Linear Programming

Larry Blume
Cornell University, IHS Vienna

These notes derive basic results in finite-dimensional linear programming using tools of convex analysis. Most sources prove these results by exploiting properties of the simplex analysis. That works, but it never did anything for me. In this case I get more from a geometric approach than from the algorithmic approach; thus these notes. One virtue of this method is that one can imagine how some of this will carry over to infinite-dimensional spaces, which is important for labor-macro people. See Gretsky, Ostroy and Zame (1992) "The Non-atomic assignment model," *Economic Theory* 2: 103-127, and their (2002) "Subdifferentiability and the duality gap," *Positivity* 6: 261-274. These notes contain four appendices. Appendix 1 finishes off a detail in the proof of Farkas lemma which is most conveniently handled once basic solutions to inequality systems have been introduced. Appendix 3 contains the proof of Theorem 11, part of the story of dual variables as shadow prices. That proof requires some deeper facts about polyhedral sets and cones which are developed in Appendix 2. Appendix 4 extends some of the ideas of these notes to convex programming, briefly reviewing Lagrangeans, saddle points and the like.

Part I: Theory

1 Convex Facts

Linear programs are the ur-example of convex optimization problems. The intuition derived from finite dimensional linear programs is useful for the study of infinite dimensional linear programs, finite and infinite dimensional convex optimization problems, and beyond. Duality theory, whose geometric intuition is apparent here, is important throughout the analysis of optimization problems. We need just a few key facts from convex analysis to get underway.

1.1 Convex Sets

Definition 1. *A subset C of a vector space is* convex *if it contains the line segment connecting any two of its elements.*

Most of the special properties of convex sets derive from the separating hyperplane theorem:

Theorem 1. *If C is a closed and convex subset of $\mathbb{R}^n$ and $x \notin C$, then there is a $p \neq 0$ in $\mathbb{R}^n$ such that $p \cdot x > \sup_{y \in C} p \cdot y$.*

The geometry of this theorem is that between the point and the set lies a hyperplane — a line in $\mathbf{R}^2$, a plane in $\mathbf{R}^3$ and so forth. There are many different statements of this theorem, having to do with, for instance, the separation of two convex sets, a distinction between strict and strong separation, and so on. See the references for statements and proofs.

The dual description of a closed convex set is an immediate consequence of the separation theorem. A closed half-space H is a set containing all vectors lying on one side of a hyperplane; that is, for some non-zero vector p and scalar α, $H = \{x : p{\cdot}x \geq \alpha\}$, or, in the other direction, the other side, $H = \{x : p \cdot x \leq \alpha\}$). The *primal* description of a convex set is the list of all vectors it contains. The *dual* description of a convex set is the list of all closed half spaces containing it. This duality is established by the following theorem:

Theorem 2. *Every closed convex set in $\mathbb{R}^n$ is the intersection of the closed half-spaces containing it.*

Proof. For a closed convex set C and each $x \notin C$ there is a vector $p_x \neq 0$ and a scalar α_x such that $p_x \cdot x > \alpha_x$ and $p_x \cdot y \leq \alpha_x$ for all $y \in C$. Define the closed half-space $H_x = \{y : p_x \cdot y \leq \alpha_x\}$. The set C is a subset of every H_x, and every x not in C is also not in the corresponding H_x. Thus $\cap_{x \notin C} H_x = C$. □

Polyhedral convex sets are a special class of convex sets: Those defined by finite numbers of linear inequalities: Given an $m \times n$ matrix A and $b \in \mathbb{R}^m$, $P = \{x : Ax \geq b\}$ is a (closed) convex polyhedron in $\mathbb{R}^n$. Constraint sets in lp's are examples of closed convex polyhedra. (Notice that the defining characteristic of polyhedral convex sets is a property of their dual description.) This special structure can be exploited to provide sharp dual characterizations of given convex sets.

Another special class of convex sets are the *convex cones.*

Definition 2. *A set Cis a* cone *if there is a vector c such that for all $y \in C$ the ray $r_c(y) = \{(1-\lambda)c + \lambda y : \lambda > 0\}$ is contained in C.*

To make sense of this definition, rewrite each ray as $c + \lambda(y - c)$ for all $\lambda > 0$ to see that $r_c(y)$ is a straight line running from c through y and ever onward.

One more definition:

Definition 3. *A cone C is* generated *by the set of vectors S if $C = \cup_{s \in S} r_0(s)$.*

A useful fact that will be demonstratedin Appendix 2 is that for a matrix A the solution set to the inequality system $Ax \geq b$ is a closed convex cone.

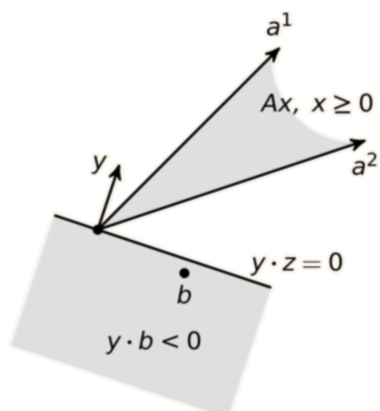
With the tools at hand we can prove theorems of a class which are central to the study of inequality systems. An important example is *Farkas' Lemma*, which describes the relationship between the solution sets to two linear equation/inequality systems.

Farkas' Lemma. *One and only one of the following alternatives is true:*

1. *The system $Ax = b$, $x \geq 0$ has a solution;*
2. *The system $yA \geq 0$, $y \cdot b < 0$ has a solution.*

Results like this are called "theorems of the alternative." It is worth noting that the first theorem of the alternative was proved by Gauss, and is known in some circles as the Fundamental Theorem of Linear Algebra. See Appendix 2 for more on this.

Farkas' Lemma is about systems of linear inequalities, but it is an algebraic expression of the following geometric statement: If b is not in the cone C generated by the columns of A, then there is a hyperplane through the origin that separates the cone from b. That is, there is a y such that $yb < 0$ and $yc \geq 0$ for all $c \in C$. Ask yourself: How does this give us the second alternative?



Proof of Farkas' Lemma. A quick calculation shows the two alternatives cannot both be true. If $yA \geq 0$ and $x \geq 0$, then $yAx \geq 0$. If $Ax = b$, then $yb \geq 0$, contra 2.

To prove that one alternative must hold, suppose the first fails, that is, $b \notin C$. We assert that C is closed. This needs to be proven, which we will do in Appendix 2.

Since C is closed the separation theorem guarantees the existence of a vector y such that $yb \leq \inf_{c \in C} yc$. Since $0 \in C$, $yb < 0$.

To see that yA is non-negative, suppose instead that the first component, $y \cdot A^1$ (the inner product of y with the first column of A) is negative. For any $x_1 > 0$, $yA(x_1, 0, \ldots, 0) = (y \cdot A^1)x_1 < 0$. By making x_1 sufficiently large, we see that $yA(x_1, 0, \ldots, 0) < b$, and yet $A(x_1, 0, \ldots 0) \in C$. This contradiction concludes the proof of the second alternative, and therefore of the lemma. □

1.2 Concave and Convex Functions

Concave and convex functions can be defined in terms of convex sets.

Definition 4. *The* subgraph *of a real-valued function on a vector space V is the set* sub f =

$\{(x, y) \in V \times \mathbb{R} : f(x) \geq y\}$. *The* supergraph *of a real-valued function on a vector space V is the set* $\sup f = \{(x, y) \in V \times \mathbb{R} : f(x) \leq y\}$.

These sets are often referred to as the *hypergraph* and *epigraph*, respectively.

These sets define concave and convex functions. Compare the following definitions to other more usual definitions.

Definition 5. *A real-valued function on a vector space V is* concave *if* $\operatorname{sub} f$ *is convex, and* convex *if* $\operatorname{super} f$ *is convex.*

Closed epi- and hypergraphs will be important for the applications they allow of separating hyperplane theorems. These properties connect to the continuity properties of convex and concave functions. Briefly, and without proof:

Definition 6. *A function $f : X \to \mathbb{R}$ from a metric space X to real numbers is* upper semi-continuous *at $x_0 \in X$ if* $\limsup_{x \to x_0} f(x) \leq f(x_0)$. *It is* lower semi-continuous *at x_0 iff* $\liminf_{x \to x_0} f(x) \geq f(x_0)$. *It is upper (lower) semi-continuous if it is upper (lower) semi-continuous at all x. Equivalently, the function f is* upper semi-continuous *for all x if for all real α, the set $\{x : f(x) \geq \alpha\}$ is closed. It is* lower semi-continuous *if for all α the set $\{x : f(x) \leq \alpha\}$ is closed.*

Clearly a function is both upper and lower semi-continuous at x_0 if and only if it is continuous at x_0. Upper semi-continuous functions can jump up but not down, and lower semi-continuous functions can jump down but not up. Readers should consider which semi-continuity property is important for, e.g., the maximum theorem. The following theorem relates these properties to sub- and supergraphs.

Theorem 3. *A function $f : X \to \mathbb{R}$ is upper semi-continuous throughout its domain iff its subgraph is closed. It is lower semi-continuous iff its supergraph is closed.*

It is often useful to think of concave and convex functions as taking values in the *extended* real numbers. In this case, for concave f, $\operatorname{dom} f = \{x : f(x) > -\infty\}$; and for convex g, $\operatorname{dom} g = \{x : g(x) < +\infty\}$. This will be important for linear programs, for describing the values of unbounded and infeasible optimization problems.

Definition 7. *A concave function $f : X \to [-\infty, +\infty]$ is* proper *if for all $x \in X$ $f(x) < +\infty$ and for some $x \in \mathbb{R}$ $f(x) > -\infty$. A convex function is proper if $-f$ (which is concave) is proper.*

Geometrically speaking, a concave (convex) function is proper if its subgraph (supergraph) is non-empty and contains no vertical line.

