An elementary proof of linear programming optimality conditions without using Farkas' lemma

Anders Forsgren
Optimization and Systems Theory
Department of Mathematics
KTH Royal Institute of Technology
Stockholm, Sweden*

Margaret H. Wright
Department of Computer Science
Courant Institute of Mathematical Sciences
New York University
New York, New York, USA†

July 7, 2014

Abstract

Although it is easy to prove the sufficient conditions for optimality of a linear program, the necessary conditions pose a pedagogical challenge. A widespread practice in deriving the necessary conditions is to invoke Farkas' lemma, but proofs of Farkas' lemma typically involve "nonlinear" topics such as separating hyperplanes between disjoint convex sets, or else more advanced LP-related material such as duality and anti-cycling strategies in the simplex method. An alternative approach taken previously by several authors is to avoid Farkas' lemma through a direct proof of the necessary conditions. In that spirit, this paper presents what we believe to be an "elementary" proof of the necessary conditions that does not rely on Farkas' lemma and is independent of the simplex method, relying only on linear algebra and a perturbation technique published in 1952 by Charnes. No claim is made that the results are new, but we hope that the proofs may be useful for those who teach linear programming.

1. Introduction

In many contexts, particularly in business and economics, linear programming (LP) is taught as a self-contained subject, including proofs of necessary and sufficient conditions for optimality. The proof of sufficient conditions is straightforward, but, as explained below, the necessary conditions are seen by many as pedagogically challenging. Broadly speaking, these conditions are proved in two different ways.

The first invokes Farkas' lemma,[1] which can be stated in a surprisingly large number of forms (for example, [15, pages 89–93]) and which itself must be proved. According to the history sketched in [3], the initial statement of Farkas' lemma was published in 1894, with its best-known exposition appearing in 1902. Despite the passage of more than a century since its correctness was established, different proofs of the lemma have continued to be devised, based on an array of motivations categorized in [3] as geometric, algebraic, and/or algorithmic.

A widely used means of proving Farkas' lemma relies on separating hyperplane theorems (for example, [9, pages 205–207], [11, pages 297–301],[2, pages 170–172]), but there can be some discomfort in bringing these more advanced topics into an LP course whose audience is familiar only with basic linear algebra. Even so, this approach is especially convenient when optimality conditions for a range of increasingly complicated constrained optimization problems are to be presented (for example, [13, pages 326–329]).

*andersf@kth.se
†mhw@cs.nyu.edu

[1] The *Chicago Manual of Style Online*, `www.chicagomanualofstyle.org`, prefers repeating the "s" after the apostrophe in indicating possession by a word ending in "s", but states that it is also correct to omit the post-apostrophe "s". We have chosen the second option.

Those who wish to keep Farkas' lemma but prefer to avoid separating hyperplanes can choose instead from a non-trivial number of "elementary" proofs of the lemma. Some of these, such as [7, 16], involve properties of linear least-squares problems. An algebraic proof related to orthogonal matrices is given in [3] (see also [14]), and [1] features a linear-algebraic approach to proving Farkas' lemma and other theorems of the alternative.

A second strategy is to prove the necessary conditions for LP optimality without explicitly calling on Farkas' lemma. This can be done in a variety of ways, for example using LP-related results such as duality (see [2, page 165], [8, pages 112-113]) or finite termination of the simplex method with a guaranteed anticycling strategy [15, page 86]. A recent proof of LP optimality conditions [10] is independent of the simplex method, relying instead on linear algebra and a perturbation technique introduced by [4] in the context of resolving degeneracy.

This paper, which falls into the second group, presents proofs of optimality conditions for linear programs expressed in a generic form that includes both equalities and inequalities; see (1.3). The problem form may seem inconsequential, especially since all known LP problem forms can be mechanically transformed into one another. But form affects substance to a perhaps surprising extent, and can have a major effect on how students and practitioners think about linear programs and algorithms for solving them.

Linear programs are (probably) most frequently expressed in textbooks using one of several variations on "standard form", of which the following is typical:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ c^Tx \quad \text{subject to} \quad Ax = b \quad \text{and} \quad x \ge 0, \tag{1.1}$$

where A is $m \times n$ with $m \le n$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and A is assumed to have rank m. Two key features of this version of standard form are that the "general" constraints involving A are all equalities, and that the only inequalities are simple lower bounds on the variables.

Standard form is very closely tied to the simplex method, which is described in many papers and books (for example, the 1963 classic [6], [5] and [17]) and which was, for almost 40 years, essentially the only method for solving linear programs. However, since the 1984 "interior-point revolution" in optimization (for example, [18, 12]), a thorough treatment of linear programming requires presentation of interior-point methods. These are easier to motivate with *all-inequality form*, which resembles a generic form for constrained optimization:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ c^Tx \quad \text{subject to} \ \ Ax \ge b, \quad \text{where } A \text{ is } m \times n. \tag{1.2}$$

A linear program in standard form (1.1) may be transformed (by reformulating the constraints and/or adding variables) into an equivalent linear program in all-inequality form (1.2) and the other way around. This paper considers a generic mixed form in which equality and inequality constraints are denoted separately:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ c^Tx \quad \text{subject to} \ \ A_{\mathcal{E}}x = b_{\mathcal{E}} \ \ \text{and} \ \ A_{\mathcal{I}}x \ge b_{\mathcal{I}}, \tag{1.3}$$

where $A_{\mathcal{E}}$ is $m_{\mathcal{E}} \times n$ with $\operatorname{rank}(A_{\mathcal{E}}) = m_{\mathcal{E}}$ and $A_{\mathcal{I}}$ is $m_{\mathcal{I}} \times n$. This form corresponds to all-inequality form when $m_{\mathcal{E}} = 0$, i.e., when $A_{\mathcal{E}}$ is empty, and to standard form when $A_{\mathcal{I}} = I_n$ (the n-dimensional identity) and $b_{\mathcal{I}} = 0$. This means that results on standard form as well as all-inequality form are immediately available from our results. We define the combined matrix A and vector b, each with $m = m_{\mathcal{E}} + m_{\mathcal{I}}$ rows, as

$$A = \begin{pmatrix} A_{\mathcal{E}} \\ A_{\mathcal{I}} \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_{\mathcal{E}} \\ b_{\mathcal{I}} \end{pmatrix}. \tag{1.4}$$

where the index sets $\mathcal{E}$ and $\mathcal{I}$ are $\mathcal{E} = \{1, 2, \ldots, m_{\mathcal{E}}\}$ and $\mathcal{I} = \{m_{\mathcal{E}} + 1, \ldots, m_{\mathcal{E}} + m_{\mathcal{I}}\}$.

Sections 2–4 contain a summary of background results that would be part of any course on linear programming; they are included to make the paper self-contained. The results in Sections 5–8 are not new in substance, but may be unfamiliar in form. In any case we hope that they might provide a useful option for proving LP optimality. For completeness, Section 9 states and proves Farkas' lemma using the results in this paper; Section 10 summarizes the logical flow of results.

2. Notation, definitions, and background results

It is assumed that $c \neq 0$ and $A \neq 0$. The ith row of A (1.4) is denoted by a_i^T and the ith component of b by b_i. The problem constraints are said to be *consistent* or *feasible* if there exists at least one $\widehat{x}$ such that $A_{\mathcal{E}}\widehat{x} = b_{\mathcal{E}}$ and $A_{\mathcal{I}}\widehat{x} \geq b_{\mathcal{I}}$, and an $\widehat{x}$ that satisfies the constraints is called a *feasible point*. An immediate result, noted explicitly for completeness, is that linearity of the constraints means that every point on the line joining two distinct feasible points $\widehat{x}$ and $\bar{x}$ is also feasible.

