1 Objective: Moving from 36900 To 36904

The objective of this module is to understand key problems that arise when moving from finite dimensional linear programs to infinite dimensional linear programs. Students are introduced to the theory of linear programming in 36900. In 36900 the primary focus is on linear and integer programs with a finite number of constraints and variables. However, there are many applications where the number of variables, and or constraints, is infinite. Infinite dimensional optimization problems are the topic of 36904. Hopefully, this module will lead to unbounded and infinite enthusiasm for taking 36904.

2 The General Linear Program

These notes are a modification of the material found in Optimization Over Ordered Vector Spaces by Kipp Martin and Chris Ryan. Let $X$ denote an arbitrary vector space (possibly infinite dimensional) and let $X'$ denote the vector space of linear functionals on defined on $X$. That is, $X'$ is the vector space of all linear functions $\phi: X \rightarrow \mathbb{R}$. You should verify that this is indeed a vector space. The vector space $X'$ is often called the algebraic dual of $X$. Likewise, $Y$ is also an arbitrary vector space with algebraic dual $Y'$. The corresponding primal-dual schema is

$$
P : \begin{bmatrix} X \\ X' \end{bmatrix} \rightarrow \begin{bmatrix} Y \\ Y' \end{bmatrix} : D \tag{2.1}$$

The paradigm expressed in (2.1) is extremely powerful. When you move to infinite dimensions, it is extremely important to be precise about which vector spaces you are working with. In the paradigm of (2.1),

- $X$ is the primal linear program variable vector space,
- $Y$ is the primal linear program constraint vector space,
- $Y'$ is the dual linear program variable evector space,
- $X'$ is the primal linear program constraint vector space.
Consider the general linear program
\[
\inf \ \langle x, \phi \rangle_X \\
\text{s.t.} \ b - A(x) \preceq_{P_Y} \theta_Y
\]

where \( b \in Y \), \( P_Y \) is a pointed (dimension of the lineality space is zero), convex cone in \( Y \) and \( A : X \to Y \) is a linear mapping of vector spaces \( X \) and \( Y \). That is, \( A(\alpha_1 x^1 + \alpha_2 x^2) = \alpha_1 A(x^1) + \alpha_2 A(x^2) \) for all \( x^1, x^2 \in X \) and \( \alpha_1, \alpha_2 \in \mathbb{R} \). The problem’s label of \((\text{GLP})\) stands for general LP as this is the most general LP formulation that we consider. In this notation, \( \phi \in X' \) and \( \langle x, \phi \rangle_X \) denotes the mapping of \( x \in X \) into \( \mathbb{R} \) under the linear functional \( \phi \). The constraints \( b - A(x) \preceq_{P_Y} \theta_Y \) mean that \( A(x) - b \) is in the cone \( P_Y \).

Classical (finite dimensional) linear programming has \( X = \mathbb{R}^n \), \( Y = \mathbb{R}^m \), and \( P_Y = \mathbb{R}^m_+ \). In this setting, \((\text{GLP})\) is the familiar:
\[
\min \ c^\top x \\
\text{s.t.} \ Ax \geq b
\]

as commonly seen in the literature, where \( A \) is an \( m \times n \) matrix, \( c \in \mathbb{R}^n \), \( x \in \mathbb{R}^n \), and \( b \in \mathbb{R}^m \). Every piece of the optimization problem when one compares \((\text{GLP})\) and \((\text{FLP})\) looks a bit different:

(i) The objective turns from \( \langle x, \phi \rangle \) to \( c^\top x \).

(ii) The constraint turns from \( b - A(x) \preceq_{P_Y} \theta_Y \) to \( Ax \geq b \).

(iii) The \( \inf \) turns into \( \min \).

We discuss these changes in detail now.