2 Introduction

The general linear program is a constrained optimization problem where objectives and constraints are all described by linear functions:

$$\begin{aligned} v_P(b) &= \max c \cdot x \\ \text{s.t. } Ax &\le b \\ x &\ge 0 \end{aligned} \tag{1}$$

where $c, x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and the matrix A is $m \times n$; m constraints in n variables. Problem (1) is called the *primal problem*, and the description in (1) is its *canonical form*. If we wanted the constraint equation $a \cdot x \ge b'$, the matrix A would have the row $-a$, and the vector b would contain corresponding coefficient $-b'$. Having both constraints $a \cdot x \le b'$ and $-a \cdot x \le -b'$ effectively imposes the equality constraint $a \cdot x = b'$. More clever, maximizing $c \cdot x - c \cdot y$ subject to $Ax - Ay \le b$ with $x, y \ge 0$ is a rewrite of the problem $\max c \cdot z$ s.t. $Az \le b$ with no positivity constraints on z. Really cleverly, the canonical form can handle absolute values of variables in both the objective function and in the constraints; exercise left to the reader. So the canonical form (and the standard form to follow) are quite expressive.

The *standard form* of a linear program is

$$\begin{aligned} v_P(b') &= \max c' \cdot x' \\ \text{s.t. } A'x' &= b' \\ x' &\ge 0 \end{aligned} \tag{2}$$

which uses only equality and non-negativity constraints.

We have already seen that a given standard-form problem, say, (2), can be rewritten in canonical form. A given problem, say, (1) in standard form, is rewritten in standard form by using *slack variables*, here variables z:

$$\begin{aligned} v_P(b) &= \max c \cdot x \\ \text{s.t. } Ax + Iz &= b \\ x &\geq 0 \\ z &\geq 0 \end{aligned} \tag{3}$$

where I is $m \times m$ and $z \in \mathbb{R}^m$. The matrix $[A\ I]$ is called the *augmented matrix* for the canonical-form problem (1).

When a canonical-form program with matrix A is put into standard form with matrix A', A' has full row rank. Without loss of generality we can assume that in *any* standard-form problem, $n \geq m$ and that A' has full row rank. To see this, start with a $k \times l$ matrix A'. Rewriting in the canonical form gives a matrix of size $2k \times l$. Rewriting the canonical in standard form using slack variables gives a matrix of size $2k \times (2k + l)$. The additional $2k$ columns correspond to slack variables, and so the submatrix with those columns is the $2k \times 2k$ identity matrix. We have thus rewritten the constraint set using a matrix with full row rank. It may seem unnatural, but it proves the point. The assumption of full row rank for standard-form problems will be maintained throughout.

We use the following terminology: The *objective function* for (1) is $f(x) = c \cdot x$. The *constraint set* is $C = \{x \in \mathbb{R}^n : Ax \leq b \text{ and } x \geq 0\}$ for programs in canonical form, and $C' = \{x \in \mathbb{R}^n : A'x = b \text{ and } x \geq 0\}$ for programs in standard form. The set C is a convex polyhedron and the set C', even more, is the intersection of an affine subspace with the non-negative orthant. A *solution* for (1) is any $x \in \mathbb{R}^n$. A *feasible solution* is a element of C (or C'). An *optimal solution* is a feasible solution at which the supremum of the objective function on C (or C') is attained. This unnatural use of the word solution to refer to any vector in the domain of the objective function dates from the earliest years of linear programming; history has locked it in.

A linear program need have no feasible solutions, and a program with feasible solutions need have no optimal solutions.

3 The Geometry of Linear Programming

The canonical-form constraint set C is a *polyhedral convex set*. In figure 1, the four corners are *vertices* of C. This constraint set is bounded, so any linear program with constraint set C will have a solution.

$$\begin{bmatrix} -3 & 1 \\ 1 & 1 \\ 1 & -2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \le \begin{bmatrix} 0 \\ 6 \\ 0 \\ -1 \end{bmatrix}$$

$$x, y \ge 0$$

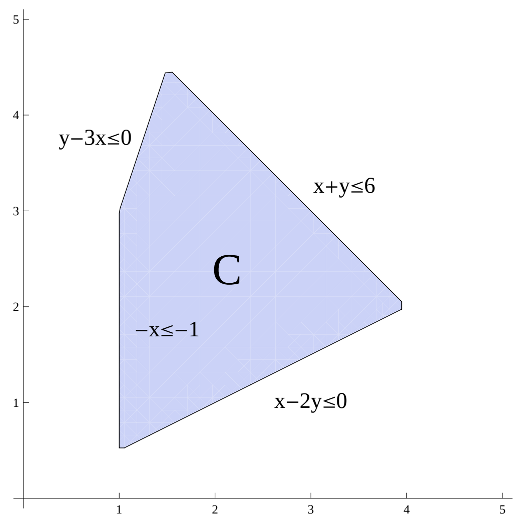


Figure 1: The set C.

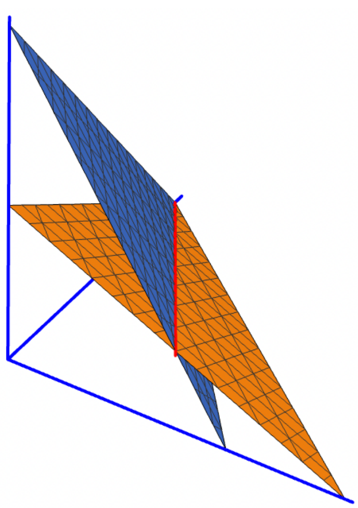


Figure 2: The Canonical Form

In this section we refer to programs in the standard form (2). Thus $C = \{x : Ax = b,\ x \ge 0\}$. The set of x satisfying the constraints $Ax = b$ is the intersection of a number of hyperplanes. These "lower dimensional planes" are called affine spaces; they are translates away from the origin of vector subspaces. Thus a standard-form constraint set is the intersection of an affine space with the non-negative orthant. Vertices will necessarily be at the orthant boundaries. In Figure 2 the intersection of two planes with the non-negative orthant are the blue and orange triangles. The constraint set C is their intersection, the red line. The vertices are the endpoints of the red line, which are in the xz and yz planes.

Definition 8. *x is a* vertex *of the polyhedron C iff there is no $y \neq 0$ such that $x + y$ and $x - y$ are both in C.*

The *vertex theorem* describes the relation between the geometry of C and solutions to (2).

Theorem 4 (Vertex Theorem)**.** *For a linear program in standard form with feasible solutions,*

1. *A vertex exists.*
2. *If $v_P(b) < \infty$ and $x \in C$, then there is a vertex x' such that $c \cdot x' \geq c \cdot x$.*

Other ways of writing a given linear program do not always have vertices. For instance, the program where $C = \{(x_1, x_2) : 0 \leq x_1 \leq 1\}$ has no vertex.

Problem 1. *Consider a linear program of the form* $\max ax + by$ *where $(x, y) \in C$, the polyhedron of Figure 1. What are the canonical and standard forms of this program? What are their vertices? Verify the conclusions of the vertex theorem for $b >=< 0$.*

Proof of Theorem 4. The proof of 1) constructs a vertex. Choose $x \in C$. If x is a vertex, we are done. If x is not a vertex, then for some $y \neq 0$, both of $x \pm y \in C$. Thus $Ay = 0$, and if $x_j = 0$ then $y_j = 0$. Let $\lambda^* \geq 0$ solve $\sup\{\lambda : x \pm \lambda y \in C\}$. Since x is not a vertex, $\lambda^* \neq 0$, and since C is closed, $x \pm \lambda^* y \in C$. One of $x \pm \lambda^* y$ has more 0's than does x. Suppose without loss of generality it is $x + \lambda^* y$. If $x + \lambda^* y$ is a vertex, we are done. If not, repeat the argument. The argument can be iterated at most n times before $x + \lambda^* y = 0$, that is, before the origin is reached, and the origin is a vertex of any standard form program for which it is feasible.

The proof of 2) has the same idea. If x is a vertex, take $x' = x$. If not, then there is a $y \neq 0$ such that $x \pm y \in C$, and this y has the properties given above. If $cy \neq 0$, without loss of generality, take y such that $cy > 0$.

If $cy = 0$, choose y such that for some j, $y_j < 0$. (Since $y \neq 0$, one of y and $-y$ must have a negative coefficient.)

Now consider $x + \lambda y$ with $\lambda > 0$. Then

$$c(x + \lambda y) = cx + \lambda cy \geq cx.$$

It cannot be the case that $y_j \geq 0$ for all j. If so, then by construction $cy > 0$, and $x + \lambda y$ is in C for all $\lambda \geq 0$. Since $cy > 0$, $c(x + \lambda y) \to +\infty$ as $\lambda \to +\infty$, so $v_P(b) = +\infty$, the problem is unbounded.

Therefore there is a j such that $y_j < 0$. For large enough λ, $x + \lambda y \not\geq 0$. Let λ^* denote the largest λ such that $x + \lambda y$ is feasible. Then $x + \lambda^* y$ is in C, has at least one more zero, and $c(x + \lambda^* y) \geq cx$.

Now repeat the argument. If the problem is bounded, case 2 happens at most n times before a vertex is reached. $\square$

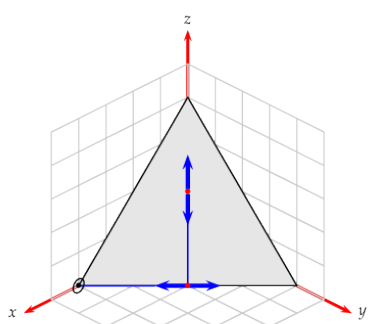


Figure 3: Finding a vertex.