Optimality subject to constraints is inherently a relative condition, involving comparison of objective function values at a possible optimal point $\widehat{x}$ with those at other feasible points. In such a comparison, *active* constraints play a crucial role. The ith constraint is said to be *active* at a feasible point $\widehat{x}$ if $a_i^T\widehat{x} = b_i$. At a feasible point, the equality constraints $A_{\mathcal{E}}\widehat{x} = b_{\mathcal{E}}$ must be active, but an inequality constraint may be active or inactive (strictly satisfied). The set of indices of inequality constraints active at a feasible point $\widehat{x}$ is denoted by $\bar{\bar{\mathcal{I}}}(\widehat{x})$, and $\overset{>}{\mathcal{I}}(\widehat{x})$ denotes the set of indices of the inactive inequality constraints at $\widehat{x}$. We use $\mathcal{A}(\widehat{x})$ to denote the set of active constraints, which means that $\mathcal{A}(\widehat{x}) = \mathcal{E} \cup \bar{\bar{\mathcal{I}}}(\widehat{x})$. Let $\bar{\bar{A}}_{\mathcal{I}}(\widehat{x})$ denote the matrix of rows of $A_{\mathcal{I}}$ corresponding to active inequality constraints, and similarly for $\bar{\bar{b}}_{\mathcal{I}}(\widehat{x})$, so that, by definition, $\bar{\bar{A}}_{\mathcal{I}}(\widehat{x})\widehat{x} = \bar{\bar{b}}_{\mathcal{I}}(\widehat{x})$. The *active-constraint matrix* $\bar{\bar{A}}(\widehat{x})$ then consists of $A_{\mathcal{E}}$ and $\bar{\bar{A}}_{\mathcal{I}}(\widehat{x})$:

$$\bar{\bar{A}}(\widehat{x}) = \begin{pmatrix} A_{\mathcal{E}} \\ \bar{\bar{A}}_{\mathcal{I}}(\widehat{x}) \end{pmatrix}. \tag{2.1}$$

The following definition, in which positivity of α^i is crucial, allows us to characterize feasible directions at a feasible point.

Definition 2.1. (Feasible direction.) *The n-vector p is a feasible direction for the constraints $A_{\mathcal{E}}x = b_{\mathcal{E}}$ and $A_{\mathcal{I}}x \geq b_{\mathcal{I}}$ at the feasible point $\widehat{x}$ if $p \neq 0$ and there exists $\alpha^i > 0$ such that $\widehat{x} + \alpha p$ is feasible for $0 < \alpha \leq \alpha^i$, where α^i may be infinite.*

By linearity, the value of the ith constraint when moving from a feasible point $\widehat{x}$ to $\widehat{x} + \alpha p$, where $p \neq 0$ and $\alpha > 0$, is given by

$$a_i^T(\widehat{x} + \alpha p) = a_i^T\widehat{x} + \alpha a_i^T p. \tag{2.2}$$

Relation (2.2) shows that, to maintain feasibility with respect to an equality constraint i, p must satisfy $a_i^T p = 0$. When i is an *inactive* inequality constraint, (2.2) implies that, even if $a_i^T p < 0$, constraint i will remain satisfied at $\widehat{x} + \alpha p$ if $\alpha > 0$ is sufficiently small. But if i is an active inequality constraint, so that $a_i^T\widehat{x} = b_i$, it follows from (2.2) that $\widehat{x} + \alpha p$ will be feasible with respect to constraint i for $\alpha > 0$ only if $a_i^T p \geq 0$.

Although inactive inequality constraints do not have an immediate local effect on feasibility, they may limit the size of the step that can be taken along a feasible direction p.

Definition 2.2. (The maximum feasible step.) *Given a feasible point $\widehat{x}$ and a feasible direction p, let $\mathcal{D}(\widehat{x}, p)$ (for "decreasing") denote the set of indices of inequality constraints that are inactive at $\widehat{x}$ for which $a_i^T p < 0$:*

$$\mathcal{D}(\widehat{x}, p) \triangleq \{i \mid i \in \overset{>}{\mathcal{I}}(\widehat{x}) \ \text{ and } \ a_i^T p < 0\}.$$

If $\mathcal{D}(\widehat{x}, p) \neq \emptyset$, the positive scalar σ^i (the step to constraint i along p) is defined as

$$\sigma^i \triangleq \frac{b_i - a_i^T\widehat{x}}{a_i^T p} \quad \text{for } i \in \overset{>}{\mathcal{I}}(\widehat{x}) \text{ and } a_i^T p < 0. \tag{2.3}$$

The maximum feasible step, denoted by $\widehat{\sigma}$, is the smallest such step, $\widehat{\sigma} \triangleq \min_{i \in \mathcal{D}(\widehat{x}, p)} \sigma^i$. Any inequality constraint i for which $\sigma^i = \widehat{\sigma}$, of which there may be more than one, becomes active at $\widehat{x} + \widehat{\sigma}p$. If $\mathcal{D}(\widehat{x}, p) = \emptyset$, $\widehat{\sigma}$ is taken as $+\infty$.

In addition to feasibility, optimality conditions need to ensure that the objective function is as small as possible.

Definition 2.3. (Descent direction.) *The vector p is a descent direction for the objective function $c^T x$ if $c^T p < 0$.*

Definition 2.4. (Feasible descent direction.) *The direction p is a feasible descent direction at a feasible point $\widehat{x}$ if $A_{\mathcal{E}} p = 0$, $\bar{A}_{\mathcal{I}}(\widehat{x}) p \geq 0$, and $c^T p < 0$. No feasible descent direction exists at $\widehat{x}$ if there are no feasible directions or if $c^T p \geq 0$ for all feasible directions p.*

Obvious optimality conditions can now be stated in terms of existence or non-existence of a feasible descent direction.

Lemma 2.5. (Necessary and sufficient optimality conditions—Version I.)
When minimizing $c^T x$ subject to $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \geq b_{\mathcal{I}}$, the feasible point x^* is optimal if and only if no feasible descent direction exists at x^*.

Proof. The "only if" result follows because existence of a feasible descent direction p at a feasible point x^* implies that there is a positive α such that $x^* + \alpha p$ is feasible and $c^T(x^* + \alpha p) = c^T x^* + \alpha c^T p < c^T x^*$. Hence, x^* cannot be optimal.

To show the "if" result, assume that x^* is feasible but not optimal. Then there is a feasible point $\widetilde{x}$ such that $c^T \widetilde{x} < c^T x^*$. Since $\widetilde{x}$ is feasible, we must have $A_{\mathcal{E}} \widetilde{x} = b_{\mathcal{E}}$ and $\bar{A}_{\mathcal{I}}(x^*) \widetilde{x} \geq \bar{b}_{\mathcal{I}}(x^*)$. In addition, it holds that $A_{\mathcal{E}} x^* = b_{\mathcal{E}}$ and $\bar{A}_{\mathcal{I}}(x^*) x^* = \bar{b}_{\mathcal{I}}(x^*)$. Hence, $c^T(\widetilde{x} - x^*) < 0$, $A_{\mathcal{E}}(\widetilde{x} - x^*) = 0$ and $\bar{A}_{\mathcal{I}}(x^*)(\widetilde{x} - x^*) \geq 0$, so that $\widetilde{x} - x^*$ is a feasible descent direction by Definition 2.1. ∎

Although Lemma 2.5 gives necessary and sufficient conditions for LP optimality, its usefulness is limited because it offers no way to verify these conditions. This is the point in teaching LP where Farkas' lemma usually enters the picture, but we now take a different route to the necessary and sufficient conditions for LP optimality.

3. Multipliers and optimality

An important feature of constrained optimization problems is the implicit existence of quantities that do not appear in the problem statement yet play a crucial role in optimality conditions. These quantities consist of m (Lagrange) *multipliers*, or *dual variables*, one for each constraint, that connect the objective and the constraints. The next result shows that existence of a multiplier with certain properties produces a lower bound on the objective value in the feasible region.

Proposition 3.1. (Lower bound on LP objective.) *Assume that $x \in \mathbb{R}^n$ satisfies $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \geq b_{\mathcal{I}}$. Further assume that there exists a multiplier $\lambda \in \mathbb{R}^m$ such that $A^T \lambda = c$ and $\lambda_{\mathcal{I}} \geq 0$, where $\lambda_{\mathcal{I}}$ denotes the $m_{\mathcal{I}}$-vector of components of λ corresponding to inequality constraints. (No sign restrictions apply to the multipliers for equality constraints.) Then $c^T x - \lambda^T b = \lambda^T(Ax - b) \geq 0$.*

Proof. Let x be feasible. The assumed existence of λ means that we can substitute $A^T \lambda$ for c and use the facts that $\lambda_{\mathcal{I}} \geq 0$, $A_{\mathcal{E}} x - b_{\mathcal{E}} = 0$, and and $A_{\mathcal{I}} \widehat{x} - b_{\mathcal{I}} \geq 0$. We then have

$$c^T x - \lambda^T b = \lambda^T(Ax - b) = \lambda_{\mathcal{E}}^T(A_{\mathcal{E}} x - b_{\mathcal{E}}) + \lambda_{\mathcal{I}}^T(A_{\mathcal{I}} x - b_{\mathcal{I}}) \geq 0.$$

It follows that $c^T x$ is bounded below by $\lambda^T b$ for every feasible x. ∎

For a general linear program, a qualifying λ may not exist. Furthermore, there can be more than one vector λ satisfying the given conditions, each producing a different value of $\lambda^T b$. However, something special happens if $\lambda_{\mathcal{I}}^T(A_{\mathcal{I}} \widehat{x} - b_{\mathcal{I}}) = 0$, allowing us to state *sufficient* conditions for LP optimality.