(i) If \( X = \mathbb{R}^n \) then \( X \) is isomorphic to \( X' \) and the linear functional \( \phi \) takes on a form \( \phi_c \) for some \( c \in \mathbb{R}^n \) and this gives
\[
\langle x, \phi_c \rangle = c^\top x.
\]

(ii) The constraint has two changes. First, the functional notation \( A(x) \) is replaced with a vector-matrix product \( Ax \). The linear map \( A \) usually is represented by its associated matrix,\(^1\) also denoted by \( A \). In this setting, \( A \) is an \( m \times n \) matrix and \( A(x) \) is the vector-matrix product \( Ax \).

The second change is swapping the \( \preceq_P \) ordering for the \( \leq \) ordering and putting \( b \) on the right hand side. In standard linear programming formulations \( P_Y \) is the positive orthant in \( \mathbb{R}^m \), that is
\[
P_Y = \mathbb{R}^m_+ = \{ y \in \mathbb{R}^m : y_i \geq 0, i = 1, \ldots, m \}
\]
where the notation \( \geq \) is standard shorthand for \( \succeq \).

(iii) In the general case of \((\text{GLP})\), the objective function is \( \inf \) and not \( \min \). The fact that every feasible and bounded (we are referring to objective function value, not feasible region) finite dimensional linear program satisfies strong duality is something usually taken for granted. This is not always the case for infinite dimensional conic linear programs.

\(^1\) Let \( A : \mathbb{R}^n \to \mathbb{R}^m \) be a linear mapping of vector spaces. Let \( e^i \in \mathbb{R}^n \) be the \( i \)th standard basis vector for \( \mathbb{R}^n \). The matrix \( M_A \) (often simply called \( A \) despite the potential for confusion) has as its \( i \)th column the image of \( e^i \) under the mapping \( A \). That is, \((M_A)_e = A(e^i)\).
It is important to distinguish between an optimal solution and an optimal solution value. A problem with both an optimal solution value and an optimal solution is solvable. However, it is possible to have an optimal solution value, but not an optimal solution.

Example 2.1.

\[
\inf x_1 \\
x_1 + \frac{1}{i^2}x_2 \geq \frac{2}{i}, \quad i = 2, 3, \ldots
\]

Observe that \((\frac{1}{\delta}, \delta)\) is feasible for all \(\delta > 0\). This is because for every \(i \in \mathbb{N}\),

\[
(i - \delta)^2 \geq 0 \\
\Rightarrow i^2 + \delta^2 - 2\delta i \geq 0 \\
\Rightarrow i^2 + \delta^2 \geq 2\delta i \\
\Rightarrow \frac{i^2 + \delta^2}{\delta^2} \geq \frac{2\delta i}{\delta^2} \\
\Rightarrow \frac{1}{\delta} + \frac{1}{\delta^2} \geq \frac{2}{\delta}.
\]

What is the optimal solution value to this linear program? What is the optimal solution?

3 Projection and the Lagrangian Dual

We solve the finite linear program \((FLP)\) using projection by: 1) writing it as

\[
\begin{align*}
\min z \\
z - c^\top x & \geq 0 \\
Ax & \geq b,
\end{align*}
\]

2) finding a nonnegative \(v\) vector, and 3) aggregating the constraints \(Ax \geq b\) with the objective function to give

\[
\min z \\
z - c^\top x + (v^\top A)x \geq v^\top b.
\]

Since we want to minimize \(z\) we will clearly pick an \(x \in \mathbb{R}^n\) that makes

\[c^\top x + v^\top (b - Ax)\]

as small as possible. That is, for a given \(v\), we want to solve the problem

\[
\min\{c^\top x + v^\top (b - Ax) : x \in \mathbb{R}^n\}.
\]

The solution to this problem is a lower bound on the optimal primal objective function value. Of course, every nonnegative \(v\) is a valid multiplier vector so we want to solve

\[
\max_{v \geq 0} \left\{ \min\{c^\top x + v^\top (b - Ax) : x \in \mathbb{R}^n\} \right\}. \tag{3.1}
\]
Observe that 3.1 is equivalent to

$$\max_{v \geq 0} \left\{ v^\top b + \min \{(c^\top - v^\top A)x : x \in \mathbb{R}^n\} \right\}.$$ 

Written this way, it is obvious that the inner minimization will always be $-\infty$ if $c^\top - v^\top A \neq 0$. Why is this the case? Hence, we want to always pick a nonnegative $v$ such that $c^\top - v^\top A = 0$. Therefore 3.1 becomes

$$\max v^\top b$$

$$c^\top - v^\top A = 0$$

$$v \geq 0.$$

Aha! The standard finite linear programming dual.

