5. LAGRANGE MULTIPLIERS

Optimality with respect to minimization over a set $C \subset I\!R^n$ has been approached up to now in terms of the tangent cone $T_C(\bar{x})$ at a point $\bar{x}$. While this has led to important results, further progress depends on introducing, in tandem with tangent vectors, a notion of normal vectors in a broader elaboration of variational geometry. Representations of normal vectors to C relative to an explicit constraint representation for C will yield special coefficients, called Lagrange multipliers, which not only serve in the statement of necessary and sufficient conditions for a minimum but take on an intriguing life of their own.

Normal vectors: For a closed set $C \subset I\!R^n$ and a point $\bar{x} \in C$, a vector v is said to be *normal to C at $\bar{x}$ in the regular sense* (a "regular normal") if $v{\cdot}w \leq 0$ for all $w \in T_C(\bar{x})$. It is *normal to C in the general sense* (a "general normal," or just a "normal") if it can be approximated by normals in the regular sense: there exist $v^\nu \to v$ and $x^\nu \to \bar{x}$ such that v^ν is a regular normal to C at x^ν.

Interpretation of regular normals: The normal vectors v to C at $\bar{x}$ in the regular sense, apart from $v = 0$, are the vectors that make a right or obtuse angle with every tangent vector w to C at $\bar{x}$. It's not hard to show that this holds if and only if

$$v{\cdot}(x - \bar{x}) \leq o\big(|x - \bar{x}|\big) \quad \text{for } x \in C.$$

The definition of a regular normal vector v could therefore just as well be given in terms of this inequality property.

Regular normals as general normals: Any normal vector v in the regular sense is in particular a normal vector in the general sense. (Consider $v^\nu \equiv v$ and $x^\nu \equiv \bar{x}$).

Regularity of a set: The set C is called *regular* at $\bar{x}$ (one of its points) if every normal at $\bar{x}$ in the general sense is in fact a normal at $\bar{x}$ in the regular sense, i.e., if the limit process in the definition doesn't yield any more normals than are already obtained by the angle condition in the definition or regular normals.

Example of irregularity: A heart-shaped set C fails to be regular at its "inward corner point," although it's regular everywhere else.

Normal cone at a point: For a closed set C, the set of all vectors v that are normal to C at a point $\bar{x} \in C$ in the general sense is called the *normal cone* to C at $\bar{x}$ and is denoted by $N_C(\bar{x})$.

Basic properties: Always, $N_C(\bar{x})$ contains the vector 0. For every $v \in N_C(\bar{x})$ and every $\lambda > 0$, the vector λv is again in $N_C(\bar{x})$. Thus, $N_C(\bar{x})$ is truly a "cone."

Extreme cases: If $C = I\!R^n$, one has $N_C(\bar{x}) = \{0\}$. On the other hand, if C is a one-element set $\{a\}$ and $\bar{x} = a$, then $N_C(\bar{x}) = I\!R^n$. In both cases, C is regular.

Limits of general normals: The cone $N_C(\bar{x})$ always closed, and a stronger property even holds: if $v^\nu \to v$ with $v^\nu \in N_C(x^\nu)$ and $x^\nu \to \bar{x}$, then $v \in N_C(\bar{x})$.

Argument: This stems from the observation that the set of vector pairs $(x, v) \in I\!R^n \times I\!R^n$ such that v is a general normal to C at x is by definition the closure of the set of pairs (x, v) such that v is a regular normal to C at x. Hence in particular, it's a closed set in $I\!R^n \times I\!R^n$.

Normals to convex sets: A closed, convex set $C \subset I\!R^n$ is regular at every one of its points $\bar{x}$. The normal cone $N_C(\bar{x})$ consists of the vectors v such that

$$v \cdot (x - \bar{x}) \leq 0 \quad \text{for all } x \in C.$$

Proof. This follows from the earlier characterization of $T_C(\bar{x})$ in the convex case as consisting of 0 and all vectors obtainable as limits of vectors $w \neq 0$ giving feasible directions in C at $\bar{x}$. From that characterization we have that v is a regular normal at $\bar{x}$ if and only if $v \cdot w \leq 0$ for all w giving feasible directions, which because of convexity means that w is a positive multiple of a difference vectors $x - \bar{x}$ with $x \in C$, $x \neq \bar{x}$. Thus, v is a regular normal if and only if it satisfies $v \cdot (x - \bar{x}) \leq 0$ for all $x \in C$.

To verify the regularity of C at $\bar{x}$, consider a sequence of points $x^\nu \to \bar{x}$ in C and regular normals v^ν to C at these points x^ν, with v^ν converging to a vector v. Is v again a *regular* normal to C at $\bar{x}$? From what has already been demonstrated, we know that for each ν we have $v^\nu \cdot (x - x^\nu) \leq 0$ for all $x \in C$. Taking the limit in this inequality as $\nu \to \infty$ with x fixed, we get $v \cdot (x - \bar{x}) \leq 0$. This being true for an arbitrary choice of $x \in C$, we conclude, by the same criterion once more, that v really is a regular normal to C at $\bar{x}$.

Half-space property: This characterization of normals to a closed, convex set C has the following geometric interpretation. A vector $v \neq 0$ belongs to $N_C(\bar{x})$ for such a set C if and only if C is included in the closed half-space $\{x \mid v \cdot x \leq v \cdot \bar{x}\}$, which has $\bar{x}$ on its boundary hyperplane and v as an outward pointing normal.

Normals to linear subspaces: For a subspace M of $I\!R^n$, one has at every point $\bar{x} \in M$ that $N_M(\bar{x}) = M^\perp$, the set of vectors v such that $v \cdot w = 0$ for all $w \in M$. This is immediate from the fact M is convex, using the characterization above.

THEOREM 9 (basic normal cone condition for optimality). *Consider the problem of minimizing a function f_0 of class $\mathcal{C}^2$ over a closed set $C \subset I\!R^n$. Let $\bar{x} \in C$.*

(a) (necessary). *If $\bar{x}$ is locally optimal, then $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$.*

(b) (sufficient). *If $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$ with C and f_0 convex, then $\bar{x}$ is globally optimal.*

Proof. To prove (a), we'll demonstrate that $-\nabla f_0(\bar{x})$ is actually a regular normal to C at $\bar{x}$: $-\nabla f_0(\bar{x}) {\cdot} w \leq 0$ for all $w \in T_C(\bar{x})$. This inequality is trivial when $w = 0$, so consider any vector $w \neq 0$ in $T_C(\bar{x})$. By the definition of the tangent cone there must be a sequence of points $x^\nu \in C$ along with scalars $\tau^\nu \searrow 0$ such that the vectors $w^\nu = (x^\nu - \bar{x})/\tau^\nu$ converge to w; her we can suppose $x^\nu \neq \bar{x}$ inasmuch as $w \neq 0$. Eventually $f_0(x^\nu) \geq f_0(\bar{x})$ by the local optimality of $\bar{x}$. Then

$$0 \leq \frac{f_0(\bar{x} + \tau^\nu w^\nu) - f_0(\bar{x})}{\tau^\nu} \to \nabla f_0(\bar{x}) {\cdot} w$$

by the differentiability of f_0. This gives the desired inequality.

To prove (b) we simply note that in view of the description of normal cones to convex sets given above, this is no more than a restatement of the sufficient condition of Theorem 7(b) in different notation. □

Versatility of the normal cone condition: Although the condition $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$ says no more than the condition $\nabla f_0(\bar{x}) {\cdot} w \geq 0$ for all $w \in T_C(\bar{x})$ that we worked with earlier, and even a bit less if C isn't regular at $\bar{x}$, it has both conceptual and technical advantages. The tangent cone condition refers to an apparently infinite family inequalities being satisfied, which is cumbersome to think about, whereas the normal cone condition focuses on membership in a certain cone $N_C(\bar{x})$ whose nature for the common kinds of sets C might be ascertained in advance. A *calculus* of normals can be built up for this purpose. The following are some initial examples.

Fermat's rule as a special case: At any point $\bar{x}$ lying in the interior of C, one has $N_C(\bar{x}) = \{0\}$; there's no normal other than the zero vector. Therefore, if f_0 has a local minimum relative to C at such a point, it's necessary that

$$\nabla f_0(\bar{x}) = 0.$$

Argument: At an interior point $\bar{x}$, every direction is a feasible direction, so that $T_C(\bar{x}) = I\!R^n$. Moreover, this holds at all points x in a neighborhood of $\bar{x}$ too. At all such points, therefore, the only regular normal is $v = 0$, and it follows from the definition of $N_C(\bar{x})$ that $N_C(\bar{x}) = \{0\}$.

Normals to boxes: If $X = I_1 \times \cdots \times I_n$ for closed intervals I_j in $I\!R$, then X is regular at any of its points $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ (by virtue of closedness and convexity), and

$$N_X(\bar{x}) = N_{I_1}(\bar{x}_1) \times \cdots \times N_{I_n}(\bar{x}_n), \text{ where}$$
$$N_{I_j}(\bar{x}_j) = \begin{cases} (-\infty, 0] & \text{if } \bar{x}_j \text{ is the left endpoint (only) of } I_j, \\ [0, \infty) & \text{if } \bar{x}_j \text{ is the right endpoint (only) of } I_j, \\ [0, 0] & \text{if } \bar{x}_j \text{ lies in the interior of } I_j, \\ (-\infty, \infty) & \text{if } I_j \text{ is a one-point interval, consisting just of } \bar{x}_j. \end{cases}$$

In other words, the condition $z \in N_X(\bar{x})$ for $z = (z_1, \ldots, z_n)$ constitutes a list of *sign restrictions* on the coordinates of z. Depending on the mode in which $\bar{x}_j$ fulfills the constraint $x_j \in I_j$, each z_j is designated as nonpositive, nonnegative, zero, or free.