Proposition 3.2. (Sufficient conditions for LP optimality.) *Consider the linear program of minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \geq b_{\mathcal{I}}$. The feasible point x^* is optimal if a multiplier $\lambda^* \in \mathbb{R}^m$ exists with the following three properties: (i) $A^T \lambda^* = c$, (ii) $\lambda_{\mathcal{I}}^* \geq 0$, and (iii) $\lambda^{*T}(Ax^* - b) = 0$. The optimal objective value is $c^T x^* = \lambda^{*T} b$.*

Proof. Since λ^* satisfies (i) and (ii), Proposition 3.1 gives the lower bound $\lambda^{*T} b$ on the optimal value of the linear program. However, (iii) and Proposition 3.1 show that the lower bound is attained for x^*, so that x^* is optimal. ∎

The crucial relationship $\lambda^{*T}(Ax^* - b) = 0$, which means that, for every $i = 1, \dots, m$, at least one of $\{\lambda^*_i, a_i^T x^* - b_i\}$ must be zero, is called *complementarity*.

The following result shows that the complementarity condition does not directly tie the multiplier λ^* to a particular x^*. Rather, if complementarity holds for one x^*, it must hold for λ^* together with any optimal solution $\widehat{x}$.

Proposition 3.3. (Properties of an optimal LP solution.) *Consider minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$. Assume that a multiplier λ^* exists such that $A^T \lambda^* = c$, $\lambda^*_{\mathcal{I}} \ge 0$, and assume that the optimal objective value is $\lambda^{*T} b$. Then a feasible point $\widehat{x}$ is optimal if and only if $\lambda^{*T}(A\widehat{x} - b) = 0$.*

Proof. For any feasible x, Proposition 3.1 gives $c^T x - \lambda^{*T} b = \lambda^{*T}(Ax - b) \ge 0$. Hence, under the assumption that the optimal value is $\lambda^{*T} b$, a point $\widehat{x}$ is optimal if and only if $\lambda^{*T}(A\widehat{x} - b) = 0$. ∎

4. Vertices and their properties

Certain feasible points, known as *vertices*, are extremely important in linear programming.

Definition 4.1. (Vertex.) *Given the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$, the point $\widehat{x}$ is a vertex if $A_{\mathcal{E}} \widehat{x} = b_{\mathcal{E}}$, $A_{\mathcal{I}} \widehat{x} \ge b_{\mathcal{I}}$, and the active-constraint matrix $\bar{\bar{A}}(\widehat{x})$ of (2.1) has rank n.*

An immediate consequence is that there cannot be a vertex when the rank of the full constraint matrix A (1.4) is less than n. A nice feature of standard-form linear programs (1.1) is that, when the constraints are consistent, a vertex must exist because the inequality constraints consist of the n-dimensional identity. This is not true in general for all-inequality form (1.2), even when the objective function is bounded below in the feasible region; consider, for example, minimizing $x_1 + x_2$ subject to $x_1 + x_2 \ge 1$.

A vertex $\widehat{x}$ is the unique solution of the linear system formed by any nonsingular submatrix of $\bar{\bar{A}}(\widehat{x})$, which has rank n by definition. A fundamental result is that the definitions of vertex and extreme point are equivalent, where an extreme point is a feasible point that does not lie on the line segment joining two distinct feasible points. (See, for example, Section 2.2 of [2] for a detailed treatment of related topics.)

A simple combinatorial argument shows that the number of vertices is bounded above by $\binom{m}{n}$. Given the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$, the set of vertices $\mathcal{V}(A, b)$ can be found by enumerating and testing for feasibility every combination of the constraints in which the equality constraints hold with equality and a total of n constraints hold with equality. (Of course, this procedure is not practical for large m and n.)

There are two kinds of vertices. At a *nondegenerate vertex* $\widehat{x}$, exactly n constraints are active and the active-constraint matrix $\bar{\bar{A}}(\widehat{x})$ is nonsingular. At a *degenerate vertex* $\widehat{x}$, there are n linearly independent active constraints, but more than n constraints are active.

The next two small results are stated formally for later reference.

Result 4.2. *Let F be a $q \times n$ nonzero matrix with* $\operatorname{rank}(F) < n$ *whose ith row is f_i^T. Assume that p is a nonzero n-vector such that $Fp = 0$. If g is a vector such that $g^T p \ne 0$, then g^T is linearly independent of the rows of F, i.e.*

$$\operatorname{rank}\begin{pmatrix} F \\ g^T \end{pmatrix} = \operatorname{rank}(F) + 1.$$

Proof. If g^T were a linear combination of the rows of F, then $g^T = y^T F$ for some vector y. Since $Fp = 0$, substituting $y^T F$ for g^T would give $g^T p = y^T F p = 0$, contradicting our assumption that $g^T p \ne 0$. ∎

Note that the implication in Result 4.2 does not go the other way: if $Fp = 0$, $p \ne 0$, and $g^T p = 0$, then g^T can nonetheless be linearly independent of the rows of F. This can be seen by example:

$$F = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad g^T = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}, \quad \text{and} \quad p^T = \begin{pmatrix} -1 & 1 & 0 & 1 \end{pmatrix}.$$

Result 4.3. *Let D be an $m \times n$ matrix with* $\operatorname{rank}(D) = n$ *whose ith row is d_i^T. Let $\widetilde{D}$ denote a subset of rows of D such that* $\operatorname{rank}(\widetilde{D}) = r < n$*, and assume that p is a nonzero vector such that $\widetilde{D}p = 0$. Then there is at least one row d_j^T of D that is not included in $\widetilde{D}$ such that (i) $d_j^T p \neq 0$ and (ii) d_j^T is linearly independent of the rows of $\widetilde{D}$.*

Proof. Let the $r \times n$ matrix F consist of r linearly independent rows of $\widetilde{D}$, so that every row in $\widetilde{D}$ that is not in F is a non-trivial linear combination of the rows of F. Hence the assumption that $\widetilde{D}p = 0$ implies that $Fp = 0$.

Because $\operatorname{rank}(\widetilde{D}) = r$ and $\operatorname{rank}(D) = n$, we can assemble a matrix G consisting of $n - r$ rows of D that are not in $\widetilde{D}$ and that are linearly independent of the rows of $\widetilde{D}$, such that the $n \times n$ matrix

$$M = \begin{pmatrix} F \\ G \end{pmatrix} \quad \text{is nonsingular.}$$

Since $p \neq 0$, nonsingularity of M means that $Mp \neq 0$ and, since $Fp = 0$, this will be true only if $Gp \neq 0$. Given how G is defined, there must be a row d_j^T in D but not in $\widetilde{D}$ such that $d_j^T p \neq 0$, and linear independence of d_j^T follows directly from Result 4.2. ∎

Our next step is to determine when there is an *optimal* vertex for the LP (1.3). We know that a vertex can exist only if the constraints are consistent and $\operatorname{rank}(A) = n$. Using a theoretical procedure, the next lemma guarantees the existence of an *optimal* vertex under the added assumption that $c^T x$ is bounded below in the feasible region.

Lemma 4.4. (Existence of an optimal vertex.) *Consider minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \geq b_{\mathcal{I}}$, where the rank of A (1.4) is n. Let $\mathcal{V}$ denote the set of all vertices for the given constraints. Then either*

(i) $c^T x$ is bounded below in the feasible region and there is a vertex $v^* \in \mathcal{V}$ where the smallest value of $c^T x$ in the feasible region is achieved; or

(ii) $c^T x$ is unbounded below in the feasible region and there exists an n-vector p such that $A_{\mathcal{E}} p = 0$, $A_{\mathcal{I}} p \geq 0$, and $c^T p < 0$.