Returning to general problem, let’s apply the projection approach to (GLP). Write (GLP) as

$$\min z$$

$$z - \langle x, \phi \rangle_X \geq 0$$

$$A(x) \succeq_{P_Y} b$$

Now, generalize the idea of aggregating $Ax \geq b$ to aggregating $A(x) \succeq_{P_Y} b$. We aggregate the constraints $Ax \geq b$ using a nonnegative vector $v$. In terms of vector spaces, $v \in Y'$ where $Y = \mathbb{R}^m$. Let’s just aggregate $A(x) \succeq_{P_Y} b$ using a nonnegative dual functional. Choose a dual vector $\psi \in Y'$ from the dual cone

$$P_Y^+ := \{ \psi \in Y' : \langle y, \psi \rangle_Y \geq 0, \forall y \in P_Y \}. \quad (3.2)$$

The elements of the dual cone are called positive linear functionals. Using the positive linear functions, the generalization of (3.1) is the Lagrangian dual

$$\sup_{\psi \in P_Y^+} \left\{ \inf_{x \in X} \left\{ \langle x, \phi \rangle_X + \langle b - A(x), \psi \rangle_Y \right\} \right\}. \quad (3.3)$$

Key Question: How does the solution optimal solution value in (3.3) compare with the optimal solution value to (GLP)?

Theorem 3.1. (Weak Duality) If $\overline{x}$ is feasible to (GLP) and $\overline{\psi} \in P_Y^+$, then

$$\langle \overline{x}, \phi \rangle_X \geq \left\{ \inf_{x \in X} \left\{ \langle x, \phi \rangle_X + \langle b - A(x), \overline{\psi} \rangle_Y \right\} \right\}.$$  

Proof. Since $\overline{x}$ is a feasible solution (GLP), $b - A(\overline{x}) \succeq_{P_Y} \theta_Y$. But $\overline{\psi} \in P_Y^+$ so $\langle b - A(\overline{x}), \overline{\psi} \rangle_Y \leq 0$. Therefore,

$$\langle \overline{x}, \phi \rangle_X \geq \langle \overline{x}, \phi \rangle_X + \langle b - A(\overline{x}), \overline{\psi} \rangle_Y \geq \inf_{x \in X} \left\{ \langle x, \phi \rangle_X + \langle b - A(x), \overline{\psi} \rangle_Y \right\}.$$ 

\[ \square \]
Corollary 3.2. (Weak Duality) If $v(\text{GLP})$ is the optimal value of the general linear program (GLP), then

$$v(\text{GLP}) \geq \sup_{\psi \in P^+_Y} \left\{ \inf_{x \in X} \left\{ \langle x, \phi \rangle_X + \langle b - A(x), \psi \rangle_Y \right\} \right\}.$$ 

Corollary 3.2 tells us that the value of the Lagrangian dual is a lower bound on the optimal value of the general linear program.

4 The Adjoint

Now simplify the Lagrangian dual (3.3) by taking the term $\langle b, \psi \rangle_Y$ out of the inner infimum since this term does not involve $x$. This gives

$$\sup_{\psi \in P^+_Y} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \left\{ \langle x, \phi \rangle_X - \langle A(x), \psi \rangle_Y \right\} \right\}.$$  \hspace{1cm} (4.1)

Now consider the inner minimization in $x$. Notice that the the terms in angle braces $\langle x, \phi \rangle_X$ and $\langle A(x), \psi \rangle_Y$ represent linear functionals over two different vector spaces $X$ and $Y$. How can we combine them? Moreover, $x$ appears in a different way in both terms, in the latter term, $x$ enters as an argument of a linear map. How is this simplified?