Normals to the nonnegative orthant: For $\bar{x} = (\bar{x}_1, \ldots \bar{x}_n)$ in the box $I\!R^n_+$, the normal vectors $z = (z_1, \ldots, z_n)$ are given by

$$z \in N_{\mathbf{R}^n_+}(\bar{x}) \iff \begin{cases} z_j \le 0 & \text{for all } j \text{ with } \bar{x}_j = 0, \\ z_j = 0 & \text{for all } j \text{ with } \bar{x}_j > 0. \end{cases}$$

Tangents from normals under regularity: When a closed set C is regular at $\bar{x}$, the geometric relationship between tangents and normals is actually symmetric:

$$\begin{aligned} N_C(\bar{x}) &= \big\{ v \,\big|\, v{\cdot}w \le 0 \text{ for all } w \in T_C(\bar{x}) \big\}, \\ T_C(\bar{x}) &= \big\{ w \,\big|\, v{\cdot}w \le 0 \text{ for all } v \in N_C(\bar{x}) \big\}. \end{aligned}$$

Proof: We won't really need this fact, but the relationship is so striking that it's worth supplying the proof, for the record. The first of the two equations here just combines the definition of regular normals with the regularity assumption that all normals are regular. It automatically entails the inclusion $\subset$ for the two sets in the second equation. The task therefore is to show that the opposite inclusion $\supset$ holds as well. To accomplish this we fix any vector $\bar{w} \notin T_C(\bar{x})$ and aim at demonstrating the existence of a vector $\bar{v} \in N_C(\bar{x})$ such that $\bar{v}{\cdot}\bar{w} > 0$.

Replacing C by its intersection with some closed ball around $\bar{x}$ if necessary (which involves no loss of generality, since the generation of normal vectors depends only on a neighborhood of $\bar{x}$), we can suppose that C is compact. Let B stand for some closed ball around $\bar{w}$ that doesn't meet $T_C(\bar{x})$ (as exists because $T_C(\bar{x})$ is closed). The definition of $T_C(\bar{x})$, in conjunction with having $T_C(\bar{x}) \cap D = \emptyset$, implies the existence of an $\varepsilon > 0$ such that the compact, convex set $S = \big\{ \bar{x} + \tau w \,\big|\, w \in B,\ \tau \in [0, \varepsilon] \big\}$ meets C only at $\bar{x}$.

Consider any sequence $\varepsilon^\nu \searrow 0$ with $\varepsilon^\nu < \varepsilon$ along with the compact, convex sets $S^\nu = \{\bar{x} + \tau w \mid w \in B,\ \tau \in [\varepsilon^\nu, \varepsilon]\}$, which are disjoint from C.

The function $h(x,u) = \frac{1}{2}|x-u|^2$ attains its minimum over the compact set $C \times S^\nu$ at some point (x^ν, u^ν). In particular, x minimizes $h(x, u^\nu)$ with respect to $x \in C$, so by Theorem 9 the vector $-\nabla_x h(x^\nu, u^\nu) = u^\nu - x^\nu$ belongs to $N_C(x^\nu)$. Likewise, the vector $-\nabla h_u(x^\nu, u^\nu) = x^\nu - u^\nu$ belongs to $N_{S^\nu}(u^\nu)$. Necessarily $x^\nu \neq u^\nu$ because $C \cap S^\nu = \emptyset$, but $x^\nu \to \bar{x}$ and $u^\nu \to \bar{x}$ because the sets S^ν increase to S (the closure of their union), and $C \cap S = \{\bar{x}\}$.

Let $v^\nu = (u^\nu - x^\nu)/|u^\nu - x^\nu|$, so $|v^\nu| = 1$, $v^\nu \in N_C(x^\nu)$, $-v^\nu \in N_{S^\nu}(u^\nu)$. The sequence of vectors v^ν being bounded, it has a cluster point $\bar{v}$, $|\bar{v}| = 1$; without loss of generality (by passing to a subsequence of necessary) we can suppose for simplicity that $v^\nu \to \bar{v}$. Along with the fact that $v^\nu \in N_C(x^\nu)$ and $x^\nu \to \bar{x}$, this implies that $\bar{v} \in N_C(\bar{x})$. Because $-v^\nu \in N_{S^\nu}(x^\nu)$ and S^ν is convex, we also have $-v^\nu{\cdot}[u - u^\nu] \leq 0$ for all $u \in S^\nu$. Since S^ν increases to S while $u^\nu \to \bar{x}$, we obtain in the limit that $-\bar{v}{\cdot}[u - \bar{x}] \leq 0$ for all $u \in S$. Recalling the construction of S, we note that among the vectors $u \in S$ are all vectors of the form $\bar{x} + \varepsilon w$ with $w \in B$. Further, B is the closed ball of a certain radius $\delta > 0$ around $\bar{w}$, so its elements w have the form $\bar{w} + \delta z$ with $|z| \leq 1$. Plugging these expressions into the limiting inequality that was obtained, we get $-\bar{v}{\cdot}\varepsilon[\bar{w} + \delta z] \leq 0$ for all z with $|z| \leq 1$. In particular we can take $z = -\bar{v}$ (since $|\bar{v}| = 1$) and see that $-\bar{v}{\cdot}\bar{w} + \delta|\bar{v}|^2 \leq 0$. This reveals that $\bar{v}{\cdot}\bar{w} \geq \delta$, so we have reached the desired conclusion that $\bar{v}{\cdot}\bar{w} > 0$.

Polar cones: When two cones in $\mathbb{R}^n$ are in this symmetric geometric relationship, they are said to be *polar* to each other. Then, incidentally, they both must be convex. Polarity of cones generalizes orthogonality of linear subspaces.

Determining normal vectors from constraint representations: To apply the optimality condition in Theorem 9 to the feasible set C in an optimization problem having not just an abstract constraint but equations and inequalities on functions $f_1(x), \ldots, f_m(x)$, it's essential to determine in detail how $N_C(\bar{x})$ can be expressed in such cases. We could try to work immediately with the form of constraint structure exhibited in a problem $(\mathcal{P})$ in conventional format, since that's our ultimate target, but it will be easier actually to work first in the more general setting in vector notation, where the set C consists of all x satisfying $x \in X$ and $F(x) \in D$. The main case to keep in mind is the one where D is a box $J_1 \times \cdots \times J_m$, so that the condition $F(x) \in D$ means $f_i(x) \in J_i$ for $i = 1, \ldots, m$.

THEOREM 10 (normals to sets with constraint structure). *In* $\mathbb{R}^n$ *let*

$$C = \{x \in X \mid F(x) \in D\}$$

for closed sets $X \subset \mathbb{R}^n$, $D \subset \mathbb{R}^m$, *and a* $\mathcal{C}^1$ *mapping* $F : \mathbb{R}^n \to \mathbb{R}^m$, *written componentwise as* $F(x) = \big(f_1(x), \ldots, f_m(x)\big)$. *Let* $\bar{x}$ *be a point of* C *such that* X *is regular at* $\bar{x}$ *and* D *is regular at* $F(\bar{x})$ *(as holds for instance when* X *and* D *are convex, and in particular when they are boxes). Suppose the following assumption (called a constraint qualification) is satisfied at* $\bar{x}$:

$$\begin{cases} \text{the only vectors } y = (y_1, \ldots, y_m) \in N_D\big(F(\bar{x})\big) \text{ and } z \in N_X(\bar{x}) \text{ for which} \\ y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}) + z = 0 \text{ are the vectors } y = (0, \ldots, 0),\ z = 0. \end{cases}$$

Then C *is regular at* $\bar{x}$, *and the normal cone* $N_C(\bar{x})$ *consists of all vectors* v *of the form*

$$y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}) + z \quad \text{with} \quad y = (y_1, \ldots, y_m) \in N_D\big(F(\bar{x})\big),\ z \in N_X(\bar{x}).$$

Note: When $X = \mathbb{R}^n$ in Theorem 10, or more generally whenever $\bar{x}$ lies in the interior of X, the normal cone $N_X(\bar{x})$ consists of just 0, so the z terms here drop out.

Proof. The justification of this theorem is furnished in its entirety, since this level of result can't be found yet in any textbook. The argument, although lengthy, is elementary in that it only uses standard facts about sequences and continuity in combination with the concepts introduced so far in this section. Of course C is a closed set (because D is closed and F is in particular continuous), so the picture is right for speaking about tangent and normal vectors to C.