Proof. Starting with any feasible point x_0, we define an iterative sequence $\{x_k\}$ that produces a vertex $v_j \in \mathcal{V}$ such that $c^T v_j \leq c^T x_0$, unless we find an n-vector p such that $A_{\mathcal{E}} p = 0$, $A_{\mathcal{I}} p \geq 0$, and $c^T p < 0$. At x_k, $\bar{\bar{A}}_k$ denotes the active-constraint matrix $\bar{\bar{A}}(x_k)$ defined by (2.1); note that the rows of $A_{\mathcal{E}}$ are always present in $\bar{\bar{A}}_k$.

Step 0. Set $k = 0$.

Step 1. If x_k is a vertex (i.e., $\operatorname{rank}(\bar{\bar{A}}_k) = n$) then $x_k = v_j \in \mathcal{V}$ for some j. Stop; a vertex has been found such that $c^T v_j \leq c^T x_0$. Otherwise, go to Step 2.

Step 2. Since $\operatorname{rank}(\bar{\bar{A}}_k) < n$, there exists a nonzero p satisfying $\bar{\bar{A}}_k p = 0$, so that any movement along p does not alter the values of constraints active at x_k. Now we consider the inequality constraints that are inactive at x_k. It follows from Result 4.3 with $\bar{\bar{A}}_k$ playing the role of $\widetilde{D}$ that there must be at least one inactive inequality constraint index j such that $a_j^T p \neq 0$; let $\mathcal{J}_k = \{ j \mid a_j^T x_k > b_k \text{ and } a_j^T p \neq 0 \}$.

Step 3. If $c^T p = 0$, we select $j \in \mathcal{J}_k$ and set $p_k = \pm p$, choosing the sign so that $a_j^T p_k < 0$ (since either choice will satisfy $\bar{\bar{A}}_k p_k = 0$). Otherwise, if $c^T p \neq 0$, choose $p_k = \pm p$ so that, for some $j \in \mathcal{J}_k$, $a_j^T p_k < 0$ and $c^T p_k < 0$; if this is not possible it must hold that $c^T p_k < 0$ and $a_j^T p_k \geq 0$ for all $j \in \mathcal{J}_k$, in which case we exit and conclude that (ii) holds.

Applying Definition 2.2, let $\alpha_k > 0$ be the maximum feasible step along p_k. Then all the constraints inactive at x_k remain feasible at $x_{k+1} = x_k + \alpha_k p_k$, and at least one additional linearly independent inequality constraint becomes active there. Hence $\operatorname{rank}(\bar{\bar{A}}_{k+1}) > \operatorname{rank}(\bar{\bar{A}}_k)$ and $c^T x_{k+1} \leq c^T x_k$.

Step 4. Increase k to $k + 1$ and return to Step 1.

For each initial x_0 there will be no more than n executions of Step 1, since $\operatorname{rank}(A) = n$ and each pass through Step 3 increases the rank of $\bar{A}_k$ by at least one.

This procedure confirms that, for every feasible point x_0, there is either a vertex $v_j \in \mathcal{V}$ such that $c^T v_j \le c^T x_0$ or an n-vector p such that $A_{\mathcal{E}} p = 0$, $A_{\mathcal{I}} p \ge 0$, and $c^T p < 0$. Let v^* denote a vertex such that $c^T v^* \le c^T v_j$ for all v_j in the finite set $\mathcal{V}$. If there is a feasible x_0 such that $c^T x_0 < c^T v^*$, the procedure must give a p such that $A_{\mathcal{E}} p = 0$, $A_{\mathcal{I}} p \ge 0$, and $c^T p < 0$ and (ii) holds. Otherwise, $c^T x_0 \ge c^T v^*$ for all feasible x_0 and (i) holds. ∎

5. Optimality at a nondegenerate vertex

It is straightforward to derive necessary and sufficient conditions for optimality of a *nondegenerate* vertex.

Proposition 5.1. (Optimality of a nondegenerate vertex.) *Consider the linear program of minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$. Assume that x^* is a nondegenerate vertex where the active set is $\mathcal{A}(x^*)$ and that the n-vector $\bar{\bar{\lambda}}$ is the solution of $\bar{A}(x^*)^T \bar{\bar{\lambda}} = c$. Then x^* is optimal if and only if $\bar{\bar{\lambda}}_{\mathcal{I}} \ge 0$, where $\bar{\bar{\lambda}}_{\mathcal{I}}$ denotes the components of $\bar{\bar{\lambda}}$ corresponding to active inequality constraints.*

Proof. Because x^* is nondegenerate, $\bar{A}(x^*)$ is nonsingular, which means that $\bar{\bar{\lambda}}$ is unique. The "if" direction follows because, as we show next, we can define an m-vector λ^* that satisfies the sufficient conditions of Proposition 3.2. Assume that the rows of $\bar{A}(x^*)$ are ordered with indices $\{w_1, \ldots, w_n\}$, so that the jth component of $\bar{\bar{\lambda}}$ is the multiplier for original constraint w_j. The full multiplier λ^* is then defined as

$$\lambda^*_{w_j} = \bar{\bar{\lambda}}_j, \quad j = 1, \ldots, n; \quad \text{and} \quad \lambda^*_j = 0 \quad \text{if } j \notin \mathcal{A}(x^*), \tag{5.1}$$

so that the multipliers corresponding to inactive inequality constraints are zero. Hence $\lambda^{*T}(Ax^* - b) = 0$, the sufficient conditions of Proposition 3.2 are satisfied, and x^* is optimal.

For the "only if" direction, suppose that $[\bar{\bar{\lambda}}_{\mathcal{I}}]_i$ is strictly negative for some active inequality constraint. Because $\bar{A}(x^*)$ is nonsingular, there is a unique direction p satisfying $\bar{A}(x^*)p = e_i$, where e_i is the ith coordinate vector, so that p is a feasible direction. It then follows from the relation $c = \bar{A}(x^*)^T \bar{\bar{\lambda}}$ that $c^T p = \bar{\bar{\lambda}}^T \bar{A}(x^*) p = [\bar{\bar{\lambda}}_{\mathcal{I}}]_i < 0$ and p is a feasible descent direction, which means that x^* cannot be optimal. ∎

6. Optimality at a degenerate vertex

Difficulties arise in proving necessary optimality conditions for a degenerate optimal vertex because Proposition 5.1 depends on nonsingularity of the active-constraint matrix at the vertex x^*. To address these difficulties, we use properties of a *working set*, which is closely related to, but not the same as, the active set; see [11, page 339] for a more restricted definition.

Definition 6.1. (Working set.) *Given the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$, let $\bar{x}$ be a feasible point, not necessarily a vertex. Consider a set of $n_{\mathcal{W}}$ distinct indices, $\mathcal{W} = \{w_1, \ldots, w_{n_{\mathcal{W}}}\}$, where $m_{\mathcal{E}} \le n_{\mathcal{W}} \le n$ and $w_i = i$ for $i = 1, \ldots m_{\mathcal{E}}$, i.e., the first $m_{\mathcal{E}}$ indices in $\mathcal{W}$ are the indices of the equality constraints. Let W be the associated $n_{\mathcal{W}} \times n$ working matrix whose ith row is $a^T_{w_i}$, so that the first $m_{\mathcal{E}}$ rows of W are $A_{\mathcal{E}}$. Let $b_{\mathcal{W}}$ denote the vector consisting of components of b corresponding to the indices in $\mathcal{W}$. Then $\mathcal{W}$ is a* working set *at $\bar{x}$ if the following two properties apply:*

(1) Every inequality constraint whose index is in $\mathcal{W}$ is active at $\bar{x}$, i.e., $W\bar{x} = b_{\mathcal{W}}$;

(2) The rows of W are linearly independent.

At a vertex, by definition there are n linearly independent active constraints, so that it is always possible to define a nonsingular working-set matrix W and an associated unique vector $\lambda_{\mathcal{W}}$ that satisfies $W^T \lambda_{\mathcal{W}} = c$, where component j of $\lambda_{\mathcal{W}}$ corresponds to original constraint w_j. Thus components $1, \ldots, m_{\mathcal{E}}$ of $\lambda_{\mathcal{W}}$

correspond to the $m_{\mathcal{E}}$ equality constraints, and the remaining components are associated with "working" (active) inequality constraints.

