

Optimality conditions for inequality-constrained nonlinear programs

Mark S. Gockenbach

1 Introduction

I now turn to a rigorous development of the optimality conditions for

$$\min f(x) \tag{1}$$

$$s.t. \quad h(x) \geq 0, \tag{2}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$. I will assume that x^* is local minimizer of (1–2) and consider a feasible path $x : [0, a] \rightarrow \mathbb{R}^n$ such that $x(0) = x^*$. In contrast to the case of an equality-constrained NLP, it is natural to consider “one-sided” paths since x^* may lie on the boundary of the feasible region. In such a case, $\dot{x}(0)$ may be a feasible direction while $-\dot{x}(0)$ is not.

If x is any feasible path satisfying $x(0) = x^*$ and $\phi : [0, a] \rightarrow \mathbb{R}$ is defined by

$$\phi(t) = f(x(t)), \quad t \in [0, a],$$

then $t = 0$ is a local minimizer of ϕ . Since 0 is an endpoint of $[0, a]$, it need not be the case that $\phi'(0) = 0$. For example, $\phi : [0, a] \rightarrow \mathbb{R}$ defined by $\phi(t) = t$ has a local minimizer at $t = 0$ and yet $\phi'(0) = 1$. As this example suggests, the relevant condition on $\phi'(0)$ is $\phi'(0) \geq 0$. Since

$$\phi'(0) = \nabla f(x^*) \cdot \dot{x}(0),$$

the first form of the necessary condition is:

If $x : [0, a] \rightarrow \mathbb{R}^n$ is a feasible path satisfying $x(0) = x^*$, then

$$\nabla f(x^*) \cdot \dot{x}(0) \geq 0.$$

As in the case of equality constraints, it is not convenient to work directly with feasible paths. The next step, therefore, is to express the optimality condition in terms of feasible directions. Characterizing feasible directions is now considerably more complicated because it is necessary to distinguish between active and inactive constraints. I will use the notation

$$\mathcal{A}(x) = \{i : h_i(x) = 0\}$$

to denote the (indices of the) active constraints at a feasible point x .

If $h_i(x^*) > 0$ (that is, $i \notin \mathcal{A}(x^*)$), then no sufficiently small change in x^* will cause it to violate the i th constraint. In other words, every direction is feasible as far as the i th constraint is concerned. On the other hand, if $h_i(x^*) = 0$ (that is, $i \in \mathcal{A}(x^*)$), then a direction z is feasible only if it points into the feasible set or along (that is, tangent to) its boundary. In other words, z is a feasible direction, with respect to the i th constraint ($i \in \mathcal{A}(x^*)$), if

$$\nabla h_i(x^*) \cdot z \geq 0. \tag{3}$$

This inequality says that either z is tangent to the set $h_i(x) = 0$ at $x = x^*$ ($\nabla h_i(x^*) \cdot z = 0$) or else it points in a direction in which h_i increases ($\nabla h_i(x^*) \cdot z > 0$). To demonstrate the necessity of (3) directly, suppose $x : [0, a] \rightarrow \mathbb{R}^n$ is a feasible path with $x(0) = x^*$ and $\dot{x}(0) = z$. Then

$$\begin{aligned} h_i(x(0)) = 0, \nabla h_i(x(0)) \cdot \dot{x}(0) < 0 &\Rightarrow \frac{d}{dt} [h_i(x(t))]_{t=0} < 0 \\ &\Rightarrow h(x(t)) < 0 \text{ for all } t > 0 \text{ sufficiently small.} \end{aligned}$$

But, by assumption, $h(x(t)) \geq 0$ for all $t \in [0, a]$. This contradiction shows that $\nabla h_i(x^*) \cdot z < 0$ is impossible for a direction z that is feasible at x^* when $h_i(x^*) = 0$; in other words, any feasible direction z at x^* must satisfy $\nabla h_i(x^*) \cdot z \geq 0$ if $h_i(x^*) = 0$.