The next problem is to find an algebraic characterization of vertices in terms of the data A and b of the standard-form lp problem.

Definition 9. *A feasible solution to a linear programming problem in standard form is a* basic solution *if and only if the columns A^j of A such that $x_j > 0$ are linearly independent; that is, if the submatrix A_x of A consisting of the columns A^j for which $x_j > 0$ has full column rank.*

The set $\{j : x_j > 0\}$ is called the support of x.

Theorem 5. *A solution x is basic if and only if it is a vertex.*

Proof. First, suppose that x is not a vertex. Then from the arguments of the proof of Theorem 4 we know that for some $y \neq 0$, $Ay = 0$ and if $x_j = 0$, then $y_j = 0$. The second statement implies that Ay is a linear combination of the columns of A_x, and the first says that combination is 0, so the columns A^j of A_x are linearly dependent. Thus x is not a basic solution.

Next, suppose that x is not basic. Then the columns A^j are linearly dependent, and so there is a y such that if $x_j = 0$ then $y_j = 0$ — in other words, y describes a linear combination of the A^j — such that $Ay = 0$. For λ such that $|\lambda|$ is sufficiently small, $x \pm \lambda y \geq 0$. For all these λ, $x \pm \lambda y \in C$, and so x is not a vertex. □

A consequence of this theorem is that our constraint sets can have only a finite number of vertices.

Corollary 6. *The set of feasible solutions has only a finite number of vertices.*

Proof. There are only a finite number of linearly independent column vectors of A. If for some submatrix A' of independent columns, $A'x = b$ and $A'y = b$ and $x \neq y$, then $A'(x - y) = 0$, contradicting the independence of the columns. Thus each independent set is associated with at most one vertex. □

Corresponding to every linearly independent set of column vectors is at most one basic solution, because if x and x' are two distinct sets of solutions employing only the columns of x, then $A(x - x') = 0$, and so the the columns of x are dependent. A maximal linearly independent set contains at most m vectors, and so there are at most $\binom{n}{m}$ maximal linearly independent sets, so this bounds from above the number of vertices.

Together, theorems 4 and 5 say three things. First, the constraint set C for a canonical-form program has vertices. Second, vertices correspond to basic solutions of the linear inequalities defining the constraints. Third, if the linear program has a feasible/optimal solution, it has a basic (vertex) feasible/optimal solution. This last fact is important, and is known as the *Fundamental Theorem of Linear programming.*

Theorem 7 (Fundamental Theorem of Linear Programming)**.** *If the linear program* (2) *has a feasible solution, then it has a basic feasible solution. If it has an optimal solution, then it has a basic optimal solution.*

This theorem already provides a (really inefficient) algorithm for finding an optimal solution on the assumption that a feasible solution exists. If an optimal solution exists, then a basic optimal solution exists, and there are only a finite number of basic solutions to check.

As an aside, one might wonder about the relationship of vertices of the standard form to vertices of the canonical form.

Theorem 8. *The feasible solution x to the standard problem* (1) *is a vertex of the canonical form feasible set if and only if there is a z such that* (x, z) *is vertex of the constraint set for the corresponding standard form problem* (3).

Proof. Let A denote the matrix for the canonical form constraint set, and $[A\, I]$ the constraint matrix for the corresponding standard form problem. First, suppose that (x, z) is a vertex for problem (3). Then it is a basic feasible solution, and so there is a square submatrix $[A'\ I']$ of the matrix $[A\ I]$ such that

$$[A'\ I'] \begin{bmatrix} x' \\ z' \end{bmatrix} = b,$$

$[A'\ I']$ is invertible and (x', z') is the non-zero part of the (x, z).

If y and w are directions such that $x \pm y$ is feasible for (1), and $(x, z) \pm (y, w)$ is feasible for (3), then y only perturbs x' and w only perturbs z', because otherwise one direction would violate a non-negativity constraint. So it suffices to pay attention only the matrices A' and I', the vectors to x', z', and their perturbations y' and w'.

Suppose that x is not a vertex for the problem (1). Then there is a y' such that $A'(x' \pm y') \leq b$ and $x' \pm y' \geq 0$. Let $w' = -Ay'$. Then

$$[A'\ I'] \begin{bmatrix} x' \pm y' \\ z' \pm w' \end{bmatrix} = b,$$

Since (x, z) is a vertex for (3), one of $(x', z') \pm (y', w')$ must not be non-negative. We assumed that $x \pm y$ are both feasible for (1), and hence non-negative, and so one of $(z' \pm w')$ is not non-negative. Without loss of generality, assume that

$$z' + w' = z' - Ay' \ngeq 0.$$

Thus for some i, $z'_i - (Ay')_i < 0$. For this i, $(Ay')_i > z'_i$, and so $(Ax')_i + (Ay')_i > (Ax')_i + z'_i = b_i$, contradicting the feasibility of $x + y$. This contradiction establishes that x is a vertex for the problem (1).

Now suppose that x is a vertex for (1), and define z such that $Ax + z = b$. We will show that (x, z) is a vertex of (3). Again, suppose not. Suppose there are (y, w) such that $(x, z) \pm (z, w)$ are feasible for problem (3). Then $z \pm w$ are both non-negative, and so $A(x \pm y) \leq b$. Since x is a vertex, $x \pm y$ cannot be feasible for (1), and so $x \pm y$ are not both non-negative. Hence contrary to assumption, $(x, z) \pm (z, w)$ are not both feasible for (3). □

A consequence of theorem 8 and the subsequent definition of a basic solution for the canonical form is that theorem 7 also holds for problems posed in the canonical form.

4 Duality

The *dual program* for problem (1) in canonical form is

$$\begin{aligned} v_D(c) &= \min y \cdot b \\ \text{s.t. } yA &\geq c \\ y &\geq 0 \end{aligned} \tag{4}$$

Appendix 4 provides motivation for the dual program by showing that both the primal and dual programs have the same Lagrangean, and that the pairs (x^*, y^*) of a solution to the primal program and a solution to the dual program are the Lagrangean's saddle points. All the results of linear programming could be developed as an application of Lagrangean duality, but this would be a backwards way to proceed. Furthermore, the direct path exposes the geometry of the problem and leads to geometric insights that motivate Lagrangean duality for more general problems.

Problem 2. *Suppose, more generally than in* (4) *that we have a primal with some inequality and some equality constraints:*

$$\begin{aligned} v_P(b, b') &= \max c \cdot x \\ \text{s.t. } Ax &\leq b \\ A'x &= b' \\ x &\geq 0 \end{aligned} \tag{5}$$

Prove that the dual can be expressed:

$$\begin{aligned} v_D(c) &= \min y \cdot b + z \cdot b' \\ \text{s.t. } yA + zA' &\geq c \\ y &\geq 0 \end{aligned} \tag{6}$$

That is, it is like the dual for the canonical form with only inequality constraints, but with no sign constraints on the dual variables corresponding to the equality constraints.

The first, and easy observation, is the weak duality theorem.

Theorem 9 (Weak Duality)**.** *For problems* (1) *and* (4)*,* $v_P(b) \leq v_D(c)$.

Proof. Write the problem in the canonical form. For feasible solutions x and y for the primal and dual, respectively, $(yA - c) \cdot x \geq 0$, and $y \cdot (b - Ax) \geq 0$, so for all feasible primal solutions x and dual solutions y, $c \cdot x \leq y \cdot b$. □

Notice that any feasible y for the dual bounds from above the value of the primal, so if the primal is unbounded, then the feasible set for the dual must be empty. Similarly, any feasible solution for the primal bounds from below the value of the dual, so if the dual is unbounded then the feasible set for the primal problem must be empty. This is part of the Duality Theorem.

Theorem 10 (Duality Theorem)**.** *For the primal program* (1) *and the dual program* (4)*, exactly one of the following three alternatives must hold:*

1. *Both are feasible, both have optimal solutions, and their optimal values coincide.*
2. *One is unbounded and the other is infeasible.*
3. *Neither is feasible.*

Proof. We can prove this theorem by making assertions about the primal and seeing the consequences for the dual, and noting that corresponding statements hold for parallel assertions about the dual and their consequences for the primal.

There are three cases to consider: The primal program has feasible solutions and a finite value, the primal program has feasible solutions but an infinite value, and the primal program has no feasible solutions.

Suppose first that the primal program has feasible solutions and a finite value. We rewrite the problem in standard form, and we denote its constraint matrix too by A. Then $-\infty < v_P(b) < +\infty$, where $v_P(b)$ is the least upper bound of the set of feasible values for the primal. First we show that the primal has an optimal solution.

Let $\{x_n\}$ denote a sequence of feasible solutions such that $c \cdot x_n \uparrow v_P(b)$. The fundamental theorem implies that there is a sequence $\{x'_n\}$ of basic feasible solutions such that for all n $c \cdot x'_n \geq c \cdot x_n$. Thus $c \cdot x'_n$ converges to $v_P(b)$.

There are, however, only a finite number of basic feasible solutions (corollary 6), and so at least one solution x' occurs infinitely often in the sequence. Thus $c \cdot x' = v_P(b)$, and x' is optimal for the primal. If the value of the primal program is bounded, weak duality bounds the value of the dual, and a parallel argument shows that the dual program has a solution.