If x^* is an optimal nondegenerate vertex, the active set $\bar{\bar{A}}(x^*)$ is a working set with $n_{\mathcal{W}} = n$, $\lambda_{\mathcal{W}}$ is the same as the unique solution $\bar{\bar{\lambda}}$ of $\bar{\bar{A}}(x^*)^T\bar{\bar{\lambda}} = c$, and $[\lambda_{\mathcal{W}}]_i \ge 0$ for $w_i \in \bar{\bar{\mathcal{I}}}(x^*)$. But if $\widehat{x}$ is an optimal vertex that is degenerate, there can be more than one working set. This complicates optimality conditions because, even if $\widehat{\mathcal{W}}$ is a working set at a degenerate optimal vertex $\widehat{x}$, it may not be true that $[\lambda_{\widehat{\mathcal{W}}}]_i \ge 0$ for inequality constraints in the working set.

Consider, for example, the following all-inequality two-variable linear program of minimizing c^Tx subject to three inequality constraints $Ax \ge b$, with

$$c = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ 1 & \frac{5}{2} \\ 1 & -2 \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} 3 \\ 6 \\ -3 \end{pmatrix}. \tag{6.1}$$

(Note that, for this example, there are no equality constraints.) The optimal solution is a degenerate vertex $x^* = (1,2)^T$. There are three working sets, $\mathcal{W}_1 = \{1,2\}$, $\mathcal{W}_2 = \{1,3\}$, and $\mathcal{W}_3 = \{2,3\}$, and the associated multipliers are

$$\lambda_{\mathcal{W}_1} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \lambda_{\mathcal{W}_2} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad \text{and} \quad \lambda_{\mathcal{W}_3} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}.$$

Although $\mathcal{W}_1$ identifies two linearly independent rows of A, optimality of x^* cannot be determined by checking the signs of the components of $\lambda_{\mathcal{W}_1}$.

We therefore define an *optimal working set* at an optimal point as one that will confirm optimality, as happens with working sets $\mathcal{W}_2$ and $\mathcal{W}_3$ in the example.

Definition 6.2. (Optimal working set.) *Given the consistent constraints $A_{\mathcal{E}}x = b_{\mathcal{E}}$ and $A_{\mathcal{I}}x \ge b_{\mathcal{I}}$ and the objective function c^Tx, assume that $\mathcal{W}$ is a working set at the optimal point $\bar{x}$ (not necessarily a vertex), with W the associated working matrix. Then $\mathcal{W}$ is an* optimal working set *if (a) the linear system $W^T\lambda_{\mathcal{W}} = c$ is compatible (which means that $\lambda_{\mathcal{W}}$ exists and is unique), and (b) $[\lambda_{\mathcal{W}}]_i \ge 0$ if w_i is the index of an active inequality constraint.*

Note that the uniqueness mentioned in property (a) follows because the columns of W^T are linearly independent (Definition 6.1).

The next proposition shows that, if the constraints are consistent, $\operatorname{rank}(A) = n$, and c^Tx is bounded below in the feasible region, then an optimal vertex and an associated optimal working set always exist, even if the vertex is degenerate. An optimal working set is obtained from the solution of a *perturbed linear program* where, as in [10], the perturbations reflect motivation introduced in [4]; see, for example, [6] and [5, pages 34–35]. The crucial property of the perturbations is that their presence guarantees existence of an optimal *nondegenerate* vertex for the perturbed problem.

Proposition 6.3. (Existence of an optimal vertex, multiplier, and working set.) *Consider minimizing c^Tx subject to the consistent constraints $A_{\mathcal{E}}x = b_{\mathcal{E}}$ and $A_{\mathcal{I}}x \ge b_{\mathcal{I}}$, where c^Tx is bounded below in the feasible region and* $\operatorname{rank}(A) = n$. *Then there are an optimal vertex x^* and an associated optimal working set $\mathcal{W} = \{w_1, \ldots, w_n\}$ of n indices, such that the corresponding working-set matrix W is nonsingular, from which an m-dimensional multiplier λ^* can be constructed such that (i) $A^T\lambda^* = c$, (ii) $\lambda^{*T}(Ax^* - b) = 0$, and (iii) $\lambda^*_{\mathcal{I}} \ge 0$.*

Proof. The proof has two parts: analyzing a perturbed linear program, and then using the resulting nondegenerate optimal vertex for the perturbed problem to define a solution and optimal working set for the original problem.

Part 1: Solving a perturbed linear program. Consider the perturbed linear program:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^Tx \quad \text{subject to} \quad A_{\mathcal{E}}x = b_{\mathcal{E}} \ \text{ and } \ A_{\mathcal{I}}x \ge b_{\mathcal{I}} - e, \tag{6.2}$$

where $e = (\epsilon, \epsilon^2, \ldots, \epsilon^{m_{\mathcal{I}}})^T$ and $\epsilon > 0$ is arbitrary and "sufficiently small". (Note that the equality constraints are not perturbed.) Because the constraints of the original LP are consistent, so are the constraints of the perturbed problem. The objective $c^T x$ and constraint matrix A are the same in both the original and perturbed problems.

We know from Lemma 4.4 that there must be an optimal vertex, denoted by x_ϵ, for the perturbed problem. Let $\bar{\bar{A}}_\epsilon$ denote the active-constraint matrix at x_ϵ with respect to the perturbed constraints of (6.2). Without loss of generality, since x_ϵ may be degenerate, we write the active constraints at x_ϵ as

$$\bar{\bar{A}}_\epsilon x_\epsilon = \bar{\bar{b}}_\epsilon - \bar{\bar{e}}_\epsilon, \quad \text{with} \quad \bar{\bar{A}}_\epsilon = \begin{pmatrix} W_\epsilon \\ Y_\epsilon \end{pmatrix}, \quad \bar{\bar{b}}_\epsilon = \begin{pmatrix} b_{\mathcal{W}_\epsilon} \\ b_{\mathcal{Y}_\epsilon} \end{pmatrix}, \quad \text{and} \quad \bar{\bar{e}}_\epsilon = \begin{pmatrix} e_{\mathcal{W}_\epsilon} \\ e_{\mathcal{Y}_\epsilon} \end{pmatrix}, \tag{6.3}$$

where W_ϵ is $n \times n$ and nonsingular. Let $\mathcal{W}_\epsilon$ denote the set of n indices $\{w_1, \ldots, w_n\}$ of the original constraints corresponding to the rows of W_ϵ, where the matrix $A_{\mathcal{E}}$ corresponding to the equality constraints occupies the first $m_{\mathcal{E}}$ rows of W_ϵ. The remaining $n - m_{\mathcal{E}}$ rows of W_ϵ and the rows of Y_ϵ contain normals of active inequality constraints whose indices are not known in advance. The first $m_{\mathcal{E}}$ components of $e_{\mathcal{W}_\epsilon}$ are zero (since the equalities are not perturbed), followed by $n - m_{\mathcal{E}}$ distinct powers of ϵ:

$$e_{\mathcal{W}_\epsilon} = (0, \ldots, 0, \epsilon^{w_1}, \epsilon^{w_2}, \ldots, \epsilon^{w_{n-m_{\mathcal{E}}}})^T, \quad \text{with} \quad 1 \le w_i \le m_{\mathcal{I}}, \;\; i = 1, \ldots, n - m_{\mathcal{E}}. \tag{6.4}$$

We next show by contradiction that x_ϵ must be nondegenerate for all sufficiently small ϵ, i.e., that Y_ϵ must be empty. Let y^T denote the normal of an inequality constraint in Y_ϵ, and assume that it corresponds to the jth inequality constraint of the original problem. Because W_ϵ is nonsingular, there is a unique vector q such that $y^T = q^T W_\epsilon$. Consequently, since $W_\epsilon x_\epsilon = b_{\mathcal{W}_\epsilon} - e_{\mathcal{W}_\epsilon}$ (see (6.3)), we have

$$y^T x_\epsilon = q^T W_\epsilon x_\epsilon = q^T (b_{\mathcal{W}_\epsilon} - e_{\mathcal{W}_\epsilon}).$$