To summarize, if z is a feasible direction at x^* , then

$$\nabla h_i(x^*) \cdot z \geq 0 \text{ for all } i \in \mathcal{A}(x^*). \quad (4)$$

The set of all feasible directions z at x^* is called the *tangent cone* (to the feasible set) at x^* . I would like it to be the case that (4) characterizes the tangent cone at x^* . However, this is not necessarily true unless the constraint function h satisfies a constraint qualification. The first constraint qualification is simply a statement of the desired property:

Constraint qualification 1.1 *There exists a feasible path $x : [0, a] \rightarrow \mathbb{R}^n$ such that $x(0) = x^*$ and $\dot{x}(0) = z$ if and only if*

$$\nabla h_i(x^*) \cdot z \geq 0 \text{ for all } i \in \mathcal{A}(x^*). \quad (5)$$

Moreover, if (5) holds, then the feasible path x can be chosen to satisfy $h_i(x(t)) = 0$ for all $t \in [0, a]$ for each i such that $\nabla h_i(x^) \cdot z = 0$.*

Next I define a notion of regular point suitable for inequality constraints.

Definition 1.2 *Suppose $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is continuously differentiable and $x^* \in \mathbb{R}^n$ satisfies $h(x^*) \geq 0$. Then x^* is called a regular point of the constraint $h(x) \geq 0$ if*

$$\{\nabla h_i(x^*) : i \in \mathcal{A}(x^*)\}$$

is a linearly independent set.

The proof of the following theorem is very similar to the analogous proof for equality constraints and will be omitted.

Theorem 1.3 *Suppose $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is continuously differentiable and x^* is a regular point of the constraint $h(x) \geq 0$. Then Constraint Qualification 1.1 holds at x^* .*

The reader should notice that, once again, it is the constraints active at x^* that must be taken into account.

Assuming that x^* is a regular point (or, more generally, that Constraint Qualification 1.1 holds), a necessary condition for x^* to be a local minimizer of (1-2) is

$$\nabla h_i(x^*) \cdot z \geq 0 \text{ for all } i \in \mathcal{A}(x^*) \Rightarrow \nabla f(x^*) \cdot z \geq 0. \quad (6)$$

To understand the next step in the development, it may help to remember the analogous reasoning for equality constraints. I showed that, if x^* was a local minimizer under the constraint $g(x) = 0$ and the constraint qualification held at $x = x^*$, then

$$z \in \mathcal{N}(\nabla g(x^*)^T) \Rightarrow \nabla f(x^*) \cdot z = 0.$$

I then applied the Fundamental Theorem of Linear Algebra to write $\nabla f(x^*)$ as a linear combination of the columns of $\nabla g(x^*)$: $\nabla f(x^*) = \nabla g(x^*)^T \lambda^*$.

I need a result, analogous to the Fundamental Theorem of Algebra, that will allow me to put (6) into a more usable form. This result is Farkas's lemma:

Theorem 1.4 (Farkas's lemma) *For any $A \in \mathbb{R}^{n \times p}$, $c \in \mathbb{R}^n$, exactly one of the two problems has a solution:*

1. $A^T x \geq 0$, $c \cdot x < 0$;
2. $c = A\lambda$, $\lambda \geq 0$.

Farkas's lemma is rather difficult to prove; a number of different proofs have been discovered, but none is very elementary. So as not to distract the reader from the argument here, I will give a proof of Farkas's lemma after completing this discussion.

I need the following corollary of Farkas's lemma:

Corollary 1.5 *A matrix $A \in \mathbb{R}^{n \times p}$ and a vector $c \in \mathbb{R}^n$ satisfy*

$$A^T x \geq 0 \Rightarrow c \cdot x \geq 0, \quad (7)$$

if and only if there exists $\lambda \in \mathbb{R}^p$ such that $\lambda \geq 0$ and

$$c = A\lambda. \quad (8)$$

The proof is immediate from (1.4).

