3.2 Linear Programming

A special case of convex optimization is **Linear Programming (LP)**. However, this special case is of a great importance to us. It is not surprising that LP was one of the first optimization problems to be carefully studied and exactly solved; LP is associated with linear algebra, one of the richest and most popular tool sets in all of mathematics. Many later developments in other areas of optimization theory were inspired by LP. For example, the development of the theory of convex optimization to great extent owes to the understanding of LPs. Remember that we previously talked about affine functions and half-spaces, being the building blocks of convex functions and convex sets, respectively. In the same manner, linear programming is the building block of all convex optimization problems.

The main reason that we study linear optimization in this course is that it has a profound application in the theory of discrete optimization. From the early days of discrete optimization studies, it became clear that many problems of interest could be formulated as the so-called **Integer Linear Programs (ILP)**. ILPs are similar to LPs, but different in that their variable space is discrete. Later, we will study ILPs and their relation with LPs. We will see how LPs can solve ILPs or lead to approximate algorithms. Now, we focus on LP and its theory.

Definition 1. *An optimization problem is linear if:*

1. *The variable domain is $\mathbb{R}^n$ for some n.*
2. *The cost function $f(\mathbf{x})$ and the constraint functions $g_i(\mathbf{x})$ and $h_j(\mathbf{x})$ are all affine functions.*

It is popular to write an LP as

$$\begin{gathered}
\max_{\mathbf{x}=(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n} c_1x_1+c_2x_2+\ldots+c_nx_n \\
\text{subject to} \\
a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n \le b_1 \\
a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n \le b_2 \\
\vdots \\
a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n \le b_m \\
, \\
d_{11}x_1+d_{12}x_2+\ldots+d_{1n}x_n = e_1 \\
d_{21}x_1+d_{22}x_2+\ldots+d_{2n}x_n = e_2 \\
\vdots \\
d_{p1}x_1+d_{m2}x_2+\ldots+d_{mn}x_n = e_p
\end{gathered} \tag{1}$$

where the terms $a_{ij}, b_i, c_j, d_{ij}, e_i$ are all arbitrary real coefficients. As seen, it might be difficult to read an optimization, written in the form of (1). Hence, a more concise notation is often used. Let us introduce the following matrix definitions:

$$\begin{gathered}
\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \\
\mathbf{D} = \begin{pmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{p,1} & d_{p,2} & \cdots & d_{p,n} \end{pmatrix} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}
\end{gathered} \tag{2}$$

Now, the optimization in (1) is written as

$$\begin{aligned}
\max_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\
& \mathbf{D}\mathbf{x} = \mathbf{e}
\end{aligned}$$

Notice that we used the symbol "$\le$" to refer to element-wise inequality.

The main issue here is that you do not need all the matrices $\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{D}$ and $\mathbf{e}$ to express an LP. Instead, it is customary to write an LP in either one of these forms: The *standard* or *canonical form*,

$$\begin{aligned}
\max_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\
& \mathbf{x} \ge 0
\end{aligned}$$

and the *augmented form*[1],

$$\begin{aligned} \max_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{c}^\intercal\mathbf{x} \\ \text{s.t.} \quad & \mathbf{D}\mathbf{x} = \mathbf{e} \\ & \mathbf{x} \geq 0 \end{aligned}$$

To turn a general form of LP into one of these forms, the following modifications are straightforward:

- Minimization of $\mathbf{c}^\intercal\mathbf{x}$ can be turned into maximization of $-\mathbf{c}^\intercal\mathbf{x}$.
- Lower bounds $\mathbf{a}_i^\intercal\mathbf{x} \geq b_i$ can be turned into $-\mathbf{a}_i^\intercal\mathbf{x} \leq -b_i$.
- Equality constraints $\mathbf{d}_i^\intercal\mathbf{x} = e_i$ can be turned into two inequality constraints $\mathbf{d}_i^\intercal\mathbf{x} \leq e_i$ and $-\mathbf{d}_i^\intercal\mathbf{x} \leq -e_i$.
- Inequalities $\mathbf{a}_j^\intercal\mathbf{x} \leq b_i$ can be turned into equalities by introducing a new *slack variable* $s_j \geq 0$ (hence the term "augmented" form), and replacing the constraint by $\mathbf{a}_j^\intercal\mathbf{x} + s_j = b_i$.
- Unconstrained variables x_i can be replaced by $x_i' - x_i''$, $x_i' \geq 0$ and $x_i'' \geq 0$.