It will help notation to write the gradient sums in the theorem in the Jacobian form $\nabla F(\bar{x})^* y$. To understand this, recall that $\nabla F(\bar{x})$ is by definition the $m \times n$ matrix having the gradient vectors $\nabla f_i(\bar{x})$ as its rows; the transpose $\nabla F(\bar{x})^*$ therefore has them as its columns, so that

$$\nabla F(\bar{x})^* y \;=\; y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}).$$

First consider a vector $v = \nabla F(\bar{x})^* y + z$ with $y \in N_D\big(F(\bar{x})\big)$ and $z \in N_D(\bar{x})$. We'll verify that v is a *regular* normal to C at $\bar{x}$. This means showing for an arbitrary vector $w \in T_C(\bar{x})$ that $v \cdot w \leq 0$. Such a vector w is by definition the limit of a sequence of vectors $w^\nu = (x^\nu - \bar{x})/\tau^\nu$ with $x^\nu \in C$ and $\tau^\nu \searrow 0$. In that setting we have $F(x^\nu) \in D$ and, by the continuity of F, also $F(x^\nu) \to F(\bar{x})$. Furthermore, the differentiability of F at $\bar{x}$ entails having $F(x^\nu) = F(\bar{x}) + \nabla F(\bar{x})(x^\nu - \bar{x}) + o(x^\nu - \bar{x})$ and consequently

$$\frac{F(x^\nu) - F(\bar{x})}{\tau^\nu} \;=\; \nabla F(\bar{x}) w^\nu + \frac{o\big(|x^\nu - \bar{x}|\big)}{|x^\nu - \bar{x}|} |w^\nu| \;\to\; \nabla F(\bar{x}) w.$$

Thus, $\nabla F(\bar{x})w \in T_D\big(F(\bar{x})\big)$. From the assumption that $y \in N_D\big(F(\bar{x})\big)$ and D is regular at $F(\bar{x})$, we know that y is a regular normal to D at $F(\bar{x})$, hence $y{\cdot}\nabla F(\bar{x})w \le 0$. But this inner product is the same as $\nabla F(\bar{x})^* y{\cdot}w \le 0$. At the same time we have $z{\cdot}w \ge 0$ through the fact that $z \in N_X(\bar{x})$ and X is regular at X; any regular normal to X at $\bar{x}$ is in particular a regular normal to C at $\bar{x}$, because $\bar{x} \in C \subset X$. From having both $\nabla F(\bar{x})^* y{\cdot}w \le 0$ and $z{\cdot}w \ge 0$ we conclude that $v{\cdot}w \ge 0$.

We must show that conversely, under the constraint qualification and regularity assumptions, every normal vector v to C at $\bar{x}$ in the general sense can be represented as $\nabla F(\bar{x})^* y + v$ for some $y \in N_D\big(F(\bar{x})\big)$ and $z \in N_X(\bar{x})$, from which it will follow that v is regular, as just seen, so that C is regular at $\bar{x}$. But the argument will pass through an intermediate stage involving representations "approximately" like this. Specifically, we consider now only a *regular* normal vector v to C at $\bar{x}$ together with any $\varepsilon > 0$, and we demonstrate the existence of

$$\left.\begin{array}{r} x \in \mathbb{R}^n \text{ with } |x - \bar{x}| < \varepsilon \\ u \in D \text{ with } \big|u - F(x)\big| < \varepsilon \\ w \in \mathbb{R}^n \text{ with } |w| < \varepsilon \\ y \in N_D(u),\ z \in N_X(x) \end{array}\right\} \text{ such that } v = \nabla F(x)^* y + z + w.$$

The proof of this won't yet use the constraint qualification or the regularity assumptions. We take any sequence of values $\tau^\nu \searrow 0$ and define the functions φ^ν on $\mathbb{R}^n \times \mathbb{R}^m$ by

$$\varphi^\nu(x,u) = \frac{1}{2\tau^\nu}\big|x - (\bar{x} + \tau^\nu v)\big|^2 + \frac{1}{2\tau^\nu}\big|F(x) - u\big|^2 \ \ge 0.$$

These functions are continuously differentiable in x and u. Our construction will be based on analyzing the problem of minimizing φ^ν over the closed set $X \times D$, and we have to know, before proceeding, that for each ν an optimal solution to this problem exists. Temporarily fixing ν and any value $\alpha \ge 0$, we'll establish that the level set $S = \big\{(x,u) \in X \times D \,\big|\, \varphi^\nu(x,u) \le \alpha\big\}$ is bounded; this will show that the problem of minimizing φ^ν over $X \times D$ is well posed, from which the existence of an optimal solution is assured by Theorem 1. Obviously any point $(x,u) \in S$ has both $|x - (\bar{x} + \tau^\nu v)|^2 \le 2\tau^\nu\alpha$ and $|u - F(x)|^2 \le 2\tau^\nu\alpha$. Then $|x| \le \lambda := |\bar{x}| + \tau^\nu|v| + \sqrt{2\tau^\nu\alpha}$. Over the closed ball $\big\{x \,\big|\, |x| \le \lambda\big\}$ there is a maximum to the possible values of $\big|F(x)\big|$ (an expression that is continuous in x); say $\big|F(x)\big| \le \sigma$ for all such x. The inequality $|u - F(x)|^2 \le 2\tau^\nu\rho$ then yields $|u| \le \mu = \sigma + \sqrt{2\tau^\nu\alpha}$. Since every element $(x,u) \in S$ has $|x| \le \lambda$ and $|u| \le \mu$, the level set S is bounded as claimed.

We now have license to denote by (x^ν, u^ν) for each ν an optimal solution (not necessarily unique) to the problem of minimizing the function φ^ν over $X \times D$; we denote the optimal value in this problem by α^ν. Obviously

$$0 \le \alpha^\nu = \varphi^\nu(x^\nu, u^\nu) \le \varphi^\nu\big(\bar{x}, F(\bar{x})\big) = \frac{\tau^\nu}{2}|v|^2 \to 0.$$

The inequalities deduced in our investigation of level sets of φ^ν tell us at the same time that neither $|x^\nu - (\bar{x} + \tau^\nu v)|^2$ nor $|u^\nu - F(x^\nu)|^2$ can exceed $2\tau^\nu\alpha^\nu$. Since $\alpha^\nu \le (\tau^\nu/2)|v|^2$, as just seen, we must have $\big|x^\nu - (\bar{x} + \tau^\nu v)\big| \le \tau^\nu|v|$ and $\big|u^\nu - F(x^\nu)\big| \le \tau^\nu|v|$. Therefore, $x^\nu \to \bar{x}$ and $\big|u^\nu - F(x^\nu)\big| \to 0$. In addition the sequence of vectors $w^\nu = (x^\nu - \bar{x})/\tau^\nu$ is bounded because the inequality $\big|x^\nu - (\bar{x} + \tau^\nu v)\big| \le \tau^\nu|v|$, when divided by τ^ν, gives $|w^\nu - v| \le |v|$. This sequence therefore has a cluster point w. Any such cluster point w belongs by definition to the tangent cone $T_C(\bar{x})$, so it satisfies $w{\cdot}v \le 0$ because v is a regular normal at $\bar{x}$, but at the same time it satisfies $|w - v| \le |v|$. In squaring the latter we see that $|w|^2 - 2v{\cdot}w + |v|^2 \le |v|^2$, which implies $|w|^2 \le 2v{\cdot}w \le 0$. Hence actually $w = 0$. Thus, the only possible cluster point of the sequence of vectors w^ν is 0, and we conclude that $w^\nu \to 0$. Eventually then, once ν is large enough, we'll have

$$|x^\nu - \bar{x}| < \varepsilon, \qquad |u^\nu - F(x^\nu)| < \varepsilon, \qquad |w^\nu| < \varepsilon.$$

Next we use the fact that (x^ν, u^ν) minimizes φ^ν over $X \times D$. In particular the minimum of $\varphi^\nu(x^\nu, u)$ over $u \in D$ is attained at u^ν, whereas the minimum of $\varphi^\nu(x, u^\nu)$ over $x \in X$ is attained at x^ν. From part (a) of Theorem 9, therefore, we have

$$-\nabla_u\varphi^\nu(x^\nu, u^\nu) \in N_D(u^\nu), \qquad -\nabla_x\varphi^\nu(x^\nu, u^\nu) \in N_X(x^\nu).$$

with $-\nabla_u\varphi^\nu(x^\nu, u^\nu)$ actually a regular normal vector to D at u^ν. Let $y^\nu = -\nabla_u\varphi^\nu(x^\nu, u^\nu)$ and $z^\nu = -\nabla_x\varphi^\nu(x^\nu, u^\nu)$. Then $y^\nu \in N_D(u^\nu)$, $z^\nu \in N_X(x^\nu)$ and, as seen through differentiation of φ^ν in u, we have $y^\nu = [F(x^\nu) - u^\nu]/\tau^\nu$. Differentiation of φ^ν in x reveals then that $z^\nu = -w^\nu + v - \nabla F(x^\nu)^* y^\nu$. Thus, $v = \nabla F(x^\nu)^* y^\nu + z^\nu + w^\nu$. It follows that when ν is taken sufficiently large the elements $x = x^\nu$, $u = u^\nu$, $y = y^\nu$ and $w = w^\nu$ furnish the kind of approximate representation of v that was required.