To apply this corollary, I define A to be the matrix whose columns are the gradients of the active constraints at x^* and take c to be $\nabla f(x^*)$. Then (6) is equivalent to (7) and hence implies (8); that is, (6) implies

$$\text{There exist } \lambda_i^* \geq 0, i \in \mathcal{A}(x^*), \text{ such that } \nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla h_i(x^*). \quad (9)$$

Defining $\lambda_i^* = 0$ for $i \notin \mathcal{A}(x^*)$, the necessary condition can be written as follows:

Theorem 1.6 *Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}$ are continuously differentiable, and x^* is a local minimizer of*

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) \geq 0. \end{aligned}$$

If Constraint Qualification 1.1 holds at x^ , then there exists $\lambda^* \in \mathbb{R}^p$ such that x^* and λ^* satisfy*

$$\begin{aligned} \nabla f(x^*) &= \nabla h(x^*) \lambda^*, \\ h(x^*) &\geq 0, \\ \lambda^* &\geq 0, \\ \lambda_i^* h_i(x^*) &= 0 \text{ for all } i = 1, 2, \dots, p. \end{aligned}$$

The last condition is referred to as complementarity.

Of course, Theorem 9 is more useful than (6), since it expresses the necessary condition in terms of a system of equations. Defining the Lagrangian ℓ to be the function

$$\ell(x; \lambda) = f(x) - \lambda \cdot h(x),$$

the first-order necessary condition can be written as

$$\nabla \ell(x^*; \lambda^*) = 0,$$

subject to the nonnegativity constraints on λ^* and the complementarity condition.

Recalling my discussion of the logarithmic penalty function, the reader will notice that Theorem 9 is exactly what I deduced by assuming the algorithm converged.

2 Second-order necessary conditions

Second-order optimality conditions are complicated by the fact that it is necessary to distinguish not only between active and inactive constraints, but also between active constraints corresponding to positive Lagrange multipliers and active constraints corresponding to zero Lagrange multipliers. The basic principle determining the second-order necessary condition is this: If z is a feasible direction at x^* and f has zero slope in the direction of z (that is, if $\nabla f(x^*) \cdot z = 0$), then the curvature of the Lagrangian¹ must be nonnegative in that direction:

$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z \geq 0.$$

I can determine the directions in which f has zero slope by using the formula

$$\nabla f(x^*) = \nabla h(x^*) \lambda^*.$$

This shows that

$$\begin{aligned} \nabla f(x^*) \cdot z &= (\nabla h(x^*) \lambda^*) \cdot z \\ &= \lambda^* \cdot (\nabla h(x^*)^T z) \\ &= \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) \cdot z \\ &= \sum_{\lambda_i^* \neq 0} \lambda_i^* \nabla h_i(x^*) \cdot z. \end{aligned}$$

For any $i \notin \mathcal{A}(x^*)$, $\lambda_i^* = 0$ must hold because of the complementarity condition:

$$\lambda_i^* h_i(x^*) = 0 \text{ for all } i = 1, 2, \dots, p.$$

However, it may be the case that $\lambda_i^* = 0$ for certain $i \in \mathcal{A}(x^*)$ (if both λ_i^* and $h_i(x^*)$ are zero). In this regard, it is helpful to define the *strict complementarity* condition:

$$\text{Exactly one of } \lambda_i^* \text{ and } h_i(x^*) \text{ is zero.} \tag{10}$$

I define $\bar{\mathcal{A}}(x^*)$ to be the subset of $\mathcal{A}(x^*)$ consisting of those indices i such that strict complementarity holds for constraint i . If $\bar{\mathcal{A}}(x^*) = \mathcal{A}(x^*)$, then it is said simply that x^* satisfies the strict complementarity condition.

Using this notation,

$$\nabla f(x^*) \cdot z = \sum_{i \in \bar{\mathcal{A}}(x^*)} \lambda_i^* \nabla h_i(x^*) \cdot z.$$

Since $\lambda_i^* > 0$ for all $i \in \bar{\mathcal{A}}(x^*)$ and $\nabla h_i(x^*) \cdot z \geq 0$ for all $i \in \mathcal{A}(x^*)$, it follows that

$$\nabla f(x^*) \cdot z = 0$$

for a feasible direction z if and only if

$$\begin{aligned} \nabla h_i(x^*) \cdot z &= 0 \text{ for all } i \in \bar{\mathcal{A}}(x^*), \\ \nabla h_i(x^*) \cdot z &\geq 0 \text{ for all } i \in \mathcal{A}(x^*) \setminus \bar{\mathcal{A}}(x^*), \end{aligned}$$

that is, if and only if

$$\begin{aligned} \nabla h_i(x^*) \cdot z &= 0 \text{ for all } i \in \mathcal{A}(x^*) \text{ such that } \lambda_i^* > 0, \\ \nabla h_i(x^*) \cdot z &\geq 0 \text{ for all } i \in \mathcal{A}(x^*) \text{ such that } \lambda_i^* = 0. \end{aligned}$$

I define $D(x^*, \lambda^*)$ to be the set of all vectors z satisfying these conditions.