The dual for a finite dimensional linear program is much cleaner and provides motivation for the adjoint mapping. The dual of the finite linear program (FLP) is

$$\begin{align*}
\text{max } & b^\top y \\
\text{s.t. } & A^\top v = c \\
& v \geq 0.
\end{align*}$$  \hspace{1cm} (FLPD)

Even more puzzling than turning (GLP) into (FLP) is turning (4.1) into (FLPD). Consider (4.1) in the case of the finite linear program (FLP). In this case, the dual vector is $\psi \in (\mathbb{R}^m)'$. However, in view of the isomorphism between $\mathbb{R}^m$ and $\mathbb{R}^m'$ we know $\psi = \psi_v$ for some $v \in \mathbb{R}^m$ where $\langle y, \psi_v \rangle = v(y) = v^\top y$. The expression inside the inner minimization of (4.1) is

$$\langle x, \phi \rangle_X - \langle A(x), \psi \rangle_Y = \langle x, c \rangle_{\mathbb{R}^n} - \langle A(x), v \rangle_{\mathbb{R}^m} = c^\top x - v^\top (Ax) = c^\top x - (v^\top A)x = \langle x, c \rangle_{\mathbb{R}^n} - \langle x, A^\top v \rangle_{\mathbb{R}^n}.$$  

We used in the third equality above the associativity of matrices that $v^\top (Ax) = (v^\top A)x$. Here is a key observation. The vector $v$ represents a linear functional in $\mathbb{R}^m$ and the vector $A^\top v$ is a vector that represents a linear functional in $\mathbb{R}^n$. Since $c$ and $A^\top v$ are linear functionals in the same space $X'$ they can be combined.

$$\langle x, c \rangle_{\mathbb{R}^n} - \langle x, A^\top v \rangle_{\mathbb{R}^n} = \langle x, c - A^\top v \rangle_{\mathbb{R}^n}.$$
This gives a new inner minimization problem
\[ \inf_{x \in X} (c^T - v^T A)x. \]

Next, apply standard logic and observe that this optimization problem has a value of \(-\infty\) unless \(c^T - v^T A = 0\). This yields (FLPD). Can this be extended in a similar way for a general LP? The answer is a resounding YES! It is enabled by a powerful “dual” (arghh, another dual concept!) to the linear map \(A\) called its adjoint.

**Definition 4.1.** Given a linear mapping \(A : X \rightarrow Y\), there is an associated mapping \(A'\) called the adjoint that maps \(Y'\) into \(X'\). The image of \(\psi \in Y'\) under \(A'\) is denoted by \(A'(\psi)\). Now, \(A'(\psi)\) is a linear functional on the vector space \(X\) and operates on an \(x \in X\) by \(\psi \circ A(x)\), or equivalently
\[
\langle x, A'(\psi) \rangle_X = \langle A(x), \psi \rangle_Y \quad \forall x \in X, \psi \in Y'. \tag{4.2}
\]

An equivalent way to express (4.2) is
\[
A'(\psi)(x) = \psi(A(x))
\]
but the “angle brace” statement is often preferred to avoid use of the two sets of parentheses on both sides of the equation needed here.\(^2\) See Figure 1 for a second way to view the adjoint map. This Figure illustrates that \(A'\) is the composition map \(\psi \circ A\). Think of \(A'(\psi)\) as an “equivalent” notation for \(\psi \circ A\). Both \(A'(\psi)\) and \(\psi \circ A\) are linear functionals that take an element of \(x \in X\) and map it to the real numbers.

![Figure 1: A visualization of the adjoint map \(A'\).](image)

**Example 4.2.** The matrix notion of transpose is really just a special case of an adjoint. Consider a linear mapping \(A : \mathbb{R}^n \rightarrow \mathbb{R}^m\). Again abusing notation, \(A\) refers to both the mapping and its associated \(m \times n\) matrix. We ask you to keep track of which is which. The notation \(A'\) is used for the adjoint of the mapping \(A\), and \(A^\top\) for the transpose of the matrix \(A\).