An important concept in optimization is that of a *dual*:

Definition 2 (Dual)**.** *For a given maximization (minimization) problem, called the* primal problem*, a* dual problem *is a* convex[2] *minimization (maximization) problem for which the optimal solution is an upper (lower) bound to the optimal value of the primal.*

[1]Depending on who you ask, each of these terms can mean anything, and definitions vary widely, so don't be surprised if you read somewhere that the standard form is a minimization, or a maximization with equality constraints, the augmented form is also called canonical and so forth. The reason for this confusion is that what we call the standard form in this lecture is often more intuitive and arises naturally in many real-world problems. However, one of the first methods to solve LP was the *simplex method*, which operates on the augmented form, so many people consider this to be *the* standard form. Others call our standard form standard, but our augmented form canonical. Add to this that *canonical* and *standard* are two words that mean the same thing, and you got a recipe for confusion. Furthermore, geometrically it is more natural to maximize when using inequality constraints, moving away from the origin, but, on the other hand, convex optimization problems operate on convex functions and hence minimize; for maximization, the function would have to be *concave*. $\mathbf{c}^\intercal\mathbf{x}$ however is both, so we got that confusion on top of it all. The only thing that seems to be generally accepted is that the augmented form has equality constraints, since we "augment" the problem by introducing slack variables. In the end, however, it's only terminology, and it's more important to recognize different forms and the context in which they are applied, no matter what you call them.

[2]This property is not always fully enforced. As shown below, an LP has a certain kind of dual which is also convex. If the solutions are restricted to integer $\mathbf{x}$, called an *integer linear program*, the dual takes on the same form as for the LP, with additional integer constraints, so it is not convex, but still called a dual. What is always observed is the upper bound property.

Consider the LP

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \leq 5 \\ & 3x_1 + 4x_2 \leq 6 \end{aligned}$$

In order to derive an upper bound for the objective function, we could multiply the first constraint by 2,

$$2(x_1 + 2x_2) \leq 2 \cdot 5 \Rightarrow 2x_1 + 4x_2 \leq 10$$

Note that each coefficient becomes greater-equal than the ones in the objective function, and therefore

$$2x_1 + 3x_2 \leq 2x_1 + 4x_2 \leq 10$$

so we know that the optimal solution can be at most 10. Also, the factor has to be non-negative, otherwise the constrain would become $\geq$. In fact, we could combine different constraints to get a better bound: let's find some $y_i \geq 0$ to multiply constraints:

$$y_1(x_1 + 2x_2) \leq y_1 5$$

$$y_2(3x_1 + 4x_2) \leq y_2 6$$

which also implies that

$$y_1(x_1 + 2x_2) + y_2(3x_1 + 4x_2) \leq y_1 5 + y_2 6$$

Again, to make sure that the right-hand side is an upper bound to the maximization objective, we require the left one to be an upper bound as well:

$$2x_1 + 3x_2 \leq y_1(x_1 + 2x_2) + y_2(3x_1 + 4x_2)$$

Now observe what happens when we rearrange the terms:

$$2x_1 + 3x_2 \leq x_1(y_1 + 3y_2) + x_2(2y_1 + 4y_2)$$

We can split this inequality up again:

$$2x_1 \leq x_1(y_1 + 3y_2)$$

$$3x_2 \leq x_2(2y_1 + 4y_2)$$

and hence

$$2 \le y_1 + 3y_2$$
$$3 \le 2y_1 + 4y_2$$

Of course, we want to make the upper bound $5y_1 + 6y_2$ on the primal as tight as possible under the given constraints. We therefore have obtained a dual minimization problem:

$$\begin{aligned} \min \quad & 5y_1 + 6y_2 \\ \text{s.t.} \quad & y_1 + 3y_2 \ge 2 \\ & 2y_1 + 4y_2 \ge 3 \\ & \mathbf{y} \ge 0 \end{aligned}$$

It is easy to see how a general form can be obtained: For a *primal problem P*

$$\begin{aligned} \max_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{c}^\intercal \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \ge 0 \end{aligned}$$

dual problem D is defined as

$$\begin{aligned} \min_{\mathbf{y}\in\mathbb{R}^m} \quad & \mathbf{b}^\intercal \mathbf{y} \\ \text{s.t.} \quad & \mathbf{A}^\intercal \mathbf{y} \ge \mathbf{c} \\ & \mathbf{y} \ge 0 \end{aligned}$$

These definitions can be used to obtain an upper bound on the maximal solution for P by minimizing the objective function for D. The upper bound property can be formalized as follows:

Theorem 1 (Weak duality in LP). *If there exists a feasible* $\mathbf{x}$ *in P, and a feasible* $\mathbf{y}$ *in D, then* $\mathbf{c}^\intercal\mathbf{x} \le \mathbf{b}^\intercal\mathbf{y}$.