This discussion suggests the following result:

¹The reader will recall that it is the curvature of the Lagrangian ℓ , not just the objective function f , that is important. This is because if x is a feasible path, then the curvature of $\phi(t) = f(x(t))$ is determined both by the curvature of f and by the curvature of the constraints. The Hessian of ℓ accounts for both sources of curvature.

Theorem 2.1 Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are continuously differentiable, x^* is a local minimizer of

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) \geq 0, \end{aligned}$$

and λ^* is a corresponding Lagrange multiplier. If Constraint Qualification (1.1) is satisfied at x^* , then

$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z \geq 0 \text{ for all } z \in D(x^*, \lambda^*).$$

Proof: Suppose there exists $z \in D(x^*, \lambda^*)$ such that

$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z < 0. \tag{11}$$

Let $x : [0, a] \rightarrow \mathbb{R}^n$ be a feasible path satisfying $x(0) = x^*$, $\dot{x}(0) = z$ and define $\phi : [0, a] \rightarrow \mathbb{R}$ by

$$\phi(t) = f(x(t)).$$

Then

$$\phi'(0) = \nabla f(x^*) \cdot z = 0$$

since $z \in D(x^*, \lambda^*)$ and (11) implies that $\phi''(0) < 0$. Taylor's theorem applied to ϕ then shows that

$$\phi(t) < \phi(0) \text{ for all } t \text{ sufficiently small,}$$

that is,

$$f(x(t)) < f(x^*) \text{ for all } t \text{ sufficiently small.}$$

This contradicts the local optimality of x^* and shows that (11) cannot hold. QED

3 Second-order sufficient conditions

The following theorem gives sufficient conditions for x^* to be a strict local minimizer of (1–2).

Theorem 3.1 Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are continuously differentiable, and $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^p$ satisfy

$$\begin{aligned} \nabla f(x^*) &= \nabla h(x^*) \lambda^*, \\ h(x^*) &\geq 0, \\ \lambda^* &\geq 0, \\ \lambda_i^* h_i(x^*) &= 0 \text{ for all } i = 1, 2, \dots, p, \\ z \cdot \nabla^2 \ell(x^*; \lambda^*) z &> 0 \text{ for all } z \in D(x^*, \lambda^*), z \neq 0. \end{aligned}$$

Then x^* is a strict local minimizer of

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) \geq 0, \end{aligned}$$

and λ^* is a corresponding Lagrange multiplier.

The proof will be omitted.

4 Convex programs

Unlike the case of equality constraints, many nonlinear inequality constraints give rise to a convex feasible set. This is because of the following theorem:

Theorem 4.1 *Suppose $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function. Then*

$$C = \{x \in \mathbb{R}^n : h(x) \leq 0\}$$

is a convex set.

Proof: Suppose $x, y \in C$ and α, β are positive real numbers such that $\alpha + \beta = 1$. Then, by definition,

$$h(x) \leq 0, \quad h(y) \leq 0 \tag{12}$$

and, by the convexity of h ,

$$h(\alpha x + \beta y) \leq \alpha h(x) + \beta h(y). \tag{13}$$

Since $\alpha, \beta \geq 0$, (12) and (13) imply that

$$h(\alpha x + \beta y) \leq 0,$$

that is, $\alpha x + \beta y \in C$. Therefore C is convex. QED

The following theorem is also needed:

Theorem 4.2 *Let A be any set and suppose that C_a is a convex set for each $a \in A$. Then*

$$C = \bigcap_{a \in A} C_a$$

is convex.