Let’s think about the adjoint of \(A\). The adjoint of \(A\) is a mapping from \((\mathbb{R}^m)'\) to \((\mathbb{R}^n)'\). We claim the adjoint map of \(A\) is
\[
A' : (\mathbb{R}^m)' \rightarrow (\mathbb{R}^n)' \tag{4.3}
\]
\[
\psi_v \mapsto \phi_{A^\top b} \tag{4.4}
\]
\(^2\)To obscure things even further some authors drop the parentheses in the statement \(A(x)\) and write \(Ax\) when \(A\) is a linear map. We avoid this notation as much as possible, but we may sometimes fall prey to its convenience.
where $A^\top$ is the usual transpose of the matrix $A$. In other words, it takes an arbitrary linear functional of $\mathbb{R}^m$, $\psi_v$, into the linear functional of $\mathbb{R}^n$, $\phi_{A^\top v}$. A sanity check reveals that $A^\top v$ is in fact a vector in $\mathbb{R}^n$.

Let’s verify that $A' = A^\top$ is indeed the adjoint of $A$. To do this verify (4.2). That is

$$\langle A(x), \psi_v \rangle_{\mathbb{R}^m} = \langle x, A'(\psi_v) \rangle_{\mathbb{R}^n} \quad (4.5)$$

for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$. Start by developing the right-hand side:

$$\langle A(x), \psi_v \rangle_{\mathbb{R}^m} = \sum_{i=1}^m \left( \sum_{k=1}^n a^k(i)x_kv(i) \right)$$

$$= \sum_{i=1}^m \left( \sum_{k=1}^n a^k(i)x_kv(i) \right)$$

$$= \sum_{k=1}^n \left( \sum_{i=1}^m a^k(i)x_kv(i) \right)$$

$$= \sum_{k=1}^n \left( \sum_{i=1}^m a^k(i)v(i) \right)$$

$$= \left( A^\top v \right)^\top x = \langle x, A^\top v \rangle_{\mathbb{R}^n}$$

This verifies that $A^\top$ is indeed the adjoint of $A$. Since the adjoint $A'$ is determined entirely by the matrix $A^\top$, we often confound notation and say the adjoint of $A$ is the matrix $A^\top$. As long as you can keep track, then no worries!

With a new understanding of the adjoint, continue with (4.1) which is reproduced here for convenience, and finish the derivation of the Lagrangian dual.

$$\sup_{\psi \in P^+_Y} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \{ \langle x, \phi \rangle_X - \langle A(x), \psi \rangle_Y \} \right\}. \quad (4.6)$$

The second term in the inner minimization $\langle A(x), \psi \rangle_Y$ is the definition of the adjoint in (4.2) and equals $\langle x, A'(\psi) \rangle_X$. Collecting terms in $x$, rewrite (4.6)

$$\sup_{\psi \in P^+_Y} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \{ \langle x, \phi - A'(\psi) \rangle_X \} \right\}. \quad (4.7)$$

Finally, employ a key observation! The inner minimization shoots off to $-\infty$ unless $\phi - A'(\psi) = \theta_X$, where $\theta_X$ is the linear functional that takes every element in $X$ to 0. That is we **project out the $x$-variables**. The algebraic dual as a generalization of (FLPD) is

$$\sup_{\psi \in P^+_Y} \{ \langle b, \psi \rangle \} \quad \text{s.t. } \phi - A'(\psi) = \theta_X, \quad \psi \in P^+_Y. \quad (GLPD)$$

A huge problem in infinite dimensional optimization is trying to characterize an adjoint $A'(\psi)$. This can be really hard, if not impossible, to do.
5 Semi-Infinite Programming

Consider the semi-infinite problem where $I = \mathbb{N}$, the set of natural numbers. Write the problem as

$$\sup \sum_{k=1}^{n} c_k x_k \geq \sum_{k=1}^{n} a_k^i x_k, \quad i \in \mathbb{N}. \quad (5.1)$$

In this formulation the $a_k$ for $k = 1, \ldots, n$ and $b$ are functions with domain $\mathbb{N}$. Thus, in the paradigm given in (2.1), $X = \mathbb{R}^n$ and $Y = \mathbb{R}^N$ (the vector space of all real sequences). What is $X'$ and $Y'$?