Next we show that the values of the primal and dual programs coincide. Weak duality says that the value of the dual is at least that of the primal. To prove equality, we show the existence of feasible solutions to the dual program that approximate the value of the primal program arbitrarily closely. We use Farkas lemma to prove this. Consider the following linear inequality system:

$$\begin{bmatrix} A \\ c \end{bmatrix} x = \begin{bmatrix} b \\ v_P(b) + \epsilon \end{bmatrix},$$
$$x \geq 0.$$

This system has a solution for $\epsilon = 0$, namely, an optimal solution to problem (1), but no solutions for $\epsilon > 0$. Accordingly, Farkas lemma says that for any $\epsilon > 0$, the following alternative system has a solution (y, α):

$$\begin{bmatrix} y & \alpha \end{bmatrix} \begin{bmatrix} A \\ c \end{bmatrix} \geq 0 \tag{7}$$

$$\begin{bmatrix} y & \alpha \end{bmatrix} \begin{bmatrix} b \\ v_P(b) + \epsilon \end{bmatrix} < 0 \tag{8}$$

If $\alpha = 0$, then $(y, 0)$ also solves the alternative for $\epsilon = 0$, a contradiction. If $\alpha > 0$, then $y \cdot b + \alpha(v_P(b) + \epsilon) < 0$. This will also hold for $\epsilon = 0$, a contradiction. So $\alpha < 0$. Since the system is homogeneous, there is a y such that $(y, -1)$ solves the system. Then (7) implies that y is feasible for the dual (4), and (8) implies that $y \cdot b < v_P(b) + \epsilon$. This and weak duality imply that $v_P(b) \leq v_D(c) < v_P(b) + \epsilon$ for all $\epsilon > 0$, and so $v_P(b) = v_D(c)$.

The weak duality theorem and the comment following its proof dispenses with the second case, that of an unbounded primal (dual) program. Any feasible solution to the dual (primal) program bounds the value of the primal (dual) program from above (below). So if the value of the primal (dual) program is $+\infty$ $(-\infty)$, then the dual (primal) program can have no feasible solutions.

We demonstrate the last alternative of the theorem by example. Consider the program

$$\begin{aligned} v_P(b) &= \max(1, 1) \cdot (x_1, x_2) \\ \text{s.t.} & \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \\ & x \geq 0. \end{aligned}$$

The feasible set for the dual program are those (y_1, y_2) that satisfy the inequalities

$$(y_1, y_2) \cdot \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \geq \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$y \geq 0.$$

Neither inequality system has a solution. □

5 Complementary Slackness

We remain in the canonical form. Let A_i and A^j denote the ith row and jth column of A.

Theorem 11 (Complementary Slackness)**.** *Suppose that x^* and y^* are feasible for the primal and dual problems, respectively. Then they are optimal for their respective problems if and only if for each constraint i in the primal program,*

$$y_i^*(b_i - A_i \cdot x^*) = 0$$

and similarly for each constraint j of the dual program,

$$(y^* \cdot A^j - c)x_j^* = 0.$$

In each case, if the constraint is slack then the corresponding dual variable is 0. Complementary slackness hints at the interpretation of dual variables as shadow prices. In this case, when a constraint is not binding, there is no gain (loss) to slackening (tightening) it.

Proof. Suppose that x^* and y^* are feasible solutions that satisfy the complementary slackness conditions. Then $y^* \cdot b = y^* \cdot Ax^*$ and $y^*A \cdot x^* = c \cdot x^*$, so $y^* \cdot b = c \cdot x^*$, and optimality follows from the duality theorem.

If x^* and y^* are optimal, then since $Ax^* \leq b$ and y^* is non-negative, $y^*Ax^* \leq y^* \cdot b$. Similarly, $cx^* \leq y^*Ax^*$. The duality theorem has $c \cdot x^* = y^* \cdot b$, so both inequalities are tight. □

6 Dual Programs and Shadow Prices

The role of dual variables in measuring the sensitivity of optimal values to perturbations of the constraints is an important theme in optimization theory. A familiar example is the equality of the Lagrange multiplier for the budget constraint in a utility optimization problem and the marginal utility of income. This fact is typically explained by the envelope theorem and the first-order conditions for utility maximization. Such results, however, are understood most deeply as consequences of convexity assumptions.

Consider again the primal program

$$
\begin{aligned}
\mathcal{P}(b): \quad v_P(b) &= \max c \cdot x \\
\text{s.t. } Ax &\leq b \\
x &\geq 0
\end{aligned}
\tag{1}
$$

and its dual program

$$
\begin{aligned}
\mathcal{D}(c): \quad v_D(c) &= \min y \cdot b \\
\text{s.t. } yA &\geq c \\
y &\geq 0
\end{aligned}
\tag{4}
$$

where we now emphasize the dependence of the problems on the constraints. The goal of this section is to describe the value functions $v_P(b)$ and $v_D(c)$.

The domains $\operatorname{dom} v_P$ and $\operatorname{dom} v_D$ are the sets of vectors $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ for which the feasible sets are not empty. They are closed convex cones.[1]

Theorem 12. *$v_P(b)$ is concave, and $v_D(c)$ is convex. Their (respective) sub- and supergraphs are closed.*

We prove the first claim. The proof of the second is similar.

[1] Recall that by definition, the value $v_P(b) > -\infty$ if and only if $b \in \operatorname{dom} v_P$, and $v_D(c) < +\infty$ if and only if $c \in \operatorname{dom} v_C$.

Proof. The value function is clearly increasing and linearly homogeneous. If (b', v') and (b'', v'') are in the subgraph of v_P, then b' and b'' are in $\operatorname{dom} v_P$, and there are feasible x' and x'' such that $v' \leq c \cdot x'$ and $v'' \leq c \cdot x''$. For any $0 \leq \lambda \leq 1$, $\lambda x' + (1 - \lambda)x''$ is feasible for $\lambda b' + (1 - \lambda)b''$, so the λ-combination of (b', v') and (b'', v'') is also in the subgraph.

Closure of the hypograph of v_P and the epigraph of v_D follows from the maximum theorem. The maximum theorem applied to b in $\operatorname{dom} v_P$ asserts that v_P is continuous in its domain. (Why? Think consumer theory.) Let $(b_n, v_n) \in \operatorname{sub} v_P$ converge to (b, v) with $b \in \operatorname{dom} v_P$. Then $v_n \leq v_P(b_n)$ and $v_P(b_n)$ converges to $v_P(b)$, so $v \leq v_P(b)$. Conclude that $(b, v) \in \operatorname{sub} v_P(b)$. The same argument applies to v_D. □

The sets $\operatorname{dom} v_P$ and $\operatorname{dom} v_D$ are the values b and c, respectively, for which the problems have feasible solutions. Notice that the preceding proofs did not depend in any way on finiteness of the value functions. The next theorem states that, for a given objective function $c \cdot x$, either the problem is bounded for all constraint vectors b or unbounded for all b.

Theorem 13. *Either $v_P(b) < +\infty$ for all b or $v_P(b) = \infty$ for all b. Similarly, either $v_D(c) > -\infty$ for all c or $v_D(c) = -\infty$ for all c.*

The proof is in an appendix. The implication of this is that either programs with a given constraint matrix A are unbounded for all possible constraint values, or for none.

One would not expect the value functions of linear programs to be smooth because basic solutions bounce from vertex to vertex as the constraint set changes. (This suggests piecewise linearity.) But since the relevant functions are concave and convex, super- and subdifferentials will be the appropriate notions of derivatives. Just for review, here are the definitions.

Definition 10. *The* superdifferential *of a concave function f on $\mathbb{R}^n$ at $x \in \operatorname{dom} f$ is the set $\partial f(x) = \{s : f(y) \leq f(x) + s \cdot (y - x) \quad \text{for all } y \in \mathbb{R}^n\}$. The* subdifferential *of a convex function g on $\mathbb{R}^n$ at $x \in \operatorname{dom} g$ is the set $\partial g(x) = \{t : g(y) \geq g(x) + t \cdot (y - x) \quad \text{for all } y \in \mathbb{R}^n\}$. Each $s \in \partial f(x)$ ($t \in \partial g(x)$) is a* supergradient (subgradient) *of f (g).*

The inequalities in the definition are called the super- and subgradient inequalities, respectively, for the concave and convex case. It is worth recalling that if a concave function is differentiable at a point x, then the graph of the function lies everywhere on or below the tangent line through $\big(x, f(x)\big)$. So the gradient describes a supporting hyperplane to the subgraph at this point. The supergradient just generalizes this concept to all concave functions. (And similarly for subgradients of convex functions.)

The main result is that solutions to the dual program are shadow prices for the primal constraints, and vice versa.

Theorem 14. *If either $v_P(b)$ or $v_D(c)$ is finite, then*

1. *$v_P(b) = v_D(c)$,*
2. *programs $\mathcal{P}(b)$ and $\mathcal{D}(c)$ have optimal solutions x^* and y^*, respectively, and*
3. *$\partial v_P(b)$ is the set of optimal solutions to the dual program, and $\partial v_D(c)$ is the set of partial solutions to the primal program.*

The only new part is the statement about sub- and superdifferentials. Theorem 13 implies that if $v_P(b)$ is finite for any b in $\operatorname{dom} v_P$, then it will be for all $b \in \operatorname{dom} v_p$, and similarly for v_D on its domain, and so the super- and subdifferentials exist everywhere in the respective domains.

Proof of Theorem 14. We will prove the characterization of solutions to the dual program as supergradients of the primal value function. The other proof proceeds in a parallel manner.

One direction is simple. If y^* solves the dual program to $\mathcal{P}(b)$, then $y^*A \geq c$ and $y^* \cdot b = v_P(b)$. Furthermore, y^* is feasible for any other dual program with objective b', so $y^* \cdot b' \geq v_P(b')$. Thus

$$y^* \cdot b' - v_P(b') \geq 0 = y^* \cdot b - v_P(b),$$

which implies the supergradient inequality.