Proof. Consider the constraint $\mathbf{A}\mathbf{x} \le \mathbf{b}$ in P. We can left-multiply it with any non-negative vector $\mathbf{v} \ge 0$, and the inequality $\mathbf{v}^\intercal\mathbf{A}\mathbf{x} \le \mathbf{v}^\intercal\mathbf{b}$ holds. Equivalently, if $\mathbf{c} \le \mathbf{A}^\intercal\mathbf{y}$, then $\mathbf{c}^\intercal \le (\mathbf{A}^\intercal\mathbf{y})^\intercal = \mathbf{y}^\intercal\mathbf{A}$. We can right-multiply this with any non-negative vector $\mathbf{w} \ge 0$ to obtain $\mathbf{c}^\intercal\mathbf{w} \le \mathbf{y}^\intercal\mathbf{A}\mathbf{w}$. Now assume that the feasible regions in P and D are not empty. Then there exist $\mathbf{x}, \mathbf{y}$, and we can pick $\mathbf{y} = \mathbf{v}$ and $\mathbf{x} = \mathbf{w}$, to obtain $\mathbf{c}^\intercal\mathbf{x} \le \mathbf{y}^\intercal\mathbf{A}\mathbf{x} \le \mathbf{b}^\intercal\mathbf{y}$. □

The difference between the two objectives is called the *duality gap*. In fact, for linear programs, there is an even stronger theorem, which we are not going to prove:

Theorem 2 (Strong duality of LP). *If P has a feasible optimal solution* $\mathbf{x}^*$*, then D has an optimal feasible solution* $\mathbf{y}^*$*, and* $\mathbf{c}^\intercal\mathbf{x}^* = \mathbf{y}^{*\intercal}A\mathbf{x}^* = \mathbf{b}^\intercal\mathbf{y}^*$. *If P is unbounded, then D is infeasible. If D is unbounded, then P is infeasible. If either one is infeasible, the other is infeasible or unbounded.*

In other words, the duality gap in LP is zero, except when both are infeasible, then it is infinite. Strong duality tells us that we can solve a maximization problem by solving its dual and vice versa. We can quickly decide if a problem is unbounded. Also, in LP, the dual of the dual is the primal (try for yourself!).

There is another way to determine whether a solution to a primal is optimal, using a property called *complementary slackness*. Consider the first equality in the strong duality theorem, $\mathbf{c}^\intercal\mathbf{x}^* = \mathbf{y}^{*\intercal}A\mathbf{x}^*$. This can be written as follows:

$$\begin{aligned} \mathbf{c}^\intercal\mathbf{x}^* &= \mathbf{y}^{*\intercal}A\mathbf{x}^* \\ \mathbf{c}^\intercal\mathbf{x}^* - \mathbf{y}^{*\intercal}A\mathbf{x}^* &= 0 \\ (\mathbf{c}^\intercal - \mathbf{y}^{*\intercal}A)\mathbf{x}^* &= 0 \\ (\mathbf{c} - \mathbf{A}^\intercal\mathbf{y}^*)^\intercal\mathbf{x}^* &= 0 \end{aligned}$$

Since $\mathbf{y}^*$ is feasible, we know that $\mathbf{c} - \mathbf{A}^\intercal\mathbf{y}^* \le 0$, and therefore $c_j - \sum_i a_{ij}y_i^* \le 0$ for all j. On the other hand, we know that $x_j^* \ge 0$ for all j, since $\mathbf{x}^*$ is also feasible. Since

$$(\mathbf{c} - \mathbf{A}^\intercal\mathbf{y}^*)^\intercal\mathbf{x}^* = \sum_j x_j^* \left(c_j - \sum_i a_{ij}y_i^* \right) = 0$$

and none of its terms can be positive, we have

$$\forall j : x_j^* \left(c_j - \sum_i a_{ij}y_i^* \right) = 0 \tag{3}$$

because there can be no positive terms to cancel out any negative ones. This means that in order to satisfy eq. (3), for every j, at least one of $\left(c_j - \sum_i a_{ij}y_i^*\right)$ and x_j must be 0. $c_j - \sum_i a_{ij}y_i^* = 0$ means that $\sum_i a_{ij}y_i^* \ge c_j$ is satisfied with equality.

