite set of vertices. This reduction motivates a "greedy-like heuristic" solver for linear programs with bounded feasible sets Ω, which is called the *simplex method*. The simplex method is the following procedure:

- Start at a vertex $\vec{v}$ of Ω.
- While there is a neighboring vertex $\vec{w}$ with $\vec{c}^\top \vec{v} > \vec{c}^\top \vec{w}$, move to it, i.e., set $\vec{v} \leftarrow \vec{w}$.
- When there are no neighboring vertices with better optima, stop and return $\vec{v}$.

There are (rather more technical) modifications one can make to this algorithm to solve linear programs with unbounded feasible sets. But the main idea is just the same as gradient descent: iteratively search locally for another point with better objective value, and move to it.

8.2 Quadratic Programs

Definition 190 (Quadratic Program)
A *quadratic program* (QP) is an optimization problem with a quadratic objective and affine constraints. A *standard form quadratic program* is an optimization problem of the following form:

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top H \vec{x} + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & A\vec{x} \leq \vec{y} \\ & C\vec{x} = \vec{z}, \end{aligned} \tag{8.52}$$

where $H \in \mathbb{S}^n$.

In the standard form, we do not lose any generality by enforcing $H \in \mathbb{S}^n$. In particular, for any H we have

$$\frac{1}{2}\vec{x}^\top H \vec{x} + \vec{c}^\top \vec{x} = \frac{1}{2}\vec{x}^\top \left(\frac{H + H^\top}{2} \right) \vec{x} + \vec{c}^\top \vec{x} \tag{8.53}$$

whence the matrix $\frac{H+H^\top}{2}$ (i.e., the *symmetric part* of H) is always symmetric. So if we have a non-symmetric H we can just replace it with its symmetric part, and thus obtain a standard form quadratic program.

Quadratic programs may or may not be convex.

Proposition 191

Consider the following standard form quadratic program:

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & A\vec{x} \leq \vec{y} \\ & C\vec{x} = \vec{z}, \end{aligned} \tag{8.54}$$

where $H \in \mathbb{S}^n$. The following are equivalent:

(a) The problem is convex.

(b) $H \in \mathbb{S}^n_+$.

Proof. The gradient and Hessian of the objective are

$$\nabla\left(\frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x}\right) = \frac{1}{2}(H + H^\top)\vec{x} + \vec{c} = H\vec{x} + \vec{c} \tag{8.55}$$

$$\nabla^2\left(\frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x}\right) = H \tag{8.56}$$

so the objective function is convex if and only if $H \in \mathbb{S}^n_+$. Thus the problem is a convex problem if and only if $H \in \mathbb{S}^n_+$ and the constraint set is convex. But the constraint set is always convex because it is defined by a set of linear equations and inequalities. Thus the problem is convex if and only if $H \in \mathbb{S}^n_+$. □

Example 192. Let us consider the unconstrained quadratic program

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n}\left(\frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x}\right) \tag{8.57}$$

where $H \in \mathbb{S}^n$, and solve it in a variety of cases, namely that where $H \notin \mathbb{S}^n_+$, $H \in \mathbb{S}^n_+$ with $\vec{c} \in \mathcal{N}(H) \setminus \{\vec{0}\}$, and $H \in \mathbb{S}^n_+$ with $\vec{c} \in \mathcal{N}(H)^\perp = \mathcal{R}(H^\top) = \mathcal{R}(H)$.

Case 1. Suppose that $H \notin \mathbb{S}^n_+$. Then H has a negative eigenvalue λ; let $\vec{v}$ be any corresponding unit eigenvector. Then, if we choose $\vec{x}_t = t \cdot \vec{v}$, we get

$$p^\star \leq \lim_{t \to \infty}\left(\frac{1}{2}\vec{x}_t^\top H\vec{x}_t + \vec{c}^\top \vec{x}_t\right) \tag{8.58}$$

$$= \lim_{t \to \infty}\left(\frac{1}{2}t^2\vec{v}^\top H\vec{v} + t\vec{c}^\top \vec{v}\right). \tag{8.59}$$

To bound the terms inside the limit, we compute

$$\vec{v}^\top H\vec{v} = \vec{v}^\top(\lambda\vec{v}) = \lambda\vec{v}^\top\vec{v} = \lambda \tag{8.60}$$

$$\vec{c}^\top \vec{v} \leq \|\vec{c}\|_2 \|\vec{v}\|_2 = \|\vec{c}\|_2. \tag{8.61}$$

Thus we have

$$p^\star \leq \lim_{t \to \infty}\left(\frac{1}{2}\lambda t^2 + t\|\vec{c}\|_2\right). \tag{8.62}$$

Since $\lambda < 0$, the term inside the limit is a concave (i.e., downward facing) quadratic function of t, and so its limit as $t \to \pm\infty$ is $-\infty$. Thus $p^\star \leq -\infty$ so $p^\star = -\infty$.