By assumption, x_ϵ is an optimal vertex for the perturbed problem, so that $y^T x_\epsilon \ge [b_{\mathcal{I}}]_j - \epsilon^j$ (since constraint j is an inequality). But our further assumption (6.3) that the constraints in Y_ϵ are active at x_ϵ for all sufficiently small ϵ implies that this relation is an *equality*, i.e., $y^T x_\epsilon = [b_{\mathcal{I}}]_j - \epsilon^j$. Substituting $q^T (b_{\mathcal{W}_\epsilon} - e_{\mathcal{W}_\epsilon})$ for $y^T x_\epsilon$ and rearranging, we obtain

$$q^T b_{\mathcal{W}_\epsilon} - [b_{\mathcal{I}}]_j - q^T e_{\mathcal{W}_\epsilon} + \epsilon^j = 0.$$

The left-hand side of this relation is a polynomial in ϵ, in which $q^T b_{\mathcal{W}_\epsilon}$ and $[b_{\mathcal{I}}]_j$ are independent of ϵ and the inner product $q^T e_{\mathcal{W}_\epsilon}$ is a linear combination of the distinct powers of ϵ from (6.4), none of which is equal to j, and there is a term ϵ^j. Such a polynomial can equal zero only when ϵ is exactly equal to one of the polynomial's roots. Hence equality cannot hold when ϵ is allowed to be any arbitrarily small positive value, and we obtain a contradiction. The same argument applies for all the constraints in $\mathcal{Y}_\epsilon$, so that $y^T x_\epsilon > [b_{\mathcal{I}}]_j - \epsilon^j$ for all $j \in \mathcal{Y}_\epsilon$. It follows that Y_ϵ is empty and that only the n constraints in $\mathcal{W}_\epsilon$ are active, confirming that x_ϵ is a nondegenerate optimal vertex with active set $\bar{\bar{A}}_\epsilon = W_\epsilon$.

Letting $\bar{\bar{\lambda}}_\epsilon$ denote the necessarily unique solution of $W_\epsilon^T \bar{\bar{\lambda}}_\epsilon = c$, it follows from the "only if" direction of Proposition 5.1 that the components of $\bar{\bar{\lambda}}_\epsilon$ corresponding to active inequality constraints are nonnegative:

$$W_\epsilon^T \bar{\bar{\lambda}}_\epsilon = c \quad \text{and} \quad [\bar{\bar{\lambda}}_\epsilon]_i \ge 0 \;\; \text{when} \;\; w_i \in \bar{\bar{\mathcal{I}}}_\epsilon. \tag{6.5}$$

Part 2. Defining an optimal solution for the original problem. We now show that the working set $\mathcal{W}_\epsilon$ for the perturbed problem is an *optimal* working set for the original problem; see Definition 6.2. Taking $\mathcal{W} = \mathcal{W}_\epsilon = \{w_1, \ldots, w_n\}$ and $W = W_\epsilon$, we define x^* as the (unique) solution of $W x^* = b_{\mathcal{W}}$, so that the n linearly independent constraints represented in W are active at x^*.

Let y^T denote the normal of any constraint not in W, and assume that it corresponds to the jth inequality constraint in the original problem. It remains to show that $y^T x^* \ge [b_{\mathcal{I}}]_j$, i.e., that x^* is feasible with respect to the corresponding original inequality constraint. The proof of Part 1 shows that there is a unique q such that $y^T = q^T W$ and that $q^T b_{\mathcal{W}} - [b_{\mathcal{I}}]_j - q^T e_{\mathcal{W}} + \epsilon^j > 0$ for all sufficiently small ϵ. Since $y^T x^* = q^T b_{\mathcal{W}}$, we have

$$y^T x^* - [b_{\mathcal{I}}]_j - q^T e_{\mathcal{W}} + \epsilon^j > 0, \tag{6.6}$$

which has two consequences.

(i) A result from [6, Lemma 1, Chapter 10] says that a polynomial in $\epsilon > 0$ will be positive for all sufficiently small ϵ if and only if the coefficient of the smallest power of ϵ is positive. The cited result implies that, if the constant term $y^T x^* - [b_{\mathcal{I}}]_j$ of the polynomial in (6.6) is nonzero, it must be positive, in which case original constraint j is inactive at x^*.

(ii) If $y^T x^* - [b_{\mathcal{I}}]_j = 0$, then by definition original constraint j is active at x^*. (This case applies when x^* is degenerate.)

In either case, $y^T x^* \geq [b_{\mathcal{I}}]_j$ and x^* is feasible with respect to all of the original inequality constraints $A_{\mathcal{I}} x \geq b_{\mathcal{I}}$.

The remaining ingredient needed to verify that x^* and $\mathcal{W}$ are optimal involves multipliers. Since the nonsingular working matrix W has been taken as W_ϵ and $W^T \overline{\lambda}_\epsilon = c$, we can define an m-vector λ^*, where $\lambda^*_{\mathcal{W}}$ denotes the vector of components of λ^* associated with constraints in the working set:

$$\lambda^*_{\mathcal{W}} = \overline{\lambda}_\epsilon \quad \text{and} \quad \lambda^*_i = 0 \;\text{ if }\; i \neq \mathcal{W}, \tag{6.7}$$

noting that $\lambda^*_i \geq 0$ if the associated constraint is an inequality in $\mathcal{W}$; see (6.5). Thus we have obtained an optimal vertex x^* and an optimal working set. Using the optimal working set, a multiplier λ^* can be defined satisfying the sufficient optimality conditions of Proposition 3.2. ∎

The just-completed proof shows that the perturbed LP is guaranteed to have a nondegenerate optimal vertex, but this vertex will in general depend on the value of ϵ and the ordering of the powers of ϵ in the perturbed constraints. This non-uniqueness of x_ϵ and $\mathcal{W}_\epsilon$ is illustrated in Figure 1 for the all-inequality linear program (6.1). The contours of the linear objective are labeled as "ϕ". The optimal degenerate vertex $x^* = (1, 2)^T$ for the original LP is shown on the left, where the constraints include a thin shading on the infeasible side. In the remaining two figures, the constraints have been perturbed and the thickness of the shading reflects the size of the perturbation. The value of ϵ is deliberately taken as $\frac{1}{2}$ so that the effects can easily be seen. In the middle figure, constraints 1, 2, and 3 are perturbed respectively by ϵ, ϵ^2, and ϵ^3, producing a single nondegenerate vertex where constraints 2 and 3 are active. In the rightmost figure, the constraint perturbations are ϵ^2, ϵ, and ϵ^3, creating two distinct nondegenerate vertices, with constraints 1 and 3 active at the optimal vertex (which differs from the optimal vertex in the middle figure). Our earlier analysis of (6.1) showed that the optimal working sets are indeed $\mathcal{W}_3 = \{2, 3\}$ and $\mathcal{W}_2 = \{1, 3\}$, shown respectively in the middle and rightmost figures.

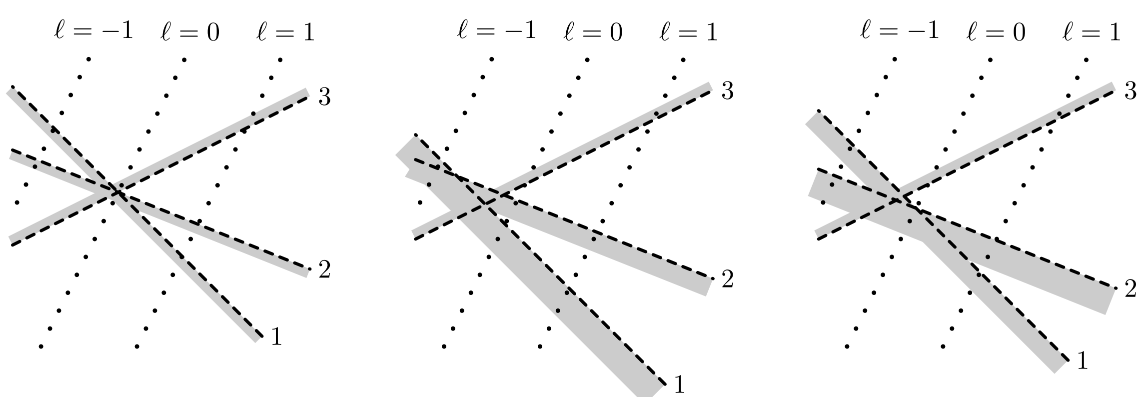

Figure 1: Effects of perturbing the constraints at a degenerate optimal vertex.