Definition 3. *An inequality constraint* $\mathbf{a}^\intercal x \le b$ *is said to be* binding *or* active *if for some given* $\mathbf{x}'$ *it is satisfied with equality, i.e.* $\mathbf{a}^\intercal\mathbf{x}' = b$*. It is said to* have slack *if it is not binding.*

An analogous argument can be made for $\mathbf{y}^{*\intercal}A\mathbf{x}^* = \mathbf{b}^\intercal\mathbf{y}^*$. This yields the following important result:

Theorem 3 (Complementary slackness). *If optimal solutions* $\mathbf{x}^*, \mathbf{y}^*$ *exist for the primal and dual LPs P and D, then the following holds:*

				$\max \mathbf{c}^\intercal \mathbf{x}$				
				0		0		
				$\leq$		$\geq$		
				x_1	x_2	x_3		
$\min \mathbf{b}^\intercal \mathbf{y}$	0	$\leq$	y_1	a_{11}	a_{12}	a_{13}	$\leq$	b_1
			y_2	a_{21}	a_{22}	a_{23}	$=$	b_2
	0	$\geq$	y_3	a_{31}	a_{32}	a_{33}	$\geq$	b_3
				$\geq$	$=$	$\leq$		
				c_1	c_2	c_3		

Figure 1

- *If the i-th primal constraint* $\sum_j a_{ij} x_j^* \leq b_i$ *has slack, then the i-th dual variable* y_i^* *is zero.*
- *If the j-th dual constraint* $\sum_i a_{ij} y_i^* \geq c_j$ *has slack, then the j-th primal variable* x_j^* *is zero.*
- *If the i-th dual variable* y_i^* *is positive, then the i-th primal constraint is* binding.
- *If the j-th primal variable* x_j^* *is positive, then the j-th dual constraint is* binding.

There is an easy way to express duality for LPs in arbitrary form (fig. 1), which follows from the way we transformed constrains into different forms, as described before. Notice for instance that the dual variable of an equality constraint is unconstrained (here: y_2 and x_2). For a $\geq$-constraint in a maximization problem, the dual variable is non-positive (y_3). For inequality constraints that point into the opposite direction than what we would expect from the standard form, the bounds of the dual variables also change direction (here: $x_3 \leq 0$, because c_3 is bounded from below instead of above, and $y_3 \leq 0$, because b_3 is bounded from above instead of below). Note that this general duality contains our standard form as a special case. Some authors also consider the special case in which all primal (dual) variables are unconstrained. These pairs are sometimes called *asymmetric duals*:

$$\begin{array}{rlcrl} \max & \mathbf{c}^\intercal \mathbf{x} & \qquad & \min & \mathbf{b}^\intercal \mathbf{y} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} & & \text{s.t.} & \mathbf{A}^\intercal \mathbf{y} = \mathbf{c} \\ & & & & \mathbf{y} \geq 0 \end{array}$$

and

$$\begin{array}{rlcrl} \max & \mathbf{c}^\intercal \mathbf{x} & \qquad & \min & \mathbf{b}^\intercal \mathbf{y} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} & & \text{s.t.} & \mathbf{A}^\intercal \mathbf{y} \geq \mathbf{c} \\ & \mathbf{x} \geq 0 & & & \end{array}$$

Primal constraints and dual variables are in rows (i), dual constraints and primal variables are in columns (j). For instance, if y_1^* is positive, then $a_{11}x_1^* + a_{12}x_2^* + a_{13}x_3^* = b_1$. If that's not the case, $\mathbf{x}^*$ and $\mathbf{y}^*$ cannot be optimal solutions.

Example 1. *Consider the primal*

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{R}^2} \; & x_1 - x_2 \\ \text{s.t.} \; -2x_1 + x_2 &\leq 2 \\ x_1 - 2x_2 &\leq 2 \\ x_1 + x_2 &\leq 5 \\ \mathbf{x} &\geq 0 \end{aligned}$$

and its dual

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{R}^3} \; & 2y_1 + 2y_2 + 5y_3 \\ \text{s.t.} \; -2y_1 + y_2 + y_3 &\geq 1 \\ y_1 - 2y_2 + y_3 &\geq -1 \\ \mathbf{y} &\geq 0 \end{aligned}$$

The claim is that $\mathbf{x} = (1,4)^\intercal$ *is the optimal solution. First, one can easily check that* $\mathbf{x}$ *is a feasible solution in* P*. Now, note that the first and third constraint in* P *are binding, which means* y_1 *and* y_3 *would have to be non-zero. On the other hand, the second constraint in* P *has slack, so* y_2 *has to be zero. Now, since* x_1 *and* x_2 *are non-zero, the first two constraints in the dual should be binding, so the dual should take the form*

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{R}^2} \; & 2y_1 + 5y_3 \\ \text{s.t.} \; -2y_1 + y_3 &= 1 \\ y_1 + y_3 &= -1 \\ \mathbf{y} &\geq 0 \end{aligned}$$

which yields $\mathbf{y} = \left(-\frac{2}{3}, 0, -\frac{1}{3}\right)^\intercal$*. Since the negative values violate the constraint* $\mathbf{y} \geq 0$*, the dual is not feasible and hence* $(1,4)^\intercal$ *cannot be the optimal solution to* P*.*