Now we are ready for the final stage of argument. We consider a general vector $v \in N_C(\bar{x})$ and aim at proving the existence of $y \in N_D(\bar{x})$ and $z \in N_D(\bar{x})$ such that $v = \nabla F(\bar{x})^* y + z$. Fix any sequence of values $\varepsilon^\nu \searrow 0$. From the definition of normal vectors in general sense, we know there exist sequences $\bar{x}^\nu \to \bar{x}$ in C and $v^\nu \to v$ with v^ν a regular normal to C at $\bar{x}^\nu$. Then, on the basis of the intermediate fact just established, there exist for each ν

$$\left.\begin{array}{r} x^\nu \in I\!R^n \text{ with } |x^\nu - \bar{x}^\nu| < \varepsilon^\nu \\ u^\nu \in D \text{ with } \big|u^\nu - F(\bar{x}^\nu)\big| < \varepsilon^\nu \\ w^\nu \in I\!R^n \text{ with } |w^\nu| < \varepsilon^\nu \\ y^\nu \in N_D(u^\nu),\ z^\nu \in N_X(x^\nu) \end{array}\right\} \text{ such that } v^\nu = \nabla F(x^\nu)^* y^\nu + z^\nu + w^\nu.$$

There are two cases to distinguish: either the sequence of vectors y^ν has a cluster point y, or it has no bounded subsequences at all, meaning that $|y^\nu| \to \infty$. In the case of a cluster point y, we have in the limit that $y \in N_D\big(F(\bar{x})\big)$, since $F(\bar{x}^\nu) \to F(\bar{x})$ and $\nabla F(\bar{x}^\nu) \to \nabla F(\bar{x})$ by the continuity of F and its first partial derivatives. The corresponding subsequence of the z^ν's is then bounded and must have a cluster point too, say z. We have $z \in N_D(\bar{x})$ and consequently from taking limits in the equation $v^\nu = \nabla F(x^\nu)^* y^\nu + z^\nu + w^\nu$ that $v = \nabla F(\bar{x})^* y + z$, as needed. In the contrary case, where $|y^\nu| \to \infty$, the vectors $\bar{y}^\nu = y^\nu/|y^\nu|$ and $\bar{z}^\nu = z^\nu/|y^\nu|$, which like y^ν and z^ν belong to $N_D\big(F(x^\nu)\big)$ and $N_X(x^\nu)$, have unit length and the sequence of pairs $(\bar{y}^\nu, \bar{z}^\nu)$ therefore has a cluster point $(\bar{y}, \bar{z})$ with $|\bar{y}| = 1$ and $|\bar{z}| = 1$. Again we get $\bar{y} \in N_D\big(F(\bar{x})\big)$ and $\bar{z} \in N_D(\bar{x})$. In dividing the equation $v^\nu = \nabla F(x^\nu)^* y^\nu + z^\nu + w^\nu$ by $|y^\nu|$ and taking the limit as $\nu \to \infty$, we see that $\nabla F(\bar{x})^* \bar{y} + \bar{z} = 0$. But $\bar{y} \neq 0$, so this is impossible under the constraint qualification in the theorem. Only the first case is viable. This finishes the proof. □

Tangents to sets with constraint structure: Under the assumptions of Theorem 10, the tangent cone to C at $\bar{x}$ has the formula:

$$T_C(\bar{x}) = \big\{ w \in T_X(\bar{x}) \,\big|\, \nabla F(\bar{x}) w \in T_D\big(F(\bar{x})\big) \big\}.$$

This follows from the regularity of C at $\bar{x}$, which we earlier saw implies that $T_C(\bar{x})$ consists of the vectors w satisfying $v \cdot w \leq 0$ for all $v \in N_C(\bar{x})$. The formula for such vectors v in Theorem 10, together with the expressions of $T_X(\bar{x})$ and $T_D\big(F(\bar{x})\big)$ in terms of $N_X(\bar{x})$ and $N_D\big(F(\bar{x})\big)$, based again on regularity, gives the result.

Lagrange multipliers: The coefficients y_i in this representation of a normal vector v are called *Lagrange multipliers* associated with the constraint functions f_i at $\bar{x}$.

Sign restrictions on Lagrange multipliers: In the central case where the set D in Theorem 10 is a box $J_1 \times \cdots \times J_m$, corresponding to constraints $f_i(x) \in J_i$ in the specification of C, one has $N_D\big(F(\bar{x})\big) = N_{J_1}\big(f_1(\bar{x})\big) \times \cdots \times N_{J_m}\big(f_m(\bar{x})\big)$. Then the condition $y \in N_D\big(F(\bar{x})\big)$ imposed in Theorem 10 means that

$$\begin{cases} y_i \leq 0 & \text{if } f_i(\bar{x}) \text{ is the left endpoint (only) of } J_i, \\ y_i \geq 0 & \text{if } f_i(\bar{x}) \text{ is the right endpoint (only) of } J_i, \\ y_i = 0 & \text{if } f_i(\bar{x}) \text{ lies in the interior of } J_i, \\ y_i \text{ free} & \text{if } J_i \text{ is a one-point interval, consisting just of } f_i(\bar{x}). \end{cases}$$

Why the constraint qualification is needed in general: Without the qualification in Theorem 10 being satisfied, the formula in that theorem can fall short of expressing all the vectors $v \in N_C(\bar{x})$. This can already be seen from examples in $\mathbb{R}^2$. Let C be the set of all $x = (x_1, x_2)$ satisfying $f_1(x) \le 0$ and $f_2(x) \le 0$ for the functions

$$f_1(x) = (x_1 - 1)^2 + (x_2 - 1)^2 - 2, \qquad f_2(x) = (x_1 + 1)^2 + (x_2 + 1)^2 - 2,$$

which corresponds to $D = \mathbb{R}^2_- = (-\infty, 0] \times (-\infty, 0]$ and $X = \mathbb{R}^2$. In this case C is the intersection of two disks, both of radius $\sqrt{2}$ and centered at $(1, 1)$ and $(-1, -1)$ respectively. These touch only at the origin, so that C consists of just $\bar{x} = (0, 0)$, with $N_C(\bar{x}) = \mathbb{R}^2$. We have $N_D\big(F(\bar{x})\big) = N_D(0, 0) = \mathbb{R}^2_+$ and $N_X(\bar{x}) = \{(0, 0)\}$, so the only vectors v covered by the formula in Theorem 10 are those expressible as $y_1(-1, -1) + y_2(1, 1) = (y_2 - y_1)(1, 1)$ for some choice of $y_1 \ge 0$ and $y_2 \ge 0$. Such vectors merely constitute a line in $\mathbb{R}^2$, not all of $\mathbb{R}^2$.

Example of normal and tangent spaces to smooth manifolds: Suppose C is given by a system of equations $f_i(x) = 0$ for $i = 1, \ldots, m$, and $\bar{x}$ is a point of C where the gradient vectors $\nabla f_i(\bar{x})$ are linearly independent, i.e., the Jacobian matrix $\nabla F(\bar{x})$ for the mapping F with component functions f_i has rank m. This is the classical case in differential geometry in which C is a *smooth manifold* around $\bar{x}$. In the framework of Theorem 10 we have $X = \mathbb{R}^n$, whereas D is the singleton set consisting of just the origin of $\mathbb{R}^m$, which is the value of $F(\bar{x})$. Hence $N_X(\bar{x})$ consists only of the origin of $\mathbb{R}^n$ while $N_D\big(F(\bar{x})\big) = \mathbb{R}^m$. The constraint qualification of Theorem 10 reduces to the condition that the only combination of coefficients $y_i \in \mathbb{R}$ with $y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}) = 0$ is the 0 coefficients. This is precisely the linear independence of the constraint gradients, which we have supposed. It follows that

$$N_C(\bar{x}) = \big\{ v = y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}) \,\big|\, y_i \in \mathbb{R} \big\},$$

which is the linear subspace of $\mathbb{R}^n$ spanned by all the vectors $\nabla f_i(\bar{x})$, whereas

$$T_C(\bar{x}) = \big\{ w \,\big|\, \nabla f_1(\bar{x}) \cdot w = 0, \ldots, \nabla f_m(\bar{x}) \cdot w = 0 \big\},$$

which is the linear subspace of $\mathbb{R}^n$ orthogonal to the vectors $\nabla f_i(\bar{x})$. These are the classical normal and tangent spaces associated with the smooth manifold C at $\bar{x}$.

Normals to sets with linear constraint structure only: The constraint qualification of Theorem 10 is superfluous in the case where X and D are boxes and the mapping F is affine (its component functions f_i are affine). The normal cone formula holds then at every point $\bar{x} \in C$ without need for any extra assumption. So too does the corresponding tangent cone formula indicated above.

Significance: This fact is important in linear and quadratic programing. Therefore its linear-algebraic proof will be supplied in full for the theoretically minded.

Proof: A shift of notation will greatly simplify matters. The condition $F(x) \in D$ refers to constraints $f_i(x) \in J_i$ for certain closed intervals J_i, each f_i being affine, but likewise the condition $x \in X$ refers to similar constraints $g_j(x) \in I_j$ for certain closed intervals I_j, where $g_j(x) = e_j{\cdot}x$ for the vector e_j having jth component 1 but all other components 0. We can regard the g_j constraints as a supplementary batch of f_i constraints and think of C as specified in that one manner. The normal cone formula comes out the same either way. Thus, without loss of generality we can focus on the case where no g_j constraints enter in at all, i.e., $X = I\!R^n$.

Having done this, we can proceed also to drop any constraint $f_i(x) \in J_i$ that's inactive because $f_i(\bar{x})$ lies in the interior of J_i; such a constraint has no effect on C around $\bar{x}$, and the multiplier y_i it contributes has to be 0. Among the active constraints remaining, some will correspond to equations—where J_i is a one-point interval containing just $f_i(\bar{x})$—while others will correspond to inequalities where $f_i(\bar{x})$ is only the right endpoint or only the left endpoint of J_i. By a switch of signs we can convert the left endpoint cases to right endpoint cases. Then, with only right endpoints active, we can suppress the left endpoints as irrelevant. In other words, we can take the intervals J_i corresponding to inequalities all to be of the form $(-\infty, c_i]$, where $f_i(\bar{x}) = c_i$. Each f_i is given by an affine expression $f_i(x) = a_i{\cdot}x - b_i$, and we see that (through a change of coordinates if necessary with x replaced by $x' = x - \bar{x}$) we can arrange that $\bar{x} = 0$, $c_i = 0$.