Proof: Suppose $x, y \in C$ and α, β are positive real numbers such that $\alpha + \beta = 1$. Then, by definition of C , $x, y \in C_a$ for all $a \in A$. By the convexity of each C_a , it follows that $\alpha x + \beta y \in C_a$ for all $a \in A$. But then

$$\alpha x + \beta y \in \bigcap_{a \in A} C_a = C.$$

QED

Putting the previous two results together yields the following corollary:

Corollary 4.3 *Suppose $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ has the property that each component h_i is a convex function. Then*

$$\{x \in \mathbb{R}^n : h(x) \leq 0\}$$

is a convex set.

For convenience, when treating inequality-constrained NLPs, I have written the inequality constraints as $h(x) \geq 0$. (This is convenient because it causes the Lagrange multipliers to be nonnegative.) Therefore, the NLP

$$\min f(x) \tag{14}$$

$$s.t. \quad h(x) \geq 0 \tag{15}$$

is a convex program if f is convex and each component of $-h$ is convex, that is, if each component of h is *concave*. In the interests of conciseness, I will refer to a vector-valued function h as convex (or concave) if each of its components is convex (or concave).

Assuming, then, that f is convex and h is concave, I have already proved a general theorem that shows that every local minimizer of (14–15) is in fact a global minimizer. I now prove the stronger result that the first-order necessary conditions are actually sufficient for an inequality-constrained convex program. So I suppose that x^* , λ^* satisfy the first-order necessary conditions for (14–15). Then x^* is a stationary point of the Lagrangian $\ell(\cdot; \lambda^*)$. Since

$$\ell(x; \lambda^*) = f(x) - \lambda^* \cdot h(x) = f(x) + \sum_{i=1}^p \lambda_i^* (-h_i(x)),$$

the following theorem shows that $\ell(\cdot, \lambda^*)$ is convex:

Theorem 4.4 *If $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex for $i = 1, 2, \dots, k$ and c_i is nonnegative for $i = 1, 2, \dots, k$, then $f : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by*

$$f = \sum_{i=1}^k c_i f_i$$

is also convex.

The proof is immediate from the definition of convexity and will be omitted.

Since $\ell(\cdot; \lambda^*)$ is convex and x^* is a stationary point, it follows that

$$\ell(x^*; \lambda^*) \leq \ell(x; \lambda^*) \text{ for all } x \in \mathbb{R}^n.$$

But $f(x^*) = \ell(x^*)$ by the complementarity condition and, since $\lambda^* \geq 0$ and $h(x) \geq 0$ for all feasible x ,

$$\ell(x; \lambda^*) \leq f(x) \text{ for all feasible } x.$$

Then, for any feasible x ,

$$f(x^*) = \ell(x^*; \lambda^*) \leq \ell(x; \lambda^*) \leq f(x),$$

which shows that x^* is a global minimizer of the convex program.

5 Proof of Farkas's lemma

There are many proofs of Farkas's lemma in the literature; the reader may consult Dax [3] or Broyden [1] for two recent proofs based on elementary ideas. If A is any $n \times p$ matrix and $c \in \mathbb{R}^n$, then Farkas's lemma states that exactly one of the following two problems has a solution:

$$A^T x \geq 0, \quad c \cdot x < 0, \tag{16}$$

$$c = A\lambda, \quad \lambda \geq 0. \tag{17}$$

If $x \in \mathbb{R}^n$ satisfies (16) and $\lambda \in \mathbb{R}^p$ satisfies (17), then

$$c \cdot x = (A\lambda) \cdot x = \lambda \cdot (A^T x) \geq 0$$

(since $\lambda \geq 0$, $A^T x \geq 0$). But this contradicts that $c \cdot x < 0$. Thus it is not possible to solve both (16) and (17).

To show that one of (16) or (17) must have a solution, I will consider the constrained least-squares problem

$$\min \quad \|A\lambda - c\|^2 \tag{18}$$

$$s.t. \quad \lambda \geq 0. \tag{19}$$

Obviously (17) has a solution if and only if (18–19) has a *zero-residual* solution, that is, a solution λ with $A\lambda = c$.

The following lemma gives the optimality conditions for (18–19). These conditions are precisely the first-order conditions described above, and they are necessary and sufficient because (18–19) is a convex problem. However, so as not to engage in circular reasoning (after all, Farkas’s lemma was used above to derive the optimality conditions), the necessity and sufficiency of these conditions are proved directly.