Next we take up the other direction, proving that any supergradient of the primal value function at b solves the program dual to $\mathcal{P}(b)$. In doing so we first assert that $v_P(b) < +\infty$ for all $b \in \mathbb{R}^m$. It is $-\infty$ off $\operatorname{dom} v_P$, and theorem 13 and the hypothesis of the theorem imply that v_P is finite everywhere on its domain.

The value functions for primal and dual are linear homogeneous as well as concave/ convex. In this case, the super- and subgradient inequalities take on a special form.

Lemma 1. *(i) if $y \in \partial v_P(b)$ then $y \cdot b' \geq v_P(b')$ for all $b' \in \mathbb{R}^n$, with equality at $b' = b$.*
(ii) if $y \in \partial v_D(c)$ then $y \cdot b' \leq v_D(b')$ for all $b' \in \mathbb{R}^n$, with equality at $b' = b$.

Proof. As usual, we only prove (i). And to prove (i) we need only worry about $b \in \operatorname{dom} v_P$ since the value of v_P is $-\infty$ otherwise. Homogeneity of v_P implies that y is a supergradient of v_P at b if and only if it is a supergradient of v_P at αb. (Apply the supergradient inequality to $(1/\alpha)b'$.) Thus for all $\alpha > 0$, the supergradient inequality says

$$\begin{aligned} v_P(b') &\leq v_P(\alpha b) + y \cdot (b' - \alpha b) \\ &= y \cdot b' + \alpha(v_P(b) - y \cdot b), \end{aligned}$$

and so $v_P(b') \leq y \cdot b'$ for all $b' \in \operatorname{dom} v_P$.

The supergradient inequality also has that

$$0 = v_P(0) \leq v_P(b) + y \cdot (0 - b),$$

so in addition, $y \cdot b \leq v_P(b)$. □

Lemma 1 breaks the subgradient inequality into two pieces. The next lemma shows that a vector y is feasible for the dual if and only if it satisfies the first piece. The following lemma shows that y is optimal if and only if, in addition, it satisfies the second piece.

Lemma 2. *For any $y \in \mathbb{R}^m$, if $y \cdot b' \geq v_P(b')$ for all $b' \in \mathbb{R}^m$, then $y \in \mathbb{R}^m$ is feasible for the dual.*

Proof. Suppose that $y \cdot b' \geq v_P(b')$ for all $b' \in \mathbb{R}^m$. For any $b' \geq 0$, $y \cdot b' \geq v_P(b') \geq v_P(0) = 0$, so $y \geq 0$.[2] For any $x \geq 0$, $y \cdot Ax \geq v_P(Ax) \geq c \cdot x$. (The last step follows because $x \geq 0$ is feasible for the problem with constraint $Ax' \leq b$ with $b = Ax$.) Thus $y \cdot A \geq c$. □

[2]Recall that $v_P(b)$ is nondecreasing in b.

If y is a supergradient of $v_P(b)$, then y is feasible for the dual program and $y \cdot b = v_P(b) = v_D(c)$, this last from the duality theorem, and so y is optimal for the dual program. $\square$

Part II: Applications

1 Markov Chains

The existence of a stationary distribution for a finite state Markov chain is also a consequence of Farkas' lemma. Suppose that P is a transition probability matrix for an n-state Markov chain; thus P is $n \times n$ and p_{ij} is the probability of transiting from state i today to state j tomorrow. If the value of the state at time t is described by a probability distribution π_t, then the distribution of the state at time $t+1$ is $\pi_{t+1} = \pi_t P$.

A stationary distribution for a Markov chain is a probability distribution on states that self-replicates; that is, a probability distribution π^* such that $\pi^* P = \pi^*$. The existence of a stationary distribution is guaranteed by Farkas lemma. A stationary distribution solves the equation

$$\pi \begin{pmatrix} P - I & e \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix} \qquad \pi \geq 0$$

where e is the vector of 1s. If no solution to this system exists, then Farkas lemma requires that

$$\begin{pmatrix} P - I & e \end{pmatrix} x \geq 0 \qquad \begin{pmatrix} 0 & 1 \end{pmatrix} x < 0.$$

for some $x \in \mathbb{R}^{n+1}$. Write $x = (x_1, \ldots, x_n, y)$. Then

$$\sum_j p_{ij} x_j - x_k + y \geq 0, \qquad y < 0,$$

from which we get for all i,

$$p_{ij} x_j - x_i > 0.$$

Let $x_k = \max_k x_k$. Then in particular,

$$\sum_j p_{kj} x_j - x_k > 0.$$

The p_{kj} are non-negative and sum over j to 1, so

$$0 = x_k - x_k = \sum_j p_{kj} x_k - x_k \geq \sum_j p_{kj} x_j - x_k > 0$$

which is impossible.

The interesting part of Markov chains comes from the fact that under additional conditions on P, the invariant distribution is unique, and under still further conditions, π_t converges to π^*. One assumes that P is primitive, and then everything follows from the Perron-Frobenius Theorem.

2 Arbitrage

The "no-arbitrage principle" is a fundamental idea in the theory of asset-pricing. The setup is this: There are m states of nature, conditions which determine the payoff of financial assets. There are n assets. Asset j pays off a_{ij} in state i. The *asset returns matrix* is the $m \times n$ matrix A whose j'th column gives the return of asset j in each of the m states. Assets can be combined into *portfolios*. A portfolio is a vector $x \in \mathbb{R}^n$ whose j'th component is the holding of asset j. Components of x can be positive or negative, representing *long* and *short* positions, respectively. The payoff of portfolio x is the vector Ax, whose i'th component is the amount of wealth generated by the portfolio in each state. This construction assumes any trader can buy or sell any asset. We are all I-bankers.

The no-arbitrage principle states that any portfolio that pays off non-negative returns in every state must have a non-negative cost. Suppose $p > 0$ is the vector of asset prices; asset j costs p_j per unit. The cost of portfolio x is px, and the no-arbitrage principle can be restated as the system

$$Ax \geq 0 \qquad px < 0 \tag{9}$$

has no solution. The no-arbitrage principle states that the second alternative in Farkas lemma does not hold, so we conclude that the first must. that is to say, there is a vector $\pi \in \mathbb{R}^m_+$ such that $\pi A = p$. And since $p > 0$, $\pi > 0$; π is semi-positive.

By changing currency units, we can take the sum $\sum_j p_j$ to be any positive number we want. It is conventional, then, to rescale prices so that $\sum_i \pi_i = 1$. So normalized, there are several ways of interpreting π. An Arrow security is a security that pays off 1 unit of return in a given state, and 0 in all other states. If Arrow securities are available, that is if they already exist or if they can be made synthetically by devising portfolios of existing securities that pay off this way, then π_i is said to be the *state price* of the state-i Arrow security. It would be surprising to actually have Arrow securities at hand. The more typical situation is when they can be manufactured. Whether or not Arrow securities for all states can be constructed depends upon the span of the column vectors — the dimension of the column space. If the matrix A has full row rank m, markets are said to be *complete* and the state price vector is uniquely determined by the asset price p.

A second interpretation is to think of π as a probability distribution. In this case the no-arbitrage condition says that the price of any asset is equal to its expected return under distribution p. In this interpretation π is called the *risk-neutral* probability distribution. This idea is the basis for the martingale approach to asset pricing.

A better statement of the no-arbitrage principle is that a portfolio that generates a semi-positive return vector, $Ax > 0$, has a positive price. Parallel to equation (9), this condition is to say that.

$$Ax > 0 \qquad px \leq 0 \tag{10}$$

has no solution. Stiemke's Theorem is a variation on Farkas' Lemma which states that if (10) has no solution, then there is a $\pi \in \mathbb{R}^m$ which solves

$$\pi A = p \qquad \pi \gg 0.$$

The tightening of the no-arbitrage condition leads to a positive state price for every state.[3]

3 The Non-Substitution Theorem

The nonsubstitution theorem describes the production possibility frontier for constant returns to scale economies. Here is a version for polyhedral production technologies. These technologies are described by two $n \times m$ non-negative matrices, A and B, where $m > n$, and a non-negative vector $a \in \mathbb{R}^m_+$; $1 \leq i \leq n$ indexes goods and $1 \leq j \leq m$ indexes production

[3] The usual statement of Stiemke's theorem is that one and only one of $Mx > 0$ and $yM = 0, y \gg 0$ has a solution. Adjoin to A the row $-p$ to get the inequality we do not want to satisfy, and let $\pi = y_{n+1}^{-1}(y_1, \ldots, y_n)$ to get the result.

processes. Column $A_{\cdot j}$ lists the inputs necessary to run activity j at "level 1", column $B_{\cdot j}$ lists the outputs from running process j at "level 1", and a_j is the labor input required to run activity j at "level 1". Input demand and output supply scales linearly in the level of activity usage. If the activities are run at levels described by the activity vector x, then *gross output* is Bx, the input requirement is Ax, and *net output* is $y = (B-A)x$. Labor usage is constrained by a labor endowment L. Net output y is efficient if there is no feasible net output $z > y$, that is, if $(B-A)x = z$, then $ax > L$.

The content of the non-substitution theorem is that given some assumptions, the output possibility frontier is the intersection of a hyperplane with the non-negative orthant. Consequently, except at the boundary relative prices are uniquely determined by the technology — the positive vector orthogonal to the hyperplane.

One key assumption is that there is no joint production.

Axiom 1. *Each $B_{\cdot j}$ column has a single non-zero element.*

With no loss of generality, we can scale each activity vector so that the non-zero element of B is 1. A second assumption is that the technology can actually produce all goods:

Axiom 2. *There is a net output $y \gg 0$ and an activity vector $x \geq 0$ such that $(B-A)x = y$.*

In the literature, technologies satisfying this assumption are called *productive*

A *technology* is a list τ of activities such that the matrix $B_\tau = [B_{\cdot j}]_{j\in\tau}$ is the identity matrix. That is, one column for each commodity. The corresponding input requirements matrix A_τ and labor requirements vector a_τ are $n \times n$ and of dimension n, respectively, and each such collection (A_τ, B_τ, a_τ) describes a *simple Leontief model*.