In summary, we can revert to the case of C being specified by

$$a_i{\cdot}x \in \begin{cases} (-\infty, 0] & \text{for } i \in [1, s], \\ [0, 0] & \text{for } i \in [s+1, m], \end{cases}$$

with all these constraints active at $\bar{x} = 0$. In this situation the tangent cone is

$$T_C(\bar{x}) = \big\{ w \,\big|\, a_i{\cdot}w \le 0 \text{ for } i \in [1, s],\ a_i{\cdot}w = 0 \text{ for } i \in [s+1, m] \big\},$$

whereas the set claimed to be $N_C(\bar{x})$ is

$$N = \big\{ y_1 a_1 + \cdot + y_m a_m \,\big|\, y_i \ge 0 \text{ for } i \in [1, s],\ y_i \text{ free for } i \in [s+1, m] \big\}.$$

It's evident that each $v \in N$ has $v{\cdot}w \leq 0$ for all $w \in T_C(\bar{x})$, so v is a regular normal at $\bar{x}$. Hence $N \subset N_C(\bar{x})$. The challenge is to show $N \supset N_C(\bar{x})$.

Let Y be the set of vectors y with $y_i \geq 0$ for $i \in [1,s]$ but y_i free for $i \in [s+1,m]$, and let $Y_0 = \big\{ y \in Y \,\big|\, \sum_{i=1}^m y_i a_i = 0 \big\}$. Let I be the set of indices $i \in [1,s]$ such that there exists $y \in Y_0$ with $y_i > 0$. There must actually be a vector $\tilde{y} \in Y_0$ with $\tilde{y}_i > 0$ for all $i \in I$ simultaneously. (Taking separate vectors that do the trick for each $i \in I$ individually, add them to get $\tilde{y}$.) For every $w \in T_C(\bar{x})$ the vector $\tilde{v} := \tilde{y}_1 a_1 + \cdots + \tilde{y}_m a_m$ satisfies $0 \geq \tilde{v}{\cdot}w = \sum_{i=1}^m \tilde{y}_i a_i{\cdot}w$. From the description of $T_C(\bar{x})$ we see this implies $a_i{\cdot}w = 0$ for all $i \in I$. Thus the inequalities describing $T_C(\bar{x})$ must, for $i \in I$, hold as equations, and this is true then too for the system specifying C (since in our set-up these are the same). Reordering indices, we can suppose $I = [r+1, \ldots, s]$ for a certain r.

Choose a maximal linearly independent subset of $\big\{ a_i \,\big|\, i \in [r+1,m] \big\}$, denoting the corresponding set of indices by I_0; then $\big\{ a_i \,\big|\, i \in I_0 \big\}$ is a basis for the subspace of $\mathbb{R}^n$ spanned by $\big\{ a_i \,\big|\, i \in [r+1,m] \big\}$. In particular, every vector in $\big\{ a_i \,\big|\, i \in [r+1,m] \big\}$ can be expressed as a linear combination of the vectors in $\big\{ a_i \,\big|\, i \in I_0 \big\}$. Then C is equally well specified by the alternative system

$$a_i{\cdot}x \in \begin{cases} (-\infty, 0] & \text{for } i \in [1,r], \\ [0,0] & \text{for } i \in I_0. \end{cases}$$

The constraint qualification of Theorem 10 for this system demands that the only coefficients giving $\sum_{i=1}^r y_i a_i + \sum_{i \in I_0} y_i a_i = 0$ with $y_i \geq 0$ for $i \in [1,r]$ be the 0 coefficients. Is this fulfilled? Consider any such coefficients y_i for $i \in [1,r] \cup I_0$ and augment them by 0 coefficients for the remaining indices i so as to get a vector $y = (y_1, \ldots, y_m)$ with $\sum_{i=1}^n y_i a_i = 0$. Taking $y' = y + \lambda \tilde{y}$ for $\lambda > 0$ sufficiently large, we get $\sum_{i=1}^m y_i' a_i = 0$ with $y_i' \geq 0$ for all $i \in [1,s]$. The maximality in the selection of I and $\tilde{y}$ ensures that for $i \in [1,r]$ not only $\tilde{y}_i = 0$ but $y_i' = 0$. Hence $y_i = 0$ for all $i \in [1,r]$, and we are left with having $\sum_{i \in I_0} y_i a_i = 0$. But the vectors a_i for $i \in I_0$ are linearly independent, so this implies $y_i = 0$ for all $i \in I_0$ as well. The condition does therefore hold, as needed, so the formula of Theorem 10 is applicable to the alternative representation of C at $\bar{x}$.

By this formula the vectors $v \in N_C(\bar{x})$ have the form $v = \sum_{i=1}^r y_i a_i + \sum_{i \in I_0} y_i a_i$ for coefficients that are nonnegative for $i \in [1,r]$ but free for $i \in I_0$. Again we can augment these by 0 coefficients in the remaining indices to get $v = \sum_{i=1}^m y_i a_i$ with $y_i \geq 0$ for $i \in [1,r]$. Then with $y' = y + \lambda \tilde{y}$ for large enough $\lambda > 0$ we have $v = \sum_{i=1}^m y_i' a_i$ with $y_i' \geq 0$ for $i \in [1,s]$. This confirms that $v \in N$. Hence $N_C(\bar{x}) \subset N$, and we are done.

Normals to polyhedral sets: Sets with linear constraint structure only are of course polyhedral sets, by definition. Earlier we had a formula for the tangent cones to such a set C in the case of a constraint representation of the type

$$x \in C \quad \Longleftrightarrow \quad \begin{cases} a_i \cdot x \leq b_i & \text{for } i \in [1, s], \\ a_i \cdot x = b_i & \text{for } i \in [s+1, m], \\ x \in X, & \text{with } X \text{ a box,} \end{cases}$$

which entails no loss of generality. Now we can obtain a corresponding formula for the normal cones to such a set C. It's only necessary to call upon Theorem 10, and in doing so we are permitted to ignore the constraint qualification in that theorem, because of the argument just given. Then for any $\bar{x} \in C$ we have

$$v \in N_C(\bar{x}) \quad \Longleftrightarrow \quad \begin{cases} v \;=\; y_1 a_1 + \cdots + y_m a_m + z \ \text{ for some choice of} \\ z \in N_X(\bar{x}) \ \text{ and } \begin{cases} y_i \geq 0 & \text{for } i \in [1, s] \text{ with } a_i \cdot \bar{x} = b_i, \\ y_i = 0 & \text{for } i \in [1, s] \text{ with } a_i \cdot \bar{x} < b_i, \\ y_i \ \text{free} & \text{for } i \in [s+1, m]. \end{cases} \end{cases}$$

Here we are taking $F(x) = \big(f_1(x), \ldots, f_m(x)\big)$ with $f_i(x) = a_i \cdot x - b_i$ and $D = J_1 \times \cdots \times J_m$ with $J_i = (-\infty, 0]$ for $i \in [1, s]$ but $J_i = [0, 0]$ for $i \in [s+1, m]$.

Generalization: Just as easily we could replace the particular inequalities and equations in this constraint representation for C by conditions of the form $a_i \cdot x \in J_i$ for any closed intervals J_i. The effect would merely be to alter the pattern of sign restrictions on the y_i's.

Minimization with linear constraints: When a function f_0 of class $\mathcal{C}^1$ is minimized over a polyhedral set C, and C is furnished with this kind of representation, it follows from Theorem 9 and the formula just developed that there must be an expression

$$-\nabla f_0(\bar{x}) \;=\; \bar{y}_1 a_1 + \cdots + \bar{y}_m a_m + \bar{z}$$

in terms of coefficients $\bar{y}_i$ and a vector $\bar{z}$ satisfying the conditions indicated. Much the same is true for nonlinear constraints, except that the constraint qualification in Theorem 10 can't usually be avoided as an assumption needed to make the normal cone formula work.

Lagrange multipliers in the description of optimality: In combining Theorem 9 with Theorem 10, it's possible to deduce optimality conditions of great generality in which normal vector expressions involving Lagrange multipliers play a leading role. Here's the pattern. Suppose C is a set having a constraint representation of the general kind in Theorem 10, and consider a problem in which a $\mathcal{C}^1$ function f_0 is minimized over C. Let $\bar{x}$ be locally optimal. By Theorem 9 we must have $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$. By Theorem 10, under the additional assumption that the constraint qualification in that theorem is fulfilled at $\bar{x}$, there must exist a representation

$$-\nabla f_0(\bar{x}) = \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) + \bar{z}$$

in which the coefficient vector $(\bar{y}_1, \ldots, \bar{y}_m)$ belongs to the cone $N_D\big(F(\bar{x})\big)$ and the vector $\bar{z}$ belongs to the cone $N_X(\bar{x})$. This equation can be written instead as

$$\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) + \bar{z} = 0,$$

but even better as

$$-\big[\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x})\big] = \bar{z}.$$

Therefore, the local optimality of $\bar{x}$ implies, under the constraint qualification in Theorem 10, the existence of Lagrange multipliers $\bar{y}_i$ such that

$$-\big[\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x})\big] \in N_X(\bar{x}) \quad \text{with } (\bar{y}_1, \ldots, \bar{y}_m) \in N_D\big(F(\bar{x})\big).$$

When D is a box, the final condition on $(\bar{y}_1, \ldots, \bar{y}_m)$ corresponds to simple sign restrictions on the $\bar{y}_i$'s, as already explained. Our goal now is to elaborate this result in the conventional setting.