Case 2. Suppose that $H \in \mathbb{S}_+^n$, and suppose that $\vec{c} \in \mathcal{N}(H) \setminus \{0\}$. Then H has (at least) an eigenvalue λ equal to 0, and in particular, by the spectral theorem $\vec{c}$ can be written as a linear combination of eigenvectors of H with eigenvalue 0. Let $\vec{v}$ be any unit eigenvector with eigenvalue 0 such that $\vec{c}^\top \vec{v} \neq 0$. Let $\vec{x}_t = -t \cdot \operatorname{sgn}\left(\vec{c}^\top \vec{v}\right) \cdot \vec{v}$. Then

$$\begin{aligned}
p^\star &\leq \lim_{t\to\infty} \left(\frac{1}{2}\vec{x}_t^\top H \vec{x}_t + \vec{c}^\top \vec{x}_t\right) && (8.63)\\
&= \lim_{t\to\infty} \left(\frac{1}{2}t^2 \vec{v}^\top H \vec{v} - t \cdot \operatorname{sgn}\left(\vec{c}^\top \vec{v}\right) \cdot \vec{c}^\top \vec{v}\right) && (8.64)\\
&= \lim_{t\to\infty} \left(\frac{1}{2}t^2 \vec{v}^\top \vec{0} - t \cdot \left|\vec{c}^\top \vec{v}\right|\right) && (8.65)\\
&= \lim_{t\to\infty} \left(-t \cdot \left|\vec{c}^\top \vec{v}\right|\right) && (8.66)\\
&= \lim_{t\to\infty} t \cdot \underbrace{\left(-\left|\vec{c}^\top \vec{v}\right|\right)}_{<0} && (8.67)\\
&= -\infty. && (8.68)
\end{aligned}$$

Thus $p^\star \leq -\infty$ so $p^\star = -\infty$.

Case 3. Suppose that $H \in \mathbb{S}_+^n$ with $\vec{c} \in \mathcal{R}(H)$. Then there is nonzero $\vec{x}_0$ such that $\vec{c} = -H\vec{x}_0$. Rewriting the objective, we obtain

$$\begin{aligned}
\frac{1}{2}\vec{x}^\top H \vec{x} + \vec{c}^\top \vec{x} &= \frac{1}{2}\vec{x}^\top H \vec{x} - \vec{x}_0^\top H \vec{x} && (8.69)\\
&= \frac{1}{2}\vec{x}^\top H \vec{x} - \vec{x}_0^\top H \vec{x} + \frac{1}{2}\vec{x}_0^\top H \vec{x}_0 - \frac{1}{2}\vec{x}_0^\top H \vec{x}_0 && (8.70)\\
&= \frac{1}{2}\left(\vec{x}^\top H \vec{x} - 2\vec{x}_0^\top H \vec{x} + \vec{x}_0^\top H \vec{x}_0\right) - \frac{1}{2}\vec{x}_0^\top H \vec{x}_0 && (8.71)\\
&= \frac{1}{2}(\vec{x} - \vec{x}_0)^\top H (\vec{x} - \vec{x}_0) - \frac{1}{2}\vec{x}_0^\top H \vec{x}_0. && (8.72)
\end{aligned}$$

Since $H \in \mathbb{S}_+^n$, the minimizer is at any $\vec{x}$ such that $\vec{x} - \vec{x}_0 \in \mathcal{N}(H)$. One can write this as $\vec{x} \in \vec{x}_0 + \mathcal{N}(H)$. A particular solution in terms of problem parameters is $\vec{x} = -H^\dagger \vec{c}$ where $H^\dagger$ is the Moore-Penrose pseudoinverse of H. Recall that we discussed the Moore-Penrose pseudoinverse in more generality in homework where we derived the solution to the least-norm least-squares problem, but one can show that if $H = U\Lambda U^\top$ then $H^\dagger = U\Lambda^\dagger U^\top$ where $\Lambda^\dagger$ is the diagonal matrix whose entries are $\Lambda_{ii}^\dagger = \begin{cases} 1/\Lambda_{ii}, & \text{if } \Lambda_{ii} \neq 0 \\ 0, & \text{if } \Lambda_{ii} = 0 \end{cases}$.

The previous example shows that we can solve unconstrained quadratic programs directly and read off the solutions. It turns out that one can transform any quadratic program with equality constraints into an unconstrained quadratic program. So really, this analysis encapsulates a huge class of quadratic programs.