Since $X$ is a finite dimensional vector space, we can take $X'$ to also be $\mathbb{R}^n$. Indeed, that is why we write the objective function as $c^T x$. Each $\phi \in X'$ corresponds to an $n-$vector $c$. The $Y'$ is much nastier. Indeed, it cannot be characterized. However, all is not lost. Recall, that we really only need the linear functionals from the dual cone $P_+^Y$ defined in (3.2). For the case of $Y = \mathbb{R}^N$, Basu, Martin, and Ryan have characterized the linear functions $\psi \in P_+^Y$. See “On the Sufficiency of Finite Support Duals in Countable Semi-Infinite Linear Programming,” Operations Research Letters, Volume 42 (1) 16-20 (2014). It turns out that $P_+^Y \cong \mathbb{R}^N_+$ where $\mathbb{R}^N_+$ denotes the set of nonnegative sequences with finite support. In other words, we can take $\psi \in P_+^Y$ to be a nonnegative sequence with a finite number of strictly positive elements.

Given the finite support result, let’s characterize the adjoint mapping. In what follows we let $v$ be a finite support vector. If $v$ is a finite support vector, we know $\langle A(x), v \rangle$ is always well defined, i.e. a real number $r$. Therefore,

$$r = \langle A(x), v \rangle = \sum_{i=1}^{\infty} \left( \sum_{k=1}^{n} a_k^i x_k \right) v(i)$$

So what is $A'(y)$? Since $v$ is finite support, there exits an integer $m$ such that $v(i) = 0$ for all $i > m$. Therefore

$$r = \sum_{i=1}^{\infty} \left( \sum_{k=1}^{n} a_k^i x_k \right) v(i) = \sum_{i=1}^{m} \left( \sum_{k=1}^{n} a_k^i x_k \right) v(i) = \sum_{i=1}^{m} \left( \sum_{k=1}^{n} v(i) a_k^i x_k \right) = \sum_{k=1}^{m} \left( \sum_{i=1}^{n} v(i) a_k^i x_k \right) = \sum_{k=1}^{m} \left( \sum_{i=1}^{m} v(i) a_k^i \right) = \sum_{k=1}^{n} \left( \sum_{i=1}^{m} v(i) a_k^i \right) = \langle x, A^T v \rangle.$$ 

But, by definition of adjoint, $\langle A(x), v \rangle = \langle x, A'(v) \rangle$ and so we take $A^T v$ to be the adjoint of $A$. 

8
Thus, the dual linear program (GLPD) becomes
\[ \sup \sum_{i=1}^{\infty} b(i)v(i) \]
\[ \text{s.t. } \sum_{i=1}^{\infty} a^k(i)v(i) = c_k, \quad k = 1, \ldots, k \]
\[ v(i) \geq 0 \]
\[ v \in \mathbb{R}^N \]

To summarize, we know:

1. If \( Y = \mathbb{R}^N \), then \( P_Y^+ \cong \mathbb{R}_+^N \). This result allows us to characterize the adjoint operator and write the dual as (5.2).

2. If \( v \in \mathbb{R}_+^N \) is a finite support vector, it can be written as a conic combination of finite support multipliers generated by applying Fourier-Motzkin elimination to (5.1). This result allows us to conclude that if (5.2) is solvable, we will find an optimal dual vector from Fourier-Motzkin elimination.

We know by Theorem 3.1 that weak duality holds for the pair (5.1)-(5.2). What about strong duality?