Theorem 5.1 *Suppose $A \in \mathbb{R}^{n \times p}$ and $c \in \mathbb{R}^n$. Then λ solves (18–19) if and only if*

$$\begin{aligned} \lambda &\geq 0, \\ A^T r &\geq 0, \text{ where } r = A\lambda - c, \\ \lambda_i (A^T r)_i &= 0, \quad i = 1, 2, \dots, p. \end{aligned}$$

Proof: Suppose first that λ solves (18–19), let $a^{(i)}$ denote the i th column of A , and let $e^{(i)}$ denote the i th standard basis vector (that is, the i th column of the $p \times p$ identity matrix). Define $f_i : \mathbb{R} \rightarrow \mathbb{R}$ by

$$f_i(\theta) = \|A(\lambda + \theta e^{(i)}) - c\|^2 = \|\theta a^{(i)} - (A\lambda - c)\|^2 = \|\theta a^{(i)} - r\|^2.$$

Clearly $\theta = 0$ is the solution of

$$\begin{aligned} \min \quad & f_i(\theta) \\ \text{s.t.} \quad & \lambda_i + \theta \geq 0. \end{aligned}$$

If $\lambda_i > 0$, then $f'_i(0) = 0$ must hold; otherwise, $f'_i(0) \geq 0$ holds. But $f'_i(0) = 2a^{(i)} \cdot r$. Therefore, for each $i = 1, 2, \dots, p$, $a^{(i)} \cdot r \geq 0$, that is, $A^T r \geq 0$. Moreover, the complementarity condition $\lambda_i (A^T r)_i = 0$, $i = 1, 2, \dots, p$, holds. This proves the necessity of the proposed optimality conditions.

Suppose, on the other hand, that λ satisfies the given conditions. If $\tilde{\lambda} \geq 0$ and $u = \tilde{\lambda} - \lambda$, then

$$\begin{aligned} \lambda_i = 0 &\Rightarrow u_i \geq 0, \\ \lambda_i > 0 &\Rightarrow (A^T r)_i = 0, \end{aligned}$$

and hence, since $(A^T r)_i \geq 0$ for all i ,

$$u \cdot (A^T r) \geq 0.$$

But then

$$\begin{aligned} \|A\tilde{\lambda} - c\|^2 &= \|A\lambda + Au - c\|^2 \\ &= \|A\lambda - c\|^2 + 2Au \cdot (A\lambda - c) + \|Au\|^2 \\ &= \|A\lambda - c\|^2 + 2u \cdot (A^T r) + \|Au\|^2 \\ &\geq \|A\lambda - c\|^2. \end{aligned}$$

This proves the sufficiency of the given conditions. QED

If the constrained least-square problem (18–19) has a solution λ , then either $r = A\lambda - c = 0$ or $r \neq 0$. In the first case, there is a solution to (17). In the second case, $A^T r \geq 0$ and

$$\begin{aligned} c \cdot r &= (A\lambda - r) \cdot r \quad (\text{since } c = A\lambda - r) \\ &= (A\lambda) \cdot r - r \cdot r \\ &= \lambda \cdot A^T r - r \cdot r \\ &= -r \cdot r \quad (\text{since } \lambda \cdot A^T r = 0 \text{ by the complementarity condition}) \\ &< 0. \end{aligned}$$

Then $x = r$ is a solution of (16).

Therefore, to prove Farkas’s lemma, it remains only to show that the constrained least-square problem (18–19) has a solution.

Lemma 5.2 *If $A \in \mathbb{R}^{n \times p}$, $c \in \mathbb{R}^n$, and the set*

$$S = \{A\lambda : y \geq 0\}$$

is closed, then (18–19) has a solution.

Proof: If (18–19) has a solution λ , then $A\lambda$ lies in the ball

$$\overline{B}_{\|c\|}(c) = \{x \in \mathbb{R}^n : \|x - c\| \leq \|c\|\},$$

(since otherwise $\|A0 - c\| < \|A\lambda - c\|$). Assuming S is closed, the set

$$S_{\|c\|} = S \cap \overline{B}_{\|c\|}(c)$$

is closed and bounded, and hence by Weierstrass's theorem the continuous function $f(x) = \|x - c\|$ attains its minimum over $S_{\|c\|}$, say at $x^* = A\lambda$. But then λ is a solution to (18–19). QED

It remains only to prove that S is closed. The following proof is taken from Ciarlet [2].