7. Necessary and sufficient optimality conditions, version 2

The result of Proposition 6.3 allows us to state necessary and sufficient conditions for optimality of a linear program with the form (1.3) in which the constraints are consistent, $\operatorname{rank}(A) = n$ (where A is defined by (1.4)), and the objective function is bounded below in the feasible region. Note that they apply at any optimal point, whether or not it is a vertex.

Proposition 7.1. (Necessary and sufficient optimality conditions—Version II.) *Consider the linear program of minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$, where $c^T x$ is bounded below in the feasible region and $\operatorname{rank}(A) = n$. The point $\widetilde{x}$, which need not be a vertex, is optimal if and only if $\widetilde{x}$ is feasible and there exists an m-vector $\widetilde{\lambda}$ such that $A^T\widetilde{\lambda} = c$, $\widetilde{\lambda}^T(A\widetilde{x} - b) = 0$ and $\widetilde{\lambda}_{\mathcal{I}} \ge 0$. The optimal objective value is $\widetilde{\lambda}^T b$.*

Proof. The "if" part was proved in Proposition 3.2.

To confirm the "only if" part, we begin by observing that Proposition 6.3 shows that an optimal vertex x^* must exist for the given LP, with an associated optimal working set $\mathcal{W}$ that allows us to define an optimal m-component multiplier λ^* such that $A^T\lambda^* = c$, $\lambda^{*T}(Ax^* - b)$, and $\lambda^*_{\mathcal{I}} \ge 0$; see (6.7). An important point is that $\lambda^*_{\mathcal{I}}$ contains multipliers for all the inequality constraints in the problem. Proposition 3.2 shows that the optimal value is $\lambda^{*T} b$.

Now suppose that the feasible point $\widetilde{x}$ is optimal, where $\widetilde{x}$ may or may not be a vertex. Proposition 3.3 states that $\lambda^{*T}(A\widetilde{x} - b) = 0$ must hold. Hence we can take $\widetilde{\lambda} = \lambda^*$ as a multiplier for $\widetilde{x}$. Again, we stress that optimality of $\widetilde{x}$ follows directly from existence of the multiplier $\widetilde{\lambda}$. ∎

8. Identifying an optimal working set at an optimal vertex

The results proved thus far show that, when the constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$ are consistent, $\operatorname{rank}(A) = n$, and $c^T x$ is bounded below in the feasible region, an optimal vertex $\widehat{x}$ and a corresponding optimal working set $\widehat{\mathcal{W}}$ of n indices such that $\widehat{W}$ is nonsingular must exist (where the working set leads to a multiplier $\widehat{\lambda}$ that ensures optimality). We know from Proposition 7.1 that $\widehat{\lambda}$ is also an optimal multiplier for any optimal point $x^* \ne \widehat{x}$. But the active constraints at $\widehat{x}$ and x^* may be different, which means that the working set $\widehat{\mathcal{W}}$ may not be a valid working set for x^* because $\widehat{W}x^* \ne b_{\widehat{\mathcal{W}}}$; see the example following the proof of Proposition 8.1.

The next result shows that given a specific optimal vertex x^*, a corresponding optimal nonsingular working-set matrix W of of n indices exists satisfying Definition 6.2. The result relies on Proposition 3.3, which shows that the multiplier associated with an optimal vertex satisfies the sufficient optimality conditions for any other optimal point.

Proposition 8.1. (Existence of an optimal working set.) *For the linear program of minimizing $c^T x$ subject to the consistent constraints $A_{\mathcal{E}} x = b_{\mathcal{E}}$ and $A_{\mathcal{I}} x \ge b_{\mathcal{I}}$, where $\operatorname{rank}(A) = n$ and $c^T x$ is bounded below in the feasible region, suppose that an optimal vertex x^* is given. Then there is an associated optimal working set $\mathcal{W}$ of n indices, such that the corresponding working-set matrix W is nonsingular.*

Proof. Proposition 6.3 guarantees existence of an optimal vertex $\widehat{x}$, an optimal working set $\widehat{\mathcal{W}}$ containing n indices such that the corresponding working-set matrix $\widehat{W}$ is nonsingular, and an m-dimensional optimal vector $\widehat{\lambda}$. If $x^* = \widehat{x}$, we can take $\mathcal{W} = \widehat{\mathcal{W}}$ and nothing more is needed. If $x^* \ne \widehat{x}$, we show next how to use $\widehat{\mathcal{W}}$ to construct an optimal working set $\mathcal{W}$ for x^*.

The working set $\widehat{\mathcal{W}}$ contains precisely n indices, which include those of the equality constraints plus a selection of active inequality constraints. Defining $\widehat{\mathcal{W}}_+$ as the set of indices of the equality constraints plus the indices i of inequality constraints with positive multipliers $\widehat{\lambda}_i$, let $\widehat{W}_+$ denote the associated submatrix of $\widehat{W}$, i.e., the matrix whose rows correspond to indices in $\widehat{\mathcal{W}}_+$. Nonsingularity of $\widehat{W}$ implies that $\widehat{W}_+$ has full row rank. Since $\widehat{x}$ and x^* are both optimal, we know from Proposition 3.3 that complementarity is satisfied for all constraints at both $\widehat{x}$ and x^*, which means that, if an inequality constraint in $\widehat{\mathcal{W}}$ has a positive multiplier, then that constraint must be active at both $\widehat{x}$ and x^*. In addition, all equality constraints are satisfied at both $\widehat{x}$ and x^*. We therefore conclude that $\widehat{W}_+$ is a submatrix of $\bar{A}(x^*)$. Defining $\widehat{\mathcal{W}}_0$ as the set of indices i of inequality constraints that are active at x^* for which $\widehat{\lambda}_i = 0$ and letting $\widehat{W}_0$ denote the corresponding matrix, it follows that

$$\bar{\bar{A}}(x^*) = \begin{pmatrix} \widehat{W}_+ \\ \widehat{W}_0 \end{pmatrix}.$$

Consequently, since x^* is a vertex, $\bar{\bar{A}}(x^*)$ has full column rank. As $\widehat{W}_+$ has full row rank, we may therefore create a nonsingular $n \times n$ working-set matrix W as a nonsingular $n \times n$ submatrix of $\bar{\bar{A}}(x^*)$ that contains $\widehat{W}_+$ and let $\mathcal{W}$ denote the associated indices. ∎

For example, consider a three-variable all-inequality LP with six constraints $Ax \ge b$, where

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 5 \\ 3 \\ 4 \\ -2 \\ -\frac{1}{2} \end{pmatrix}, \quad \text{and} \quad c = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}. \tag{8.1}$$

Two degenerate vertices, $x^* = (2, 1, 1)^T$ and $\widehat{x} = (3, \frac{1}{2}, 1)^T$, are optimal, and the optimal objective is $c^T x^* = 7$. Suppose that $\widehat{x}$ is the optimal vertex produced by Proposition 6.3. The active set at $\widehat{x}$ is $\mathcal{A}(\widehat{x}) = \{1, 2, 5, 6\}$, and $\widehat{\mathcal{W}} = \{1, 2, 5\}$ is an optimal working set, with

$$\widehat{W}\widehat{x} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ \frac{1}{2} \\ 1 \end{pmatrix} = b_{\widehat{\mathcal{W}}} = \begin{pmatrix} 1 \\ 5 \\ -2 \end{pmatrix} \quad \text{and} \quad \widehat{\lambda}_{\widehat{\mathcal{W}}} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

The associated 6-component optimal multiplier is $\widehat{\lambda} = (2, 1, 0, 0, 0, 0)^T$. Using the notation in the proof of Proposition 8.1, $n_+ = 2$ and $\widehat{\mathcal{W}}_+ = \{1, 2\}$.

Now consider finding an optimal working set at x^*. As shown in the proof, constraints 1 and 2 must be active at x^* (and indeed they are), but constraints 5 and 6 are not. The active set at x^* is $\mathcal{A}(x^*) = \{1, 2, 3, 4\}$ and $\widehat{W}x^* \ne b_{\widehat{\mathcal{W}}}$, so that $\widehat{\mathcal{W}}$ is not an optimal working set for x^*, even though $\widehat{\lambda}$ is an optimal multiplier.