### 6 Countably Infinite Linear Programs

**Example 6.1.** This example is from “A Linear Programming Approach to Nonstationary Infinite Horizon Markov Decision Processes,” by Arhis Ghate and Robert L. Smith. See *Operations Research* 61 (2) 413-425 2013. The primal linear program is

\[ \min \sum_{i=1}^{\infty} \frac{1}{2i} x_{i,i+1} \]
\[ x_{1,2} = 1 \]
\[ x_{i,i+1} - x_{i-1,i} = 0, \quad i = 2, 3, \ldots, \]
\[ x_{i,i+1} \geq 0, \quad i = 1, 2, \ldots \]

Ghate and Smith form what they call the natural dual. They write this as

\[ \max v_1 \]
\[ v_i - v_{i+1} \leq \frac{1}{2i}, \quad i = 1, 2, \ldots, \]

Ghate and Smith go on to observe that there is only one feasible solution, \( x_{i,i+1} = 1 \) for \( i = 1, 2, \ldots \). This gives a primal solution value of 1. Next they observe that \( v_i = \theta \) for \( i = 1, 2, \ldots \), is a feasible dual solution. Therefore, the value of dual is strictly greater than the value of the primal for \( \theta > 1 \). Huh? What? We violate weak duality! How can that be? I am not smart enough to figure this out, you will have to do so for homework.
7 Summary

The general linear program for the primal-dual schema

\[
P : \begin{bmatrix} X \\ X' \end{bmatrix} \rightarrow \begin{bmatrix} Y \\ Y' \end{bmatrix} : \mathbb{D}
\]

is

\[
\inf (x, \phi)_X \\
\text{s.t. } \theta_Y \succeq_P b - A(x).
\]

(7.1)

The corresponding finite linear program where \( X = \mathbb{R}^n \) and \( Y = \mathbb{R}^m \) is

\[
\inf (x, \phi)_{\mathbb{R}^n} \\
\text{s.t. } \theta_{\mathbb{R}^m} \succeq_{P_{\mathbb{R}^m}} b - A(x).
\]

(7.2)

Use the fact \( \mathbb{R}^n \cong (\mathbb{R}^n)' \) and represent a linear functional \( \phi \in (\mathbb{R}^n)' \) with \( c \in \mathbb{R}^n \) and define \( P_{\mathbb{R}^m} \) to be the nonnegative cone in \( \mathbb{R}^m \). This gives

\[
\inf c^\top x \\
\text{s.t. } 0 \geq b - Ax.
\]

(7.3)

Now write the linear programs in the format we use for aggregation and projection. The general linear program is

\[
\inf z \\
\text{s.t. } z \geq (x, \phi)_X \\
\quad \theta_Y \succeq_P b - A(x)
\]

(7.4)

and the corresponding finite linear program is

\[
\inf z \\
\text{s.t. } z \geq c^\top x \\
\quad 0 \geq b - Ax.
\]

(7.5)

We aggregate by applying a linear functional \( \psi \in P_{\mathbb{Y}}^+ \) to both sides of the constraint \( \theta_Y \succeq_P b-A(x) \) and adding the result to the constraint \( z \geq (x, \phi)_X \). This gives

\[
z + \langle \theta_y, \psi \rangle_Y \geq (x, \phi)_X + \langle b - A(x), \psi \rangle_Y
\]

which reduces to

\[
z \geq (x, \phi)_X + \langle b - A(x), \psi \rangle_Y.
\]

(7.6)

In the case of finite linear programming, \( Y = \mathbb{R}^m \) so we associate \( \psi \in Y' \) with the \( m \)-vector \( v \). Using a nonnegative \( v \) to aggregate the constraints \( 0 \geq b - A(x) \) with \( z \geq c^\top x \) gives

\[
z \geq c^\top x + v^\top (b - Ax).
\]

By (7.7), the aggregation \( (x, \phi)_X + \langle b - A(x), \psi \rangle_Y \) provides a lower bound on the value of \( z \). The objective is to find the infimum of \( z \). Hence, for a given \( \psi \), we want an \( x \in X \) that makes \( (x, \phi)_X + \langle b - A(x), \psi \rangle_Y \) as small as possible. That is, for a given \( \psi \), we solve the problem

\[
\inf_{x \in X} \{ (x, \phi)_X + \langle b - A(x), \psi \rangle_Y \} = \langle b, \psi \rangle_Y + \inf_{x \in X} \{ (x, \phi)_X - \langle A(x), \psi \rangle_Y \}.
\]
In the finite case this is
\[ v^T b + \inf_{x \in \mathbb{R}^n} \{ c^T x - v^T (Ax) \}. \]