Efficient production requires minimizing labor input. Labor cost minimization is described by the linear program

$$\begin{aligned} \lambda(y) &= \min a \cdot x \\ \text{s. t. } (B-A)x &\geq 0 \\ x &\geq 0 \end{aligned} \tag{11}$$

Here x is a vector of activity levels, and y is a net output vector. We will also solve the same problem for given technologies τ:

$$\begin{aligned} \lambda_\tau(y) &= \min a_\tau \cdot x \\ \text{s.t. } (B_\tau - A_\tau)x &\geq 0 \\ x &\geq 0 \end{aligned} \tag{12}$$

Let $P(L)$ denote the set of net-outputs that can be produced with labor endowment L. The non-substitution theorem is stated formally thus:

Theorem 15. *If there is no joint production and if the technology is productive, then there is a vector $p > 0$ such that for any labor endowment L, $P(L) = \{y \geq 0 : py \leq L\}$.*

Consider problem (11) with $y = e$. It describes a minimization problem whose value is bounded below by 0, so an optimal solution exists. Conclude from the Fundamental Theorem that a basic optimal solution exists. The basic optimal solution will of necessity use only one activity for each good, so its support, the set of indices of positive coefficients, is a technology: Call it τ. Since $e >> 0$ is a feasible net output for τ, it is not hard to prove that $(I - A_\tau)^{-1}$ exists and maps the non-negative orthant into itself. That is, technology τ can produce any non-negative net output vector given enough labor. Therefore $\lambda_\tau(y) = a_\tau(I - A_\tau^{-1})y$ for any non-negative y. Note well that $\lambda_\tau(y)$ is linear in y.

Now we show that $\lambda_\tau(y) = \lambda(y)$; that technology τ is the efficient technology for producing any net output $y \geq 0$. First we show that for any commodity i the efficient technology to produce e^i, the net output vector with 1 for good i and 0 for all other commodities. That is, $\lambda_\tau(e^i) \leq \lambda(e^i)$. For if not, the Fundamental Theorem implies the existence of a different technology θ for which $\lambda_\theta(e^i) < \lambda_\tau(e^i)$. But if so we have the contradiction

$$\begin{aligned}\lambda(e) &= \lambda_\tau(e) \\ &= \sum_k \lambda_\tau(e^k) \\ &> \sum_{k\neq j} \lambda_\tau(e^k) + \lambda_\theta(e^j) \\ &\geq \lambda(e).\end{aligned}$$

The last inequality follows from the fact that it is certainly feasible to use θ to produce 1 unit of good j and technology τ to produce 1 unit of everything else.

To conclude the proof, observe that, again from the Fundamental Theorem, any net output y can be produced efficiently by some technology γ, and

$$\begin{aligned}\lambda_\gamma(y) &= \lambda_\gamma(\sum_j y_j e^j) \\ &= \sum_j y_j \lambda_\gamma(e^j) \\ &\geq \sum_j y_j \lambda_\tau(e^j) \\ &= \lambda_\tau(y).\end{aligned}$$

Thus τ is an efficient technology for producing any net-output vector. Given a labor endowment L, the set of feasible net outputs is then

$$P(L) = \{y : a_\tau(I - A_\tau)^{-1} y \leq L\}$$

which is the set of non-negative net output bundles lying on and underneath the hyperplane $a_\tau(I - A_\tau)^{-1} y = L$. In other words, the p of the theorem is $p = a_\tau(I - A_\tau)^{-1}$. Taking labor as the numeraire good, the 0-profit price vector is then $p = a_\tau(I - A_\tau)^{-1}$. Constant returns to scale production with only labor as a non-produced input gives a labor theory of value. Finally, if either $a \gg 0$ or if the input requirements matrix of the efficient technology is irreducible, it can be shown that $p \gg 0$.

4 Transferable Utility Matching

5 Auctions

6 Zero-Sum Games

7 Correlated Equilibrium

Appendices

Appendix 1: Deeper Facts About Cones and Polyhedra

There are two ways of describing convex cones. The cone generated by the vectors $\mathcal{A}$ is the set of all finite non-negative linear combinations of the elements of $\mathcal{A}$.

Definition 11. *A convex cone C is* generated *by the set of vectors $a \in \mathcal{A}$ if for each $c \in C$ there are vectors $a_1, \dots, a_k \in \mathcal{A}$, $k < \infty$, and scalars $\lambda_1, \dots, \lambda k$ such that $c = \sum_{j=1}^{k} \lambda_j a_j$. If $\mathcal{A}$ is finite, then C is* finitely generated, *and $c \in C$ iff $c = Ax$ for some $x \geq 0$, where the column vectors of A are the members of $\mathcal{A}$.*

Another "finite description" of a convex cone is by homogeneous inequalities:

Definition 12. *A convex cone is* polyhedral *iff there is a matrix B such that $x \in C$ iff $Bx \leq 0$.*

Finitely generated cones and polyhedral convex cones are one and the same. This fact is fundamental to the development of Farkas lemma and friends, linear and convex programming.

Theorem 16. *A convex cone is finitely generated iff it is polyhedral.*

As exciting as the proof of this theorem is, even more exciting are some of the concepts needed in its execution.

To show that a finitely generated cone is polyhedral we will need a result useful in its own right, that the projection of a polyhedron onto a lower-dimensional subspace is polyhedral. Let C denote the convex polyhedron $\left\{ (x, y) : [A\ B] \begin{bmatrix} x \\ y \end{bmatrix} \leq b \right\}$ in $\mathbb{R}^n$.

Lemma 3 (Projection Lemma)**.** *The set $\{y :$ for some x, $(x, y) \in C\}$ is polyhedral.*

Proof. The proof proceeds by projecting out one variable at a time. It suffices to show that if $C = \{x : Ax \leq b\}$, then $D = \{(x_2, \dots, x_n) :$ for some x_1, $(x_1, \dots, x_n) \in C\}$ is polyhedral. After combinining inequalities if necessary and reversing sign if necessary, the inequalities fall into three classes:

$$\begin{cases} a_{i1}x_1 + \cdots + a_{in}x_n \leq b_i & i \in I \\ a_{j1}x_1 + \cdots + a_{jn}x_n = b_j & j \in J \\ a_{k1}x_1 + \cdots + a_{kn}x_n \geq b_k & k \in K \end{cases}$$

where all the coefficients of x_1 are non-negative. If the set J is non-empty and contains an equation for which $a_{j1} > 0$, solve for x_1 in terms of $x_2, \ldots, x_n$ and substitute into the remaining equations to have a set of linear inequalities involving only these variables that defines D.

The more interesting case is when J is empty or each $a_{j1} = 0$. In this case the inequalities can be rewritten as

$$x_1 \leq a'_{i2}x_2 + \cdots + a'_{in}x_n + b'_i \qquad i \in I$$
$$a'_{k2}x_2 + \cdots + a'_{kn}x_n + b'_k \leq x_1 \quad k \in K$$

where the primed coefficients are derived from the unprimed coefficients by the obvious algebraic manipulations. These inequalities are satisfiable for some $(x_2 \ldots, x_n)$ if and only if the right hand side of every inequality in I exceeds the left hand side of every inequality in K; that is, for all $i \in I$ and $k \in K$,

$$(a'_{k2} - a'_{i2})x_2 + \cdots + (a'_{kn} - a'_{in})x_n \leq b'_i - b'_k.$$

These inequalities together with any equalities in J (if J is not empty) define a polyhedral set involving only the variables $x_2, \ldots, x_n$ that define D. □

The Projection Lemma is interesting in its own right. It says that if C is a polyhedron, then the set of all y for which *there exists* an x such that $(x, y) \in C$ is a polyhedron. This projection procedure is an example of "quantifier elimination". A class of sets $\mathcal{C}$ is *closed under quantifier elimination* if for all $C \in \mathcal{C}$ the sets $D = \{(x_2, \ldots, x_n) : \exists x_1 \ (x_1, \ldots, x_n) \in C\}$ and $E = \{(x_2, \ldots, x_n) : \forall x_1 \ (x_1, \ldots, x_n) \in C\}$ are both in $\mathcal{C}$. The projection lemma shows the *exists* piece for convex sets. The $\forall$ piece is also true. For each x_1, $C(x_1) = \{(x_2, \ldots, x_n) : (x_1, \ldots, x_n) \in C\}$ is convex. Let C_1 denote the projection of C onto its first coordinate. Then $E = \cap_{x_1 \in C_1} C(x_1)$, and the intersection of convex sets is convex. Closure of a class of sets under *quantifier elimination* is an important property for many different classes of sets. For an application to game theory of the closure under quantifier elimination of the class of sets defined by algebraic inequalities, see Blume and Zame (1994), "The Algebraic Geometry of Perfect and Sequential Equilibrium," *Econometrica* 62: 783-794.

One consequence of the projection lemma is one half of Theorem 16 .

Lemma 4. *Every finitely generated cone is polyhedral.*

Proof. If C is the cone spanned by $v_1, \ldots, v_m$, then C is the projection onto the y variables of

$$\{(x,y) : y = Vx, x \geq 0\} = \{(x,y) : y - Vx \leq 0, Vx - y \leq 0, x \geq 0\}$$

where V is the matrix with column vectors v_j. This last set is polyhedral, and so C is as well. □

The other direction, that polyhedral cones are finitely generated, is more complicated, and requires some new concepts that are of interest in their own right. If C is a convex cone, any hyperplane supporting C contains the origin. Thus a minimal set of half-spaces containing C are described by those y such that $y \cdot c \leq 0$ for all $c \in C$. The set of such y is called the polar cone of C, and gives the "dual description of C. The geometry of the relationship of cones and polar cones is shown in figure 4.