Computing the dual of a quadratic program has a similar number of cases; it is an exercise which is left to homework.

Example 193 (Linear-Quadratic Regulator)**.** Suppose we have a discrete-time dynamical system, of the form

$$\begin{aligned}
\vec{x}_{t+1} &= A\vec{x}_t + B\vec{u}_t, \qquad \forall t \geq 0 && (8.73)\\
\vec{x}_0 &= \vec{\xi}. && (8.74)
\end{aligned}$$

Then one can show that

$$\vec{x}_t = A^t \vec{\xi} + \sum_{k=0}^{t-1} A^{t-k-1} B \vec{u}_k. \tag{8.75}$$

For a fixed terminal time T, we want to reach goal state $\vec{g}$. Namely, we want to solve the problem

$$\min_{\substack{\vec{x}_0,\dots,\vec{x}_T \\ \vec{u}_0,\dots,\vec{u}_{T-1}}} \quad \|\vec{x}_T - \vec{g}\|_2^2 + \sum_{k=0}^{T-1} \|\vec{u}_k\|_2^2 \tag{8.76}$$

$$\text{s.t.} \quad \vec{x}_{t+1} = A\vec{x}_t + B\vec{u}_t, \qquad \forall t \in \{0, 1, \dots, T-1\} \tag{8.77}$$

$$\vec{x}_0 = \vec{\xi}. \tag{8.78}$$

This is a quadratic program since the objective function is a quadratic function of each $\vec{x}_t$, and the constraints are affine equations relating the $\vec{x}_t$ and the $\vec{u}_t$.

As a last note, problems with quadratic objectives and quadratic inequality constraints are called *quadratically constrained quadratic programs* (QCQPs). Like quadratic programs, QCQPs can be convex or non-convex.

8.3 Quadratically-Constrained Quadratic Programs

Definition 194 (Quadratically-Constrained Quadratic Program)
A *quadratically-constrained quadratic program* (QCQP) is an optimization problem with a quadratic objective and quadratic constraints. A *standard form quadratically constrained quadratic program* is an optimization problem of the following form:

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \frac{1}{2}\vec{x}^\top P_i\vec{x} + \vec{b}_i^\top \vec{x} + c_i \leq 0, \qquad \forall i \in \{1, \dots, m\} \\ & \frac{1}{2}\vec{x}^\top Q_i\vec{x} + \vec{d}_i^\top \vec{x} + f_i = 0, \qquad \forall i \in \{1, \dots, p\}, \end{aligned} \tag{8.79}$$

where $H, P_1, \dots, P_m, Q_1, \dots, Q_p \in \mathbb{S}^n$.

Proposition 195
Consider the following standard form quadratically-constrained quadratic program:

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top H\vec{x} + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \frac{1}{2}\vec{x}^\top P_i\vec{x} + \vec{b}_i^\top \vec{x} + c_i \leq 0, \qquad \forall i \in \{1, \dots, m\} \\ & \frac{1}{2}\vec{x}^\top Q_i\vec{x} + \vec{d}_i^\top \vec{x} + f_i = 0, \qquad \forall i \in \{1, \dots, p\}, \end{aligned} \tag{8.80}$$

where $H, P_1, \dots, P_m, Q_1, \dots, Q_p \in \mathbb{S}^n$. If $H, P_1, \dots, P_m \in \mathbb{S}^n_+$ and $Q_1 = \cdots = Q_p = 0$, then the problem is convex.

Proof. Left as exercise. □

8.4 Second-Order Cone Programs

There is one more broad class of problems that we consider in this course, called *second-order cone programs* (SOCPs). They are among the broadest class of problems that we can efficiently solve using algorithms such as the interior point method, which we may discuss later in the course.

Definition 196 (Second-Order Cone Program)

A *second-order cone program* is an optimization problem with a linear objective and affine and "second-order cone constraints", i.e., constraints which say that an affine function of $\vec{x}$ is contained in the second-order cone. A *standard form second-order cone program* is an optimization problem of the following form:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.81}$$
$$\text{s.t.} \quad \|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.82}$$

Second order cone constraints are strictly more broad than affine constraints; to encode an affine constraint $A_i \vec{x} = \vec{y}_i$ as a second-order cone constraint, pick the corresponding $\vec{b}_i = \vec{0}$ and $z_i = 0$. This makes the constraint $\|A_i \vec{x} - \vec{y}_i\|_2 \leq 0$ or equivalently $A_i \vec{x} = \vec{y}_i$.

Proposition 197

Second-order cone problems are convex optimization problems.