Lemma 5.3 *For any $A \in \mathbb{R}^{n \times p}$, the set*

$$S = \{A\lambda : y \geq 0\}$$

is closed.

Proof: I will prove the result first under the assumption that the columns of A are linearly independent. Suppose $\{s^{(k)}\}$ is a sequence in S converging to some vector s . By definition of S , there exists a sequence $\{\lambda^{(k)}\}$ in \mathbb{R}^p such that $\lambda^{(k)} \geq 0$ and $s^{(k)} = A\lambda^{(k)}$ for all k . Then $\|A\lambda^{(k)}\|^2$ is uniformly bounded, and

$$\|A\lambda^{(k)}\|^2 = \lambda^{(k)} \cdot A^T A \lambda^{(k)} \geq \lambda_{\min}(A^T A) \|\lambda^{(k)}\|^2. \quad (20)$$

Since A has full rank by assumption, $A^T A$ is positive definite and so $\lambda_{\min}(A^T A)$ is positive. Therefore, (20) implies that $\|\lambda^{(k)}\|$ is uniformly bounded. Without loss of generality, I can assume that $\lambda^{(k)} \rightarrow \lambda$ for some λ . Since $\lambda^{(k)} \geq 0$ for all k , it follows that $\lambda \geq 0$, and

$$A\lambda^{(k)} \rightarrow s, \quad A\lambda^{(k)} \rightarrow A\lambda$$

both hold. Therefore, $s = A\lambda$, $\lambda \geq 0$, and the result is proved in this case.

I will now show that if the columns of A are linearly dependent, then S can be expressed as a finite union of sets of the form $\{By : y \geq 0\}$, where $B \in \mathbb{R}^{n \times k}$ has linearly independent columns. Then, by the above reasoning, each of these sets is closed, and a finite union of closed sets is closed.

I denote the columns of A by $a^{(1)}, a^{(2)}, \dots, a^{(p)}$. By assumption, there exist numbers $\mu_1, \mu_2, \dots, \mu_p$, not all zero, such that

$$\sum_{i=1}^p \mu_i a^{(i)} = 0.$$

Clearly, multiplying by -1 is necessary, the scalars $\mu_1, \mu_2, \dots, \mu_p$ can be chosen so that at least one of the numbers is negative. I define $J = \{i : \mu_i < 0\}$ and, for any $\lambda \geq 0$,

$$t = \min \left\{ -\frac{\lambda_i}{\mu_i} : i \in J \right\}.$$

It follows that

$$A\lambda = \sum_{i=1}^p \lambda_i a^{(i)} = \sum_{i=1}^p (\lambda_i + t\mu_i) a^{(i)},$$

and, by construction, $\lambda_i + t\mu_i$ is nonnegative for each i . Moreover, for at least one value of j , $\lambda_j + t\mu_j = 0$, and therefore $A\lambda$ can be expressed using at most $p-1$ of the vectors $a^{(1)}, a^{(2)}, \dots, a^{(p)}$. This is true for every $\lambda \geq 0$, and therefore

$$S = \bigcup_{j=1}^p \left\{ \sum_{\substack{i=1 \\ i \neq j}}^p \lambda_i a^{(i)} : \lambda_i \geq 0 \text{ for all } i \neq j \right\}.$$

This argument can be applied to each set

$$\left\{ \sum_{\substack{i=1 \\ i \neq j}}^p \lambda_i a^{(i)} : \lambda_i \geq 0 \text{ for all } i \neq j \right\}$$

such that $\{a^{(i)} : i \neq j\}$ is linearly dependent, and then again as necessary, until finally S is expressed in terms of a finite number of closed sets. QED

References

- [1] C. G. Broyden. A simple algebraic proof of Farkas's lemma and related theorems. 1998.
- [2] Philippe G. Ciarlet. *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge University Press, Cambridge, 1989.
- [3] Achiya Dax. An elementary proof of Farkas' lemma. *SIAM Review*, 39(3):503–507, 1997.