Definition 13. *The* polar (dual) cone *of a cone C is the closed and convex cone $C^* = \{y : y \cdot x \leq 0\}$.*

What really makes the polar a dual description of a closed convex cone is that C and C^* can each be recovered from the other. Although the proof is short, this is important.

Theorem 17. *If C is closed and convex, then $C^{**} = C$.*

Proof. If $x \in C$, then $y \cdot x \leq 0$ for all $y \in C^*$ so $x \in C^{**}$. If $x \notin C$, then there is a y such that $y \cdot x > 0$ and $y \cdot z \leq 0$ for all $z \in C$. The last inequality says that $y \in C^*$ and then the strict inequality disqualifies x for membership in C^{**}. □

The polar cones of polyhedral cones are finitely generated.

Lemma 5. *If C is a polyhedral cone, then C^* is finitely generated.*

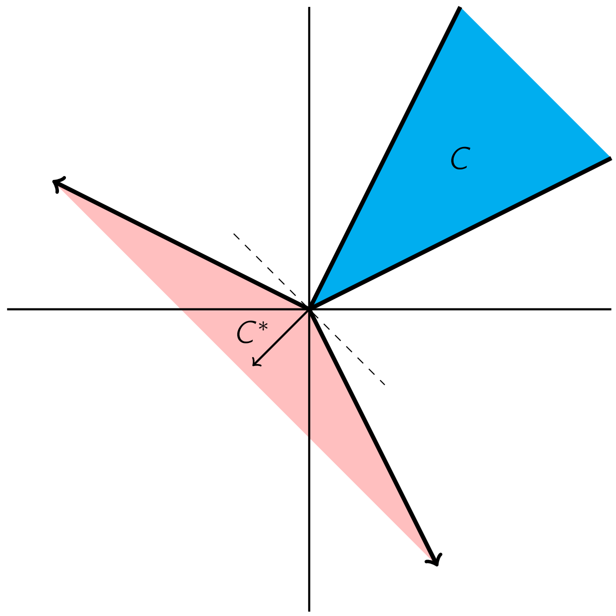


Figure 4: A cone and its polar cone.

Proof. Suppose that C is the set of all vectors x such that $Vx \leq 0$. A good guess for generators of C^* are the rows of V, since each inequality defining C is of the form $v \cdot x \leq 0$ for some row vector v of V. Let D be the cone generated by the rows of V; $x \in D$ iff $y = \sum_j \gamma_i v_i$ where the γ_i's are non-negative and the v_i's are the rows of V. First, $D \subset C^*$ since $y \cdot x = \sum_i y_i v_i \cdot x \leq 0$ for all $x \in C$.

Suppose $y \notin D$. Then there is an x such that $y \cdot x > 0$ and $d \cdot x \leq 0$ for all $d \in D$. Then $\sum_i \gamma_i v_i \cdot x \leq 0$ for all non-negative γ_i. In particular this applies to any $\gamma_i = 1$ and the others 0, so $v_i \cdot x \leq 0$ for all i. From the weak inequality we conclude that $x \in C$, and from the strong inequality that $y \notin C^*$. □

The remainder of theorem 16's proof is quick: If C is polyhedral, then C^* is finitely generated. Lemma 4 implies that C^* is polyhedral. Then C^{**} is finitely generated, and $C^{**} = C$.

Appendix 2: Farkas' Lemma and Theorem 11.

The first theorem of the alternative, due to Gauss, goes as follows. Here A is an $m \times n$ matrix and $b \in \mathbb{R}^n$.

Fundamental Theorem of Linear Algebra. *One and only one of the following alternatives holds:*

1. *There is an $x \in \mathbb{R}^n$ such that $Ax = b$.*
2. *There is a $y \in \mathbb{R}^m$ such that $yA = 0$ and $yb \neq 0$.*

Why is this true? Obviously both cannot be true; otherwise $0 \neq yb = aAx = 0x = 0$. If 1) fails, then b is not in the column space of A. So the column space is a strict subspace of $\mathbb{R}^m$. For any y in its orthogonal complement, $yAx = 0$ for all $x \in \mathbb{R}^m$, and so $yA = 0$.

Farkas' lemma is harder to prove because of the non-negativity constraint, that $x \geq 0$. The missing piece from the proof of Farkas' lemma is to show the following lemma:

Lemma A.1. *The cone $C = \{b : Ax = b, x \geq 0\}$ is closed.*

Proof. The cone C is generated by the columns of A. From Lemma A.4, C is polyhedral. Polyhedral cones are the intersection of a finite number of closed half spaces, and therefore are closed. □

To nail down the proof of Theorem 11 it helps to have the following fact, which is the Minkowski half of the Minkowski-Weyl Theorem:

Lemma A.2. *If the closed polyhedron $\{x : Ax \leq b\}$ is non-empty, then it equals the set sum $P + Q$, where P is a bounded convex set and Q is the cone $\{x : Ax \leq 0\}$.*

The converse is also true; perhaps this is the Weyl half. In summary, a set P is polyhedral if and only if it is the sum of a finitely generated convex set and a finitely generated convex cone.

Proof. Our polyhedron is $C(b) = \{x : Ax \le b\} \subset \mathbb{R}^n$. We map this set to the cone $K(b) = \{(y, t) \in \mathbb{R}^n \times \mathbb{R} : Ay - tb \le 0, t \ge 0\}$. Clearly this is a polyhedral cone in the half-space $\{(y, t) : t \ge 0\}$. Note that $C(b) = \{x : (x, t) \in K, t = 1\}$.

Since $K(b)$ is polyhedral, it is generated by a finite number of vectors $(y_1, t_1), \ldots, (y_r, t_r)$, with all the $t_r \ge 0$. This is a consequence of Theorem 16. We can rescale the vectors so as to have any positive t_i equal 1. Then without loss of generality we can suppose that the first p vectors are of the form $(w_i, 1)$ and the remaining q vectors are of the form $(z, 0)$. If $p = 0$, then $C(b)$ is empty, so $p \ge 1$. By taking $(y_r, t_r) = (0, 0)$ if necessary, we can without loss of generality assume that $q \ge 1$. The set of all vectors in $K(b)$ with last coefficient 1 is the set of all vectors $(y, t) = \sum_{i=1}^{p} \alpha_i (w_i, 1) + \sum_{j=1}^{q} \beta_j (z_j, 0)$, where the α_i and β_j are all non-negative, and $\sum_{i=1}^{p} \alpha_i = 1$. Thus take P to be the set of convex combinations of the w_i, and take Q to be the set of non-negative affine combinations of the z_j; P is clearly bounded, and Q is clearly a cone.

Finally, for any $(z, 0) \in K(b)$, by definition $Az = Az - 0b \le 0$, so $\{z : Az \le 0\} \subset Q$. On the other hand, note from the construction that if $x \in P$, then $Ax \le b$. If $Az \le 0$, then $A(x + z) \le b$. Thus $z \in Q$, so $Q = \{z : Az \le 0\}$. □

Appendix 2: Farkas Lemma

The missing piece of the proof of Farkas' lemma is to show the following lemma:

Lemma A.3. *The cone $C = \{b : Ax = b, x \ge 0\}$ is closed.*

Proof. The cone C is generated by the columns of A. From Lemma A.4, C is polyhedral. Polyhedral cones are the intersection of a finite number of closed half spaces, and therefore are closed. □

Lemma A.4 is an immediate consequence of the projection lemma. The projection lemma requires none of the apparatus of Appendix A.1, so this result really is very basic. (It requires more in the way of definitions than anything else.)

Appendix 3: Proof of Theorem 13

An easy piece of this highlights the importance of $b = 0$: $v_P(0)$ is either 0 or $= \infty$, and if the latter, then $v_P(b) = +\infty$ for all b in the domain.

Lemma A.4. *If $v_P(0) \neq 0$, then $v_P(0) = +\infty$.*

Proof. Since 0 is feasible for $\mathcal{P}(0)$, $v_P(0) \geq 0$. If there is an $x \geq 0$ such that $Ax \leq 0$ and $c \cdot x = \alpha > 0$, then for any $t > 0$, $tx \geq 0$, $Atx \leq 0$ and $c \cdot tx = t\alpha$, so $v_P(0) = +\infty$. $\square$

Lemma A.5. *If $v_P(0) = +\infty$, then for all $b \in \mathbb{R}^n$, $v_P(b) = +\infty$.*

Proof. If $v_P(0) = +\infty$, then for any $\alpha > 0$ there is an x feasible for $\mathcal{P}(0)$ such that $c \cdot x > \alpha$. If x' is feasible for $\mathcal{P}(b)$, then so is $x + x'$ and so $v_P(b) \geq c \cdot (x' + x) = c \cdot x' + \alpha$. $\square$

The remaining step is to show that if for any $b \in \mathbb{R}^n$, if $v_P(b) = +\infty$ then $v_P(0) = +\infty$. This follows from lemma A.2. If $v_P(b) = +\infty$, then there is a sequence $\{x_n\}$ of feasible solutions such that $c \cdot x_n$ converges to $+\infty$. These can be written $x_n = p_n + q_n$ as in the lemma. Since P is bounded, the $c \cdot p_n$ are bounded, and so $c \cdot q_n \uparrow +\infty$; $v_P(0) = +\infty$. $\square$

Appendix 4: Saddle Point Theory for Constrained Optimization

The duality ideas of linear programming generalize to concave and convex optimization. Here is a very brief sketch.