Constraints 1 and 2 must be part of the working set at x^*. Since $\operatorname{rank}(\bar{\bar{A}}(x^*)) = 3$, we need to add one further constraint which is active at x^* to $\widehat{\mathcal{W}}_+$. For this example, the extra constraint can be taken as constraint 3 or 4. In either case, the optimal multiplier is the same, $\lambda^* = \widehat{\lambda}$, and the linear system $Wx^* = b_{\mathcal{W}}$ is satisfied.

Note that if we seek a working set at a non-vertex optimal point, such as $\widetilde{x} = (\frac{5}{2}, \frac{3}{4}, 1)^T$ in example (8.1), then $\widetilde{\mathcal{W}} = \widehat{\mathcal{W}}_+$ is an optimal working set at $\widetilde{x}$. In fact, $\widehat{\mathcal{W}}_+$ is an optimal working set at any optimal point.

9. A proof of Farkas' lemma

For completeness, we state and prove a common form of Farkas' lemma using the results in this paper. Note that in Farkas' lemma, no requirement is imposed about the rank of the matrix involved.

Lemma 9.1. (Farkas' lemma.) *Given an $m \times n$ matrix A and an n-vector c, precisely one of the following two conditions must be true:*

(1) There exists $y \ge 0$ such that $A^T y = c$;

(2) There exists p such that $Ap \ge 0$ and $c^T p < 0$.

Proof. If y satisfies (1) and p satisfies (2) then $c^T p = y^T Ap$. Because $Ap \ge 0$ and $y \ge 0$, it follows that $y^T Ap \ge 0$, which contradicts the relation $c^T p < 0$ in (2). Hence (1) and (2) cannot both be true.

To show that one of (1) or (2) must be true, we consider the all-inequality linear program

$$\underset{p\in\mathbb{R}^n}{\text{minimize}} \;\; c^Tp \quad \text{subject to} \quad \widetilde{A}p \ge b, \quad \text{with} \quad \widetilde{A} = \begin{pmatrix} A \\ I_n \\ -I_n \end{pmatrix} \;\text{ and }\; b = \begin{pmatrix} 0 \\ -e \\ -e \end{pmatrix}, \tag{9.1}$$

where e denotes $(1, 1, \ldots, 1)^T$. The first m constraints are $Ap \ge 0$ and the last $2n$ constraints are equivalent to requiring that $-1 \le p_i \le 1$ for $i = 1, \ldots, n$.

This LP has the following properties: (i) $\widetilde{A}$ has rank n because of the presence of the two identity matrices, (ii) the constraints $\widetilde{A}p \ge b$ are consistent because $p = 0$ is feasible, and (iii) the feasible region is bounded so the objective function is bounded below. Let p^* denote an optimal solution of (9.1).

Proposition 7.1 implies that there exists a nonnegative optimal multiplier λ, which we may partition as $\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$, where λ_1 is an m-vector and λ_2 and λ_3 are n-vectors. Because λ is an optimal multiplier, we know that $\widetilde{A}^T\lambda = c$ and $c^Tp^* = \lambda^Tb$. Writing out these relations in partitioned form gives

$$\widetilde{A}^T\lambda = A^T\lambda_1 + \lambda_2 - \lambda_3 = c \;\text{ with }\; \lambda_i \ge 0, \quad \text{and} \quad c^Tp^* = \lambda^Tb = -e^T(\lambda_2 + \lambda_3). \tag{9.2}$$

Since $c^Tp = 0$ at the feasible point $p = 0$, the optimal objective value c^Tp^* must be either zero or negative. If $c^Tp^* = 0$, the second relation in (9.2) implies that $-e^T(\lambda_2 + \lambda_3) = 0$. Since λ_2 and λ_3 are both nonnegative, it follows that $\lambda_2 = \lambda_3 = 0$. Consequently, the first relation in (9.2) shows that $A^T\lambda_1 = c$, $\lambda_1 \ge 0$, which means that that case (1) of the lemma holds for $y = \lambda_1$. If $c^Tp^* < 0$, then, since $Ap^* \ge 0$, p^* satisfies relation (2) of the lemma. Consequently, exactly one of (1) and (2) has a solution. ∎

10. Summary

Assume that the constraints $A_{\mathcal{E}}x = b_{\mathcal{E}}$ and $A_{\mathcal{I}}x \ge b_{\mathcal{I}}$ are consistent, $\text{rank}(A) = n$, and c^Tx is bounded below in the feasible region. We have shown that the feasible point x^* is an optimal solution if and only if there exists an optimal multiplier λ^* such that (i) $A^T\lambda^* = c$, (ii) $\lambda^{*T}(Ax^* - b) = 0$, and (iii) $\lambda^*_{\mathcal{I}} \ge 0$. These conditions were derived through elementary proofs of the following sequence of results:

(a) If λ^* exists satisfying (i), (ii), and (iii), x^* is optimal. (Proposition 3.2.)

(b) Let x^* be an optimal point with an associated multiplier λ^* satisfying (i), (ii), and (iii). For any other optimal point $\widetilde{x} \ne x^*$, condition (ii) is satisfied with λ^* and $\widetilde{x}$, i.e., $\lambda^{*T}(A\widetilde{x} - b) = 0$, and $c^T\widetilde{x} = c^Tx^*$. (Proposition 3.3.)

(c) There always exists an optimal vertex $\widehat{x}$ and an associated multiplier $\widehat{\lambda}$ satisfying (i), (ii), and (iii). (Propositions 5.1 and 6.3.)

(d) There must be an optimal multiplier corresponding to any optimal solution. (Proposition 7.1.)

(e) There must be a nonsingular optimal working set corresponding to any optimal vertex. (Proposition 8.1.)

References

[1] Bartl, D. (2007). Farkas' Lemma, other theorems of the alternative, and linear programming in infinite-dimensional spaces: a purely linear-algebraic approach, *Linear and Multilinear Algebra* 55, 327–353.

[2] Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to Linear Optimization*, Athena Scientific, Belmont, Massachusetts.

[3] Broyden, C. G. (1998). A simple algebraic proof of Farkas's Lemma and related theorems, *Optimization Methods and Software* 8, 185–199.

[4] Charnes, A. (1952). Optimality and degeneracy in linear programming, *Econometrica* 20, 160–170.

[5] Chvátal, V. (1983). *Linear Programming*, W. H. Freeman, New York.

[6] Dantzig, G. B. (1963). *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey.

[7] Dax, A. (1997). An elementary proof of Farkas' lemma, *SIAM Review* 39, 503–507.

[8] Ferris, M. C., Mangasarian, O. L., and Wright, S. J. (2007). *Linear Programming with Matlab*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

[9] Fletcher, R. (1987). *Practical Methods of Optimization* (second edition), John Wiley & Sons, Chichester, United Kingdom.

[10] Forsgren, A. (2008). An elementary proof of optimality conditions for linear programming, Report TRITA-MAT-2008-O86, Department of Mathematics, KTH Royal Institute of Technology, Sweden.

[11] Gill, P. E. Murray, W., and Wright, M. H. (1991). *Numerical Linear Algebra and Optimization, Volume 1*, Addison–Wesley, Redwood City, California.

[12] Nemirovski, A. S. and M. J. Todd (2008). Interior-point methods for optimization, *Acta Numerica* 17, 191–234.

[13] Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization* (second edition), Springer, New York.

[14] Roos, C. and Terlaky, T. (1999). Note on a paper of Broyden, *Operations Research Letters* 15, 183–186.

[15] Schrijver, A. (1987). *Theory of Linear and Integer Programming*, John Wiley & Sons, New York, New York.

[16] Svanberg, K. (2008). Farkas' lemma derived by elementary linear algebra, Report TRITA-MAT-2008-OS7, Department of Mathematics, KTH Royal Institute of Technology, Sweden.

[17] Vanderbei, R. J. (1996). *Linear Programming: Foundations and Extensions*, Kluwer Academic Publishers, Boston, Massachusetts.

[18] Wright, M. H. (1992). Interior methods for constrained optimization, *Acta Numerica* 1, 341–407.