Of course aggregating the constraints \( \langle x, \phi \rangle_X + \langle b - A(x), \psi \rangle_Y \) using any \( \psi \in P_Y^+ \) provides a valid lower bound. Hence the strongest lower bound is given by
\[ \sup_{\psi \in P_Y^+} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \{ \langle x, \phi \rangle_X - \langle A(x), \psi \rangle_Y \} \right\} \]

which is the **Lagrangian Dual**. This provides a lower bound on the optimal primal value by weak duality. See Theorem 3.1. The problem with the Lagrangian dual above is that we would like to simplify \( c^T x - v^T (Ax) \). Unfortunately \( \phi \) and \( \psi \) are linear functionals in different dual spaces.

Taking a clue from the finite case consider
\[ \sup_{y \geq 0} \left\{ v^T b + \inf_{x \in \mathbb{R}^n} \{ c^T x - v^T (Ax) \} \right\} \]

and write
\[ c^T x - v^T (Ax) = c^T x - (v^T A)x = \langle x, c \rangle_{\mathbb{R}^n} - \langle x, A^T v \rangle_{\mathbb{R}^n} = \langle x, c - A^T v \rangle_{\mathbb{R}^n} \]

which simplifies the Lagrangian to
\[ \sup_{y \geq 0} \left\{ v^T b + \inf_{x \in \mathbb{R}^n} \{ (c - A^T v)^T x \} \right\}. \]

Taking the lead from the finite dimensional case we extend the idea of the transpose \( A^T v \) to the adjoint \( A'(\psi) \) and write
\[ \sup_{\psi \in P_Y^+} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \{ \langle x, \phi \rangle_X - \langle A(x), \psi \rangle_Y \} \right\} = \sup_{\psi \in P_Y^+} \left\{ \langle b, \psi \rangle + \inf_{x \in X} \{ \langle x, \phi - A'(\psi) \rangle_X \} \right\} \]

Finally we observe that we need to **project out the x variables**, otherwise the infimum over \( x \in X \) becomes \(-\infty\). That is we find a \( \psi \in P_Y^+ \) such that \( \langle x, \phi - A'(\psi) \rangle_X = 0 \) and write the Lagrangian dual as
\[ \sup \langle b, \psi \rangle \quad \text{s.t. } \phi - A'(\psi) = \theta_X, \quad \psi \in P_Y^+. \quad \text{(GLPD)} \]

### 8 Talking Points

1. Weak duality as derived in this note always holds! How do Ghate and Smith come up with an example that violates weak duality? What went wrong?

2. In general, the algebraic dual is **very difficult to characterize** in infinite dimensions. Getting a handle on the linear functionals \( \psi \) is challenging. It is common to form a dual based on subspaces of the algebraic dual. For example, in the case of countably infinite programs, one might take the dual variable space to be \( \ell_1 \). What effect might this have on the duality gap?
3. *Characterizing the adjoint mapping is also nontrivial.* Let’s say $X = \mathbb{R}^N$, the space of real sequences, the constraint space is also $Y = \mathbb{R}^N$, and $A : X \rightarrow Y$ is a linear map represented by matrix $A$. Is the adjoint $A^\top$?

4. One might also want to restrict the constraint space. Let’s say the constraint space is $Y = \mathbb{R}^N$. What is the effect on the duality gap of restricting $Y$ to say, an $\ell_p$ space?

5. It is absolutely critical that you are precise about:

   i) the vector space that defines the primal variables
   
   ii) the vector space that defines the primal constraint space
   
   iii) the vector space that defines the dual variables
   
   iv) the vector space that defines the dual constraint space