Let $H : X \times Y \to \mathbb{R}$, not necessarily concave.

Definition A.1. *A point $(x^*, y^*) \in X \times Y$ is a* **saddle point** *of H iff for all $x \in X$ and $y \in Y$,*

$$H(x, y^*) \leq H(x^*, y^*) \leq H(x^*, y).$$

In words, x^* maximizes H over X and y^* minimizes H over Y. Saddle points are important for convex optimization and in game theory. Here is one important feature of the set of saddle points of H: The set of saddle point pairs is a product set.

Theorem A.1. *If (x_1^*, y_1^*) and (x_2^*, y_2^*) are saddle points, then (x_1^*, y_2^*) and (x_2^*, y_1^*) are saddle points, and*

$$H(x_1^*, y_1^*) = H(x_1^*, y_2^*) = H(x_2^*, y_1^*) = H(x_2^*, y_2^*).$$

Proof. For $i, j = 1, 2$,

$$H(x_i^*, y_i^*) \leq H(x_i^*, y_j^*) \leq H(x_j^*, y_j^*)$$

so all four points take on the same value.

From this equality and the saddle point properties of $H(x_i^*, y_i^*)$ it follows that

$$H(x, y_2^*) \leq H(x_2^*, y_2^*) = H(x_1^*, y_2^*) = H(x_1^*, y_1^*) \leq H(x_1^*, y),$$

proving that (x_1^*, y_2^*) is a saddle point. A similar argument shows the same for (x_2^*, y_1^*). □

Now lets suppose that f and $g_1, \dots, g_m$ are each functions from a convex set $C \subset \mathbb{R}^n$ to $\mathbb{R}$. Denote by $g : X \to \mathbb{R}^m$ the function whose ith coordinate function is g_i. Consider the constrained optimization problem

$$\begin{aligned} \mathcal{P}: \qquad \max_x \quad & f(x) \\ \text{s.t.} \quad & g_1(x) \geq 0 \\ & \vdots \\ & g_m(x) \geq 0. \end{aligned}$$

The **Lagrangian** for this problem is the function on $C \times \mathbb{R}_+^m$

$$L(x, \lambda) = f(x) + \sum_{i=1}^m y_i g_i(x).$$

The $x \in C$ are the **primal variables** and the $y \in \mathbb{R}_+^m$ are the **dual variables**, also called **Lagrange multipliers**.

The main theorem is true for any functions, concave or not.

Theorem A.2. *Let X be a subset of $\mathbb{R}^n$, let f and $g_1, \ldots, g_m$ be functions from X to $\mathbb{R}$. If (x^*, y^*) is a saddle point of the Lagrangian $L(x, y)$ then x^* solves $\mathcal{P}$. Furthermore, for all i, $y_i g_i(x^*) = 0$ (the* **complementary slackness property***).*

Proof. The saddle point property is that for all $x \in X$ and $y \in \mathbb{R}^m_+$,

$$L(x, y^*) \leq L(x^*, y^*) \leq L(x^*, y).$$

The second inequality implies that all $y^* \cdot g_(x^*) \leq y \cdot g(x^*)$ for all $y \geq 0$, so $g(x^*) \geq 0$. (Why?) This shows that x^* is feasible. Since $0 \leq y^* \cdot g(x^*) \leq 0 \cdot g(x^*) = 0$, so $y^* \cdot g(x^*) = 0$ Since each term in the sum is non-negative, all must be 0. This proves the complementary slackness property.

The first inequality implies that for all $x \in C$,

$$f(x) + y^* g(x) \leq f(x^*) + y^* \cdot g(x^*) = f(x^*).$$

Thus for any feasible x, that is, $g(x) \geq 0$, $f(x) \leq f(x^*)$. □

Define for $x \geq 0$ the *primal functional*

$$f_p(x) = \inf_{y \geq 0} L(x, y).$$

At any infeasible $x \geq 0$, $f_p(x) = -\infty$ since some $f_i(x) < 0$. At any feasible x for which constraint i is not binding, $g_i(x) > 0$ and so minimization of $L(x, y)$ requires that $y_i = 0$. Thus on the feasible set, $f_p(x) = f(x)$. Consequently, the problem of maximizing $f_p(x)$ is that of maximizing $f(x)$ on the feasible set. In this sense the primal functional encodes the primal program. If (x^*, y^*) is a saddle point, x^* is feasible for the primal program, and for any feasible x,

$$f(x^*) = f_p(x^*) = L(x^*, y^*) \geq L(x, y^*) \geq \inf_{y \geq 0} L(x, y) = f(x)$$

and so x^* solves the primal program.[4]

In a similar manner, define for $y \geq 0$ the *dual functional*

$$f_d(y) = \sup_{x \geq 0} L(x, y).$$

The dual program is the program that minimizing the dual encodes.

Specialize this to linear programs. The Lagrangean for the primal program is

$$\begin{aligned} L(x, y) &= cx + y \cdot (b - Ax) \\ &= yb + (c - yA)x. \end{aligned}$$

the *primal functional* is

$$f_p(x) = \begin{cases} -\infty & \text{if } b - Ax \not\geq 0; \\ cx & \text{otherwise;} \end{cases}$$

for $x \geq 0$. As we have seen, maximizing the primal functional entails solving the primal program.

The *dual functional* takes on the values

$$f_d(y) = \begin{cases} +\infty & \text{if } c - yA \not\leq 0; \\ yb & \text{otherwise;} \end{cases}$$

for $y \geq 0$. The functional f_d encodes the dual program because its value is that of the dual objective function, yb, on the dual feasible set $yA \geq c$, and $+\infty$ elsewhere; thus the dual program is to minimize $f_d(y)$ on $\mathbf{R}^m_+$. The dual problem and the primal problem have the same Lagrangean. The statement that the primal Lagrangean has a saddle point is strong duality, that the value of the primal program equals that of the dual program.

The strong duality theory for linear programs shows that the Lagrangean for these problems has a saddle point. Under what conditions, generally speaking, do Lagrangeans have saddle

[4] Conversely, given appropriate constraint qualifications, if x^* solves the primal program, then there is a y^* such that (x^*, y^*) is a saddle point.

points? The first fact amounts to weak duality for the case of linear programs:

Theorem A.3. *If C and D are arbitrary subsets of two Euclidean spaces, and $H : C \times D \to [-\infty, \infty]$, then*

$$\sup_{c\in C} \inf_{d\in D} H(c, d) \geq \inf_{d\in D} \sup_{c\in C} H(c, d).$$

Proof. Define for all $c \in C$ the functional $f_p(c) = \inf_{d\in D} H(c, d)$, and let

$$\alpha = \sup_{c\in C} f_p(c) = \sup_{c\in C} \inf_{d\in D} H(c, d).$$

For each $d \in D$, $H(c, d) \geq f_p(c)$ for all $c \in C$, and so

$$\sup_{c\in C} H(c, d) \geq \sup_{c\in C} f_p(c) = \alpha.$$

Since this holds for all $d \in D$, it follows that

$$\inf_{d\in D} \sup_{c\in C} H(c, d) \geq \alpha,$$

which is the Theorem's claim. □

This abstract inequality says little about saddle points *per se.* They may fail to exist. But the preceding theorem gives a step up on this problem.

Lemma A.6. *A point $(c^*, d^*) \in C \times D$ is a saddle point of H iff*

1. $\sup_{c\in C} \inf_{d\in D} H(c, d) = \inf_{d\in D} H(c^*, d)$,
2. $\inf_{d\in D} \sup_{c\in C} H(c, d) = \sup_{c\in C} H(c, d^*)$, *and*
3. $\sup_{c\in C} \inf_{d\in D} H(c, d) = \inf_{d\in D} \sup_{c\in C} H(c, d)$.

The value of the equality in 3) is called the *saddle value* of H.

Proof. If (c^*, d^*) is a saddle point, then

$$H(c^*, d^*) = \inf_{d\in D} H(c^*, d) \leq \sup_{c\in C} \inf_{d\in D} H(c, d),$$

$$H(c^*, d^*) = \sup_{c\in C} H(c, d^*) \geq \inf_{d\in D} \sup_{c\in C} H(c, d).$$

It follows from Theorem A.3 that these inequalities are all equal, so the conditions of the Lemma are satisfied.

If the conditions of the Lemma are satisfied, the saddle value of H, denote by α the saddle value of H. Then

$$H(c^*, d^*) \leq \sup_{c \in C} H(c, d^*) = \alpha = \inf_{d \in D} H(c^*, d) \leq H(c^*, d^*),$$

and so (c^*, d^*) is a saddle point for H. $\square$

There is a general theory of saddle points for concave-convex functions. A typical result is this:

Theorem A.4. *Suppose that C and D are non-empty, compact and convex, and suppose that H is continuous. Then a sassle point exists.*

The proof is of independent interest; it is essentially the proof of the existence of Nash equilibrium.

Proof. Define $\phi_H(c) = \mathrm{argmin}_{d \in D} H(c, d)$ and $\phi^H(d) = \mathrm{argmax}_{c \in C} H(c, d)$. By virtue of the Maximum Theorem, both of these correspondences are non empty-valued and upper hemi-continuous. Thus they have closed graph, and so are compact-valued. Concave-convex implies that both are convex-valued. Finally, $\Phi : (c, d) \mapsto \phi^H(d) \times \phi_H(c)$ inherits all these properties, and hence has a fixed point (c^*, d^*). The problem $\max_{c \in C} H(c, d^*)$ is solved by c^* and the problem $\min_{d \in D} H(c^*, d)$ is solved by d^*. Thus (c^*, d^*) is a saddle-point. $\square$