Proof. Each second-order cone constraint $\|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i$ can be alternatively formulated as constraining the tuple $(A_i \vec{x} - \vec{y}_i, \vec{b}_i^\top \vec{x} + z_i) \in \mathbb{R}^{n+1}$ to lie within the second-order cone in $\mathbb{R}^{n+1}$. But this tuple is an affine transformation of $\vec{x}$, in particular

$$\begin{bmatrix} A_i \vec{x} - \vec{y}_i \\ \vec{b}_i^\top \vec{x} + z_i \end{bmatrix} = \begin{bmatrix} A_i \\ \vec{b}_i^\top \end{bmatrix} \vec{x} + \begin{bmatrix} -\vec{y}_i \\ z_i \end{bmatrix}. \tag{8.83}$$

Since the second order cone is convex and the tuple is an affine transformation of $\vec{x}$, it follows that $\{\vec{x} \in \mathbb{R}^n \mid \|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i\}$ is a convex set. Thus the feasible set is convex (as the intersection of convex sets). The objective function is linear in $\vec{x}$, so the second-order cone problem is convex. □

Example 198. Consider the following problem:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \sum_{i=1}^{m} \|A_i \vec{x} - \vec{y}_i\|_2 . \tag{8.84}$$

One can formulate this as a second-order cone program by using slack variables:

$$p^\star = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ \vec{s} \in \mathbb{R}^m}} \quad \sum_{i=1}^{m} s_i \tag{8.85}$$
$$\text{s.t.} \quad \|A_i \vec{x} - \vec{y}_i\|_2 \leq s_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.86}$$

The following problem is also very related:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \max_{i \in \{1, \ldots, m\}} \|A_i \vec{x} - \vec{y}_i\|_2 . \tag{8.87}$$

We can use a similar slack variable reformulation to formulate this problem as a second order cone program.

$$p^\star = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ s \in \mathbb{R}}} \quad s \tag{8.88}$$
$$\text{s.t.} \quad \|A_i \vec{x} - \vec{y}_i\|_2 \leq s, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.89}$$

These problems can be formulated in terms of route planning – more specifically, finding the route which minimizes the total length between waypoints (in the first problem), or the route which minimizes the maximum length between waypoints (in the second problem).

Example 199 (LPs, QPs, and QCQPs as SOCPs)**.** One can see how LPs are QPs and how QPs are QCQPs, because in each transition the set of properties becomes more permissive — first a linear objective can become a quadratic objective, then linear constraints can become quadratic constraints. It is less clear how LPs, QPs, and QCQPs are SOCPs. In this example we derive a way to write QCQPs as SOCPs, which is also applicable to LPs and QPs (since LPs and QPs are QCQPs).

Consider a QCQP of the form

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top H \vec{x} + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \frac{1}{2}\vec{x}^\top P_i \vec{x} + \vec{b}_i^\top \vec{x} + c_i \leq 0, \qquad \forall i \in \{1, \dots, m\} \\ & C\vec{x} = \vec{z}, \end{aligned} \tag{8.90}$$

where $H, P_1, \dots, P_m \in \mathbb{S}^n_+$. We use the epigraph reformulation to obtain

$$\begin{aligned} p^\star = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ t \in \mathbb{R}}} \quad & t + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \frac{1}{2}\vec{x}^\top P_i \vec{x} + \vec{b}_i^\top \vec{x} + c_i \leq 0, \qquad \forall i \in \{1, \dots, m\} \\ & \frac{1}{2}\vec{x}^\top H \vec{x} \leq t \\ & C\vec{x} = \vec{z}. \end{aligned} \tag{8.91}$$

Let us try to convert the quadratic constraint

$$\frac{1}{2}\vec{x}^\top P_i \vec{x} + \vec{b}_i^\top \vec{x} + c_i \leq 0 \tag{8.92}$$

into a second-order cone constraint. Notice that this constraint is equivalent to

$$\vec{x}^\top P_i \vec{x} + 2(\vec{b}_i^\top \vec{x} + c_i) \leq 0. \tag{8.93}$$

In order to write each term as a square, we first write $\vec{x}^\top P_i \vec{x} = \left\| P_i^{1/2} \vec{x} \right\|_2^2$. We now use a difference of squares identity:

$$(u + v)^2 - (u - v)^2 = 4uv. \tag{8.94}$$

To apply this to the above formula, we plug in $u = \frac{1}{2}$ and $v = \vec{b}_i^\top \vec{x} + c_i$, to get

$$2(\vec{b}_i^\top \vec{x} + c_i) = \left(\frac{1}{2} + \vec{b}_i^\top \vec{x} + c_i\right)^2 - \left(\frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i\right)^2. \tag{8.95}$$