Lagrange multipliers for problems in conventional format: These facts will be applied now to the feasible set C associated with a problem in conventional format:

$$(\mathcal{P}) \qquad \text{minimize } f_0(x) \text{ over all } x \in X \text{ such that } f_i(x) \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases}$$

It will be assumed throughout this discussion that the functions $f_0, f_1, \ldots, f_m$ are of class $\mathcal{C}^1$, and the set X is nonempty, closed and regular at all of its points. In particular X could be $I\!R^n$ or any closed, convex set such as a box. An inequality constraint $f_i(x) \leq 0$ will be called *active* at $\bar{x}$ if $f_i(\bar{x}) = 0$ and *inactive* if $f_i(\bar{x}) < 0$.

How this fits in: Here $C = \{x \in X \,|\, F(x) \in D\}$ for the vector mapping F with component functions f_i and the set $D = (-\infty, 0]^s \times [0, 0]^{m-s}$, which is a box. We work out the details of our theory of Lagrange multipliers for this case. First we need the corresponding specialization of the constraint qualification of Theorem 10, which will turn out to be the following.

Basic constraint qualification: For a feasible solution $\bar{x}$ to such a problem $(\mathcal{P})$, we'll say that the *basic constraint qualification* is fulfilled at $\bar{x}$ if

$$\begin{cases} \text{only the vector } y = (0, \ldots, 0) \text{ satisfies} \\ -\big[y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x})\big] \in N_X(\bar{x}) \\ \text{with } \begin{cases} y_i \geq 0 & \text{for } i \in [1, s] \text{ active at } \bar{x}, \\ y_i = 0 & \text{for } i \in [1, s] \text{ inactive at } \bar{x}, \\ y_i \text{ free} & \text{for } i \in [s+1, m]. \end{cases} \end{cases}$$

THEOREM 11 (first-order optimality in conventional format). *Let $\bar{x}$ be a feasible solution to problem $(\mathcal{P})$ (with every f_i of class $\mathcal{C}^1$ and X closed and regular).*

(a) (necessary). *If $\bar{x}$ is locally optimal and the basic constraint qualification is fulfilled at $\bar{x}$, there must be a Lagrange multiplier vector $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_m)$ such that*

$$-\big[\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x})\big] \in N_X(\bar{x})$$
$$\textit{with } \begin{cases} \bar{y}_i \geq 0 & \textit{for } i \in [1, s] \textit{ active at } \bar{x}, \\ \bar{y}_i = 0 & \textit{for } i \in [1, s] \textit{ inactive at } \bar{x}, \\ \bar{y}_i \textit{ free} & \textit{for } i \in [s+1, m]. \end{cases}$$

(b) (sufficient). *If such a vector $\bar{y}$ exists, and f_0 and C are convex (as in the case of convex programming), then $\bar{x}$ is globally optimal.*

Note: When $\bar{x}$ lies in the interior of X the gradient condition here becomes:

$$\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) = 0.$$

Proof. As already mentioned, we have here the case of $C = \{x \in X \,|\, F(x) \in D\}$ with $F(x) = \big(f_1(x), \ldots, f_m(x)\big)$ and the box $D = J_1 \times \cdots \times J_m$ formed by taking $J_i = (-\infty, 0]$ for $i \in [1, s]$ but $J_i = [0, 0]$ for $i \in [s+1, m]$. Since D is a box, it's regular at all of its points. We saw earlier how the condition $y \in N_D\big(F(\bar{x})\big)$ reduces for a box D to sign restrictions on the components y_i of y. Here these restrictions come out as the ones in the statement of the basic constraint qualification as well as in the theorem. We only have to invoke Theorem 9 using the formula for $N_C(\bar{x})$ provided by Theorem 10. □

Kuhn-Tucker conditions: A pair of vectors $\bar{x} \in I\!R^n$ and $\bar{y} \in I\!R^m$ is said to satisfy the *Kuhn-Tucker conditions* (and be a *Kuhn-Tucker pair*) for $(\mathcal{P})$ if $\bar{x}$ is a feasible solution to $(\mathcal{P})$ and $\bar{y}$ is an associated Lagrange multiplier vector as in Theorem 11:

$$-\left[\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x})\right] \in N_X(\bar{x})$$
$$\text{with } \begin{cases} \bar{y}_i \geq 0 & \text{for } i \in [1,s] \text{ active at } \bar{x}, \\ \bar{y}_i = 0 & \text{for } i \in [1,s] \text{ inactive at } \bar{x}, \\ \bar{y}_i \text{ free} & \text{for } i \in [s+1,m]. \end{cases}$$

The Kuhn-Tucker conditions thus constitute the chief mode of expressing first-order optimality for optimization problems $(\mathcal{P})$ in conventional format. Under the basic constraint qualification, the local optimality of $\bar{x}$ implies the existence of $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a Kuhn-Tucker pair. On the other hand, this property is sufficient for the global optimality of $\bar{x}$ in the case of convex programming.

Alternative expression of the sign restrictions on Lagrange multipliers: The restrictions on the $\bar{y}_i$'s in the Kuhn-Tucker conditions can fruitfully be stated in another way. In combination with the requirement that $\bar{x}$ satisfy $f_i(\bar{x}) \leq 0$ for $i \in [1,s]$ and $f_i(\bar{x}) = 0$ for $i \in [s+1,m]$, they can jointly be summarized in terms of the vector $F(\bar{x}) = \big(f_1(\bar{x}), \ldots, f_m(\bar{x})\big)$ by saying that

$$F(\bar{x}) \in N_Y(\bar{y}), \text{ where } Y := I\!R^s_+ \times I\!R^{m-s}.$$

Reason: Here Y is a box: s copies of $[0,\infty)$ times $m-s$ copies of $(-\infty,\infty)$. For $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_m)$ and $\bar{u} = (\bar{u}_1, \ldots, \bar{u}_m)$ (with $\bar{u} = F(\bar{x})$ as a special case),

$$\left\{ \begin{array}{l} \bar{y} \in Y \\ \bar{u} \in N_Y(\bar{y}) \end{array} \right\} \iff \left\{ \begin{array}{ll} \bar{y}_i \geq 0,\ \bar{u}_i \leq 0,\ \bar{y}_i \bar{u}_i = 0 & \text{for } i \in [1,s] \\ \bar{y}_i \text{ free},\ \bar{u}_i = 0 & \text{for } i \in [s+1,m] \end{array} \right\}.$$

Lagrangian function: An elegant expression of the Kuhn-Tucker conditions as a whole can be achieved in terms of the *Lagrangian* for problem $(\mathcal{P})$, which is the function

$$L(x,y) := f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) \quad \text{for } x \in X \text{ and } y \in Y = I\!R^s_+ \times I\!R^{m-s}.$$

Then $\nabla_x L(x,y) = \nabla f_0(x) + y_1 \nabla f_1(x) + \cdots + y_m \nabla f_m(x)$ and $\nabla_y L(x,y) = F(x)$.

Lagrangian form of the Kuhn-Tucker conditions: A pair of vectors $\bar{x}$ and $\bar{y}$ satisfies the Kuhn-Tucker conditions for $(\mathcal{P})$ if and only if $\bar{x} \in X$, $\bar{y} \in Y$, and

$$-\nabla_x L(\bar{x}, \bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x}, \bar{y}) \in N_Y(\bar{y}),$$

where $Y = I\!R^s_+ \times I\!R^{m-s}$. This is immediate from the observations just made.

Problems with an abstract constraint only: If there are no constraints in function form, so $(\mathcal{P})$ consists just of minimizing $f_0(x)$ over all $x \in X$ (this set being identical with C), the optimality condition furnished by Theorem 11 reduces to the one in Theorem 9. At any locally optimal solution $\bar{x}$ we must have

$$-\nabla f_0(\bar{x}) \in N_X(\bar{x}).$$

Problems with equality constraints only: Consider the case of $(\mathcal{P})$ where no inequality constraints are present and $X = I\!R^n$:

$$\text{minimize } f_0(x) \text{ subject to } f_i(x) = 0 \text{ for } i = 1, \ldots, m.$$

The basic constraint qualification turns out to mean the linear independence of the vectors $\nabla f_1(\bar{x}), \ldots, \nabla f_m(\bar{x})$. This is the case where C is a smooth manifold around $\bar{x}$, the geometry of which was explained earlier. We obtain from Theorem 11 that if $\bar{x}$ is a locally optimal solution at which such linear independence holds, there must exist multipliers $\bar{y}_1, \ldots, \bar{y}_m$ such that

$$\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) = 0 \quad \text{(with } \bar{y}_i \text{ free of restriction).}$$

Problems with inequality constraints only: Consider instead the case of $(\mathcal{P})$ where $X = I\!R^n$ and $s = m$, the problem being to

$$\text{minimize } f_0(x) \text{ subject to } f_i(x) \leq 0 \text{ for } i = 1, \ldots, m.$$

The basic constraint qualification comes out then as the requirement that

$$\begin{cases} \text{only the vector } y = (0, \ldots, 0) \text{ satisfies the conditions} \\ y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x}) = 0 \text{ with } \begin{cases} y_i \geq 0 & \text{for } i \text{ active at } \bar{x}, \\ y_i = 0 & \text{for } i \text{ inactive at } \bar{x}. \end{cases} \end{cases}$$