This gives us the constraint

$$\left\| P_i^{1/2} \vec{x} \right\|_2^2 + \left(\frac{1}{2} + \vec{b}_i^\top \vec{x} + c_i\right)^2 - \left(\frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i\right)^2 \leq 0, \tag{8.96}$$

which is equivalent to

$$\left\| P_i^{1/2} \vec{x} \right\|_2^2 + \left(\frac{1}{2} + \vec{b}_i^\top \vec{x} + c_i\right)^2 \leq \left(\frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i\right)^2. \tag{8.97}$$

Now we would like to take square roots and write things in terms of the ℓ^2-norm. For this, we need to show that $\frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i \geq 0$. This follows because, since P_i is positive semidefinite, we have $\frac{1}{2} \geq 0 \geq -\frac{1}{2}\vec{x}^\top P_i \vec{x}$, and so

$$\frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i \geq -\frac{1}{2}\vec{x}^\top P_i \vec{x} - \vec{b}_i^\top \vec{x} - c_i \geq 0. \tag{8.98}$$

Now, taking square roots and writing things in terms of the ℓ^2-norm, we have

$$\left\|\begin{bmatrix} P_i^{1/2}\vec{x} \\ \frac{1}{2} + \vec{b}_i^\top \vec{x} + c_i \end{bmatrix}\right\|_2 \leq \frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i. \tag{8.99}$$

Thus we can write the QCQP as

$$\begin{aligned} p^\star = \min_{\substack{\vec{x}\in\mathbb{R}^n \\ t\in\mathbb{R}}} \quad & t + \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \left\|\begin{bmatrix} P_i^{1/2}\vec{x} \\ \frac{1}{2} + \vec{b}_i^\top \vec{x} + c_i \end{bmatrix}\right\|_2 \leq \frac{1}{2} - \vec{b}_i^\top \vec{x} - c_i, \qquad \forall i \in \{1, \dots, m\} \\ & \left\|\begin{bmatrix} H^{1/2}\vec{x} \\ \frac{1}{2} - t \end{bmatrix}\right\|_2 \leq \frac{1}{2} + t \\ & \|C\vec{x} - \vec{z}\|_2 \leq 0. \end{aligned} \tag{8.100}$$

Below, we establish that the dual of an SOCP is an SOCP. This fact can either be proved via conic duality, or proved directly.

Theorem 200

Let $\vec{c} \in \mathbb{R}^n$, and for $i \in \{1, \dots, m\}$ let $A_i \in \mathbb{R}^{d_i \times n}$, $\vec{y}_i \in \mathbb{R}^{d_i}$, $\vec{b}_i \in \mathbb{R}^n$, and $z_i \in \mathbb{R}$. The dual of the following SOCP in standard form:

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.101}$$

$$\text{s.t.} \quad \|A_i\vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i, \qquad \forall i \in \{1, \dots, m\}. \tag{8.102}$$

can be formulated as an SOCP in standard form.

Proof via Conic Duality. Let $K_i \doteq \{(\vec{u}, r) \in \mathbb{R}^{d_i} \times \mathbb{R} \mid \|\vec{u}\|_2 \leq r\}$ denote the second-order cone in $\mathbb{R}^{d_i+1}$, and let $d = \sum_{i=1}^m d_i$. Then the standard form SOCP can be written as:

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.103}$$

$$\text{s.t.} \quad -\left(A_i\vec{x} - \vec{y}_i, \vec{b}_i^\top \vec{x} + z_i\right) \preceq_{K_i} \vec{0}, \qquad \forall i \in \{1, \dots, m\}. \tag{8.104}$$

The Lagrangian $L \colon \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ can thus be defined as follows. For each $\vec{x} \in \mathbb{R}^n$, $\vec{\lambda} = (\vec{\lambda}_1, \dots, \vec{\lambda}_m) \in \mathbb{R}^d$ (with $\vec{\lambda}_i \in \mathbb{R}^{d_i}$ for each $i \in \{1, \dots, m\}$), and $\vec{\mu} \in \mathbb{R}^m$:

$$L(\vec{x}, \vec{\lambda}, \vec{\mu}) = \vec{c}^\top \vec{x} - \sum_{i=1}^m \left[\vec{\lambda}_i^\top (A_i\vec{x} - \vec{y}_i) + \mu_i(\vec{b}_i^\top \vec{x} + z_i)\right] \tag{8.105}$$

$$= \left(\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i)\right)^\top \vec{x} + \sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i). \tag{8.106}$$