When this holds at a locally optimal solution $\bar{x}$, we conclude from Theorem 11 that there must exist multipliers $\bar{y}_1, \ldots, \bar{y}_m$ such that

$$\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) = 0 \text{ with } \begin{cases} \bar{y}_i \geq 0 & \text{for } i \text{ active at } \bar{x}, \\ \bar{y}_i = 0 & \text{for } i \text{ inactive at } \bar{x}. \end{cases}$$

THEOREM 12 (refined constraint qualification for linear constraints). *For a problem $(\mathcal{P})$ in conventional format in which X is a box and some of the constraint functions f_i are affine, the following weaker condition can be substituted for the basic constraint qualification in Theorem 11:*

$$\begin{cases} \text{the only vectors } y = (y_1, \ldots, y_m) \text{ satisfying} \\ -\big[y_1 \nabla f_1(\bar{x}) + \cdots + y_m \nabla f_m(\bar{x})\big] \in N_X(\bar{x}) \\ \text{with } \begin{cases} y_i \geq 0 & \text{for } i \in [1, s] \text{ active at } \bar{x}, \\ y_i = 0 & \text{for } i \in [1, s] \text{ inactive at } \bar{x}, \\ y_i \text{ free} & \text{for } i \in [s+1, m], \end{cases} \\ \text{have } y_i = 0 \text{ for every } i \text{ such that } f_i \text{ is } \textbf{not affine.} \end{cases}$$

Thus, if $(\mathcal{P})$ has only linear constraints, no constraint qualification is needed, and the Kuhn-Tucker conditions are necessary for the local optimality of any feasible solution $\bar{x}$.

Proof. For simplicity we can suppose the notation is chosen so that the indices of the affine inequality constraints are $i \in [1, q]$ while those of the affine equality constraints are $i \in [r+1, m]$. Let $X' = \big\{x \in X \,\big|\, f_i(x) \leq 0 \text{ for } i \in [1, q],\ f_i(x) = 0 \text{ for } i \in [r+1, m]\big\}$, so that $(\mathcal{P})$ can be identified with the problem

$$(\mathcal{P}') \qquad \text{minimize } f_0(x) \text{ over all } x \in X' \text{ such that } f_i(x) \begin{cases} \leq 0 & \text{for } i = q+1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, r. \end{cases}$$

In particular the locally optimal solutions are the same in both cases. Let $\bar{x}$ be one at which the refined constraint qualification holds.

Because the constraint system specifying X' is linear, Theorem 10 can be applied to it without need for checking any constraint qualification (as was established above). In consequence, $N_{X'}(\bar{x})$ consists of the vectors of the form

$$z' = \sum_{i=1}^{q} y_i \nabla f_i(\bar{x}) + \sum_{i=r+1}^{m} y_i \nabla f_i(\bar{x}) \text{ with } \begin{cases} y_i \geq 0 & \text{for } i \in [1, q] \text{ active at } \bar{x}, \\ y_i = 0 & \text{for } i \in [1, q] \text{ inactive at } \bar{x}, \\ y_i \text{ free} & \text{for } i \in [r+1, m]. \end{cases}$$

Our strategy is to apply Theorem 11 to $(\mathcal{P}')$ in place of $(\mathcal{P})$, using this information. We do have X' closed as well as regular at all of its points, because X' is polyhedral. The basic constraint qualification for $(\mathcal{P}')$ obliges us to examine the possibilities of having

$$-\sum_{i=q+1}^{r} y_i \nabla f_i(\bar{x}) \in N_{X'}(\bar{x}) \text{ with } \begin{cases} y_i \geq 0 & \text{for } i \in [q+1, s] \text{ active at } \bar{x}, \\ y_i = 0 & \text{for } i \in [q+1, s] \text{ inactive at } \bar{x}, \\ y_i \text{ free} & \text{for } i \in [s+1, r]. \end{cases}$$

The issue is whether this necessitates $y_i = 0$ for all $i \in [q+1, r]$. It does, through the direct combination of the formula for vectors $z' \in N_{X'}(\bar{x})$ and the refined constraint qualification being satisfied at $\bar{x}$. The Kuhn-Tucker conditions for $(\mathcal{P}')$ must therefore hold at $\bar{x}$: there must exist Lagrange multipliers $\bar{y}_i$ for $i \in [q+1, r]$ such that

$$-\big[\nabla f_0(\bar{x}) + \sum_{i=q+1}^{r} \bar{y}_i \nabla f_i(\bar{x})\big] \in N_{X'}(\bar{x}) \text{ with } \begin{cases} \bar{y}_i \geq 0 & \text{for } i \in [q+1, s] \text{ active at } \bar{x}, \\ \bar{y}_i = 0 & \text{for } i \in [q+1, s] \text{ inactive at } \bar{x}, \\ \bar{y}_i \text{ free} & \text{for } i \in [s+1, r]. \end{cases}$$

Invoking the formula for vectors $z' \in N_{X'}(\bar{x})$ once more, we end up with the additional Lagrange multipliers needed to see that the Kuhn-Tucker conditions for $(\mathcal{P})$ hold at $\bar{x}$. □

Convex programming with linear constraints: It follows from Theorem 12 along with Theorem 11 that in any convex programming problem $(\mathcal{P})$ with only linear constraints, as in linear programming and quadratic programming, the existence of a Lagrange multiplier vector $\bar{y}$ satisfying the Kuhn-Tucker conditions with $\bar{x}$ is *necessary and sufficient* for the local—indeed *global*—optimality of $\bar{x}$.

Second-order conditions: Until now in the study of Lagrange multipliers, we have been occupied with only first-order conditions. The full theory of second-order necessary conditions and sufficient conditions for local optimality is subtle and complicated. Here we'll be content with looking at a sufficient condition favored in the development of numerical methods.

THEOREM 13 (second-order optimality in conventional format). *Consider the case of problem $(\mathcal{P})$ where X is polyhedral and the functions $f_0, f_1, \ldots, f_m$ are of class $\mathcal{C}^2$. Suppose that $\bar{x}$ and $\bar{y}$ satisfy the Kuhn-Tucker conditions and, in addition,*

$$w \cdot \nabla^2_{xx} L(\bar{x}, \bar{y}) w > 0 \text{ for all } w \neq 0 \text{ such that}$$
$$w \in T_X(\bar{x}) \text{ and } \nabla f_i(\bar{x}) \cdot w \begin{cases} \leq 0 & \text{for active } i \in [1, s] \\ = 0 & \text{for inactive } i \in [s+1, m] \textbf{ and for } i = 0. \end{cases}$$

Then $\bar{x}$ is a locally optimal solution to $(\mathcal{P})$.

Proof. For a penalty parameter value $\rho > 0$ of magnitude yet to be determined, consider the problem

$$\text{minimize } f(x, u) := f_0(x) + \sum_{i=1}^{m} \bar{y}_i \big[f_i(x) - u_i\big] + \frac{\rho}{2} \sum_{i=1}^{m} \big[f_i(x) - u_i\big]^2 \text{ over } X \times D,$$

where D is the box formed by product of s intervals $(-\infty, 0]$ and $m - s$ intervals $[0, 0]$. This is a problem in which a $\mathcal{C}^2$ function f is minimized over a polyhedral set, and the sufficient condition in Theorem 6(b) is therefore applicable for establishing local optimality. Is this condition satisfied at $(\bar{x}, \bar{u})$, where $\bar{u}_i = f_i(\bar{x})$? It turns out that under our hypothesis it is, provided ρ is high enough. This will be verified shortly, but first suppose it's true, in order to see where it leads. Suppose, in other words, that $f(x,u) \geq f(\bar{x}, \bar{u})$ for all $(x,u) \in X \times D$ in some neighborhood of $(\bar{x}, \bar{u})$.

Then for feasible $x \in X$ near enough to $\bar{x}$, the vector $u(x) := \big(f_1(x), \ldots, f_m(x)\big)$ in D will (by the continuity of the f_i's) be near to $u(\bar{x}) = \bar{z}$ with $f(x, u(x)) \geq f(\bar{x}, \bar{u})$. But $f(x, u(x)) = f_0(x)$ when x is feasible, and in particular $f(\bar{x}, \bar{u}) = f_0(\bar{x})$. It follows that $f_0(x) \geq f_0(\bar{x})$ for all feasible x in some neighborhood of $\bar{x}$, and we conclude that $\bar{x}$ is locally optimal in the given problem.

We proceed now with verifying that for large values of ρ the sufficient condition in Theorem 6(b) is satisfied for the local optimality of $(\bar{x}, \bar{u})$ in the problem of minimizing f over $X \times D$. The condition in question involves first and second partial derivatives of f as well as the tangent cone to the box $X \times D$, which from the characterization given earlier for tangents to boxes can be expressed in the product form

$$T_{X \times D}(\bar{x}, \bar{u}) = T_X(\bar{x}) \times T_D(\bar{u}).$$

Specifically, the condition requires that

$$\begin{gathered}\nabla f(\bar{x}, \bar{u}) \cdot (w, z) \geq 0 \text{ for all } (w,z) \text{ in } T_X(\bar{x}) \times T_D(\bar{u}), \\ (w,z) \cdot \nabla^2 f(\bar{x}, \bar{w})(w,z) > 0 \text{ for all } (w,z) \neq (0,0) \text{ in } T_X(\bar{x}) \times T_Z(\bar{u}) \\ \qquad\qquad\qquad \text{with } \nabla f(\bar{x}, \bar{u}) \cdot (w,z) = 0.\end{gathered}$$

The first partial derivatives are

$$\begin{aligned}\frac{\partial f}{\partial x_j}(x,u) &= \frac{\partial L}{\partial x_j}(x, \bar{y}) + \rho \sum_{i=1}^m \big[f_i(x) - u_i\big] \frac{\partial f_i}{\partial x_j}(x), \\ \frac{\partial f}{\partial u_i}(x,u) &= -\bar{y}_i - \rho\big[f_i(x) - u_i\big],\end{aligned}$$

while the second partial derivatives are

$$\begin{aligned}&\frac{\partial^2 f}{\partial x_k \partial x_j}(x,u) = \frac{\partial^2 L}{\partial x_k \partial x_j}(x, \bar{y}) + \rho \sum_{i=1}^m \big[f_i(x) - u_i\big] \frac{\partial^2 f_i}{\partial x_k \partial x_j}(x) + \rho \frac{\partial f_i}{\partial x_k}(x) \frac{\partial f_i}{\partial x_j}(x), \\ &\frac{\partial^2 f}{\partial u_l \partial x_j}(x,u) = -\rho \frac{\partial f_l}{\partial x_j}(x), \qquad\qquad \frac{\partial^2 f}{\partial x_k \partial u_i}(x,u) = -\rho \frac{\partial f_i}{\partial x_k}(x), \\ &\frac{\partial^2 f}{\partial u_l \partial u_i}(x,u) = \begin{cases} 1 & \text{if } l = i \\ 0 & \text{if } l \neq i. \end{cases}\end{aligned}$$

Because $f_i(\bar{x}) - \bar{u}_i = 0$, we obtain that

$$\nabla f(\bar{x},\bar{u}){\cdot}(w,z) = \nabla_x L(\bar{x},\bar{y}){\cdot}w - \bar{y}{\cdot}z,$$
$$(w,z){\cdot}\nabla^2 f(\bar{x},\bar{u})(w,z) = w{\cdot}\nabla^2_{xx} L(\bar{x},\bar{y})w + \rho\sum_{i=1}^{m}\big[\nabla f_i(\bar{x}){\cdot}w - z_i\big]^2.$$

The sufficient condition we wish to verify (for ρ large) thus takes the form:

$$\begin{aligned}
&\nabla_x L(\bar{x},\bar{y}){\cdot}w \ge 0 \text{ for all } w \in T_X(\bar{x}), \qquad \bar{y}{\cdot}z \le 0 \text{ for all } z \in T_D(\bar{u}),\\
&w{\cdot}\nabla^2_{xx} L(\bar{x},\bar{y})w + \rho\sum_{i=1}^{m}\big[\nabla f_i(\bar{x}){\cdot}w - z_i\big]^2 > 0 \text{ for all } (w,z) \ne (0,0) \text{ with}\\
&\qquad w \in T_X(\bar{x}), \quad \nabla_x L(\bar{x},\bar{y}){\cdot}w = 0, \quad z \in T_D(\bar{u}), \quad \bar{y}{\cdot}z = 0.
\end{aligned}$$

Here the first-order inequalities merely restate the relations $-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x})$ and $\bar{y} \in N_D(\bar{u})$ (equivalent to $\bar{u} \in N_Y(\bar{y})$, as seen before), which hold by assumption. In the second-order condition we obviously do have strict inequality when $w = 0$ and $z \ne 0$, since the quadratic expression reduces in that case to $\rho|z|^2$. Therefore, we can limit attention to demonstrating strict inequality in cases where $w \ne 0$, or more specifically (through rescaling), where $|w| = 1$. From the form of D and $\bar{u}$ we know

$$\left.\begin{array}{l} z \in T_D(\bar{u}) \\ \bar{y}{\cdot}z = 0 \end{array}\right\} \iff \begin{cases} z_i \text{ free} & \text{for inactive } i \in [1,s],\\ z_i \le 0 & \text{for active } i \in [1,s] \text{ with } \bar{y}_i = 0,\\ z_i = 0 & \text{for active } i \in [1,s] \text{ with } \bar{y}_i > 0,\\ z_i = 0 & \text{for } i \in [s+1,m], \end{cases}$$

so for any $w \ne 0$ in $T_X(\bar{x})$ the minimum of the quadratic expression with respect to $z \in T_D(\bar{u})$ with $\bar{y}{\cdot}z\rangle = 0$ will be attained when

$$z_i = z_i(w) = \begin{cases} \nabla f_i(\bar{x}){\cdot}w & \text{for inactive } i \in [1,s],\\ \min\big\{0, \nabla f_i(\bar{x}){\cdot}w\big\} & \text{for active } i \in [1,s] \text{ with } \bar{y}_i = 0,\\ 0 & \text{for active } i \in [1,s] \text{ with } \bar{y}_i > 0,\\ 0 & \text{for } i \in [s+1,m]. \end{cases}$$

Thus, we can limit attention further to pairs (w,z) not only with $|w| = 1$ but also with $z_i = z_i(w)$. We'll suppose the claim for this special case is false and argue toward a contradiction.

If the claim is false, there has to be a sequence of values $\rho^\nu \to \infty$ along with vectors $w^\nu \in T_X(\bar{x})$ with $|w^\nu| = 1$ such that $\nabla_x L(\bar{x},\bar{y}){\cdot}w^\nu\rangle = 0$ and

$$w^\nu, \nabla^2_{xx} L(\bar{x},\bar{y})w^\nu + \rho^\nu\sum_{i=1}^{m}\Big[\nabla f_i(\bar{x}){\cdot}w^\nu - z_i(w^\nu)\Big]^2 \le 0,$$

and hence in particular

$$\begin{aligned} w^\nu \cdot \nabla^2_{xx} L(\bar{x}, \bar{y}) w^\nu \ &\leq \ 0, \\ \Big[\nabla f_i(\bar{x}) \cdot w^\nu - z_i(w^\nu) \Big]^2 \ &\leq \ -\frac{1}{\rho^\nu} w^\nu \cdot \nabla^2_{xx} L(\bar{x}, \bar{y}) w^\nu \ \text{ for all } i. \end{aligned}$$

Because the sequence $\{w^\nu\}_{\nu=1}^\infty$ is bounded, it has a cluster point $\bar{w}$. By the continuity of the expressions involved, and the closedness of tangent cones, we get in the limit that

$$\begin{aligned} &\bar{w} \in T_X(\bar{x}), \qquad |\bar{w}| = 1, \qquad \nabla_x L(\bar{x}, \bar{y}) \cdot \bar{w} = 0, \qquad \bar{y} \cdot \bar{z}(\bar{w}) = 0, \\ &\bar{w} \cdot \nabla^2_{xx} L(\bar{x}, \bar{y}) \bar{w} \leq 0, \qquad \big[\nabla f_i(\bar{x}) \cdot \bar{w} - z_i(\bar{w}) \big]^2 \leq 0 \ \text{ for all } i. \end{aligned}$$

The final batch of inequalities says that $\nabla f_i(\bar{x}) \cdot \bar{w} = z_i(\bar{w})$ for all i, which means

$$\nabla f_i(\bar{x}) \cdot w \begin{cases} \leq 0 & \text{for active } i \in [1, s] \text{ with } \bar{y}_i = 0, \\ = 0 & \text{for inactive } i \in [1, s] \text{ with } \bar{y}_i > 0, \\ = 0 & \text{for } i \in [s+1, m]. \end{cases}$$

These conditions along with the fact that $F(\bar{x}) \in N_Y(\bar{y})$ and

$$\nabla_x L(\bar{x}, \bar{y}) \cdot \bar{w} = \nabla f_0(\bar{x}) \cdot \bar{w} + \bar{y}_1 \nabla f_1(\bar{x}) \cdot \bar{w} + \cdots + \bar{y}_m \nabla f_m(\bar{x}) \cdot \bar{w}$$

also imply $\nabla f_0(\bar{x}) \cdot \bar{w} = 0$. We have arrived therefore at a vector $\bar{w} \neq 0$ for which the second-order condition in the theorem is violated. This finishes the proof. □

Specialization to linear constraints: If all the functions f_i are affine, the sufficient condition in Theorem 13 reduces to the one in Theorem 6(b). Indeed, the feasible set C is polyhedral then, and the indicated constraints on w describe the vectors $w \in T_C(\bar{x})$ with $\nabla f_0(\bar{x}) \cdot w = 0$. At the same time one has $\nabla^2_{xx} L(\bar{x}, \bar{y}) = \nabla^2 f_0(\bar{x})$.

Curvature of the feasible set when nonlinear constraints are present: The key feature of Theorem 13, in contrast to Theorem 6(b), is that the possible curvature of the boundary of C around $\bar{x}$ has been accounted for through the replacement of $\nabla^2 f_0(\bar{x})$ by the matrix $\nabla^2_{xx} L(\bar{x}, \bar{y}) = \nabla^2 f_0(\bar{x}) + \bar{y}_1 \nabla^2 f_1(\bar{x}) + \cdots + \bar{y}_m \nabla^2 f_m(\bar{x})$.

Caution: It's not true that the condition analogous to the one in Theorem 13, but merely with $w \cdot \nabla^2_{xx} L(\bar{x}, \bar{y}) w \geq 0$, is necessary for the local optimality in $(\mathcal{P})$, not even if the basic constraint qualification is satisfied. Second-order necessary conditions generally have to resort to a combination of several different Lagrange multiplier vectors $\bar{y}$ at the same point $\bar{x}$. More sharply developed sufficient conditions take on such a form as well. Such necessary or sufficient conditions lie beyond the scope of our efforts here.