PART I. OPTIMIZATION: CLASSICAL APPROACHES
(LECTURE 6)

Shpilev Petr Valerievich

Faculty of Mathematics and Mechanics, SPbU

September, 2025

Санкт-Петербургский государственный университет

Comments

In this lecture, we will delve into first- and second-order necessary conditions for constrained optimization, with a focus on the well-known Karush–Kuhn–Tucker (KKT) conditions. We will begin by examining the KKT conditions for a box constraint and explore the geometry behind feasible directions and tangent cones. The lecture includes a detailed proof of Lemma 5, illustrating its relevance to optimality conditions. We will also discuss Farkas' Lemma, a classical result in the theory of alternatives, and use it to establish the existence of Lagrange multipliers and verify the KKT conditions. Building on these foundations, we will investigate second-order optimality conditions, both necessary and sufficient, and their application to characterize the behavior of solutions at optimal points. The lecture concludes with a deeper exploration of the critical cone and its role in second-order conditions.

Let's consider the general problem

$$\min_{x\in\Omega} f(x), \text{ where } \Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0, \ i \in \mathcal{E}; \ c_i(x) \geq 0, \ i \in \mathcal{I}\} \qquad (*)$$

We define the Lagrangian function for this problem as follows:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(x).$$

Theorem 21 (First-Order Necessary Conditions)

Suppose that x^* is a local solution of $(*)$, that the functions f and c_i in $(*)$ are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components $\lambda_i^*, i \in \mathcal{E}\cup\mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \qquad (7a)$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \qquad (7b)$$

$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I}, \qquad (7c)$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I}, \qquad (7d)$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}\cup\mathcal{I}. \qquad (7e)$$

F-O Necessary condition
KKT
Tangent Cone & Feasible Directions
Necessary condition
Second-Order Necessary condition
Sufficient Conditions

Comments

When studying constrained optimization, one of the most fundamental tools is the Lagrangian function. It allows us to combine the objective function with the constraints into a single mathematical object. For a problem where we minimize the function $f(x)$, subject to both equality and inequality constraints, the Lagrangian is defined as $f(x)$ minus the sum over all multipliers λ_i times the constraint $c_i(x)$. This construction captures how constraints influence the optimal point.

The central result is the first-order necessary conditions, which describe what must hold at any local solution. If a point x^* is indeed a local minimizer, and if the objective and constraint functions are continuously differentiable, and if the linear independence constraint qualification, or LICQ, is satisfied at x^*, then there exists a multiplier vector λ^* with remarkable properties. Together, the pair x^* and λ^* must satisfy several conditions. First, the gradient of the Lagrangian with respect to x must vanish at x^*, which expresses stationarity. Second, all equality constraints must hold exactly. Third, all inequality constraints must be satisfied. Fourth, the multipliers associated with inequalities must be nonnegative. Finally, the complementarity condition must hold: for every constraint, either the multiplier or the constraint value is zero.

These conditions together are powerful because they transform a constrained problem into algebraic relationships between gradients and multipliers. They give us the necessary structure to analyze and eventually compute optimal solutions. They are called "first-order" because they rely only on first derivatives, making them both general and computationally useful.

The conditions (7a)–(7d) are often known as the *Karush–Kuhn–Tucker conditions*, or KKT conditions for short.

The conditions (7d) are called *complementarity conditions*. They imply that either constraint i is active or $\lambda_i^* = 0$, or both.

In particular, for inequality constraints:

- If constraint i is inactive at x^* (i.e. $i \notin \mathcal{A}(x^*)$), then $\lambda_i^* = 0$.
- Thus, we can rewrite (7a) as:

$$0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*).$$

Definition: Strict Complementarity

Given a local solution x^* of $(*)$ and a vector λ^* satisfying (7a)–(7d), we say that *strict complementarity conditions* holds if:

- For each index $i \in \mathcal{I}$, exactly one of λ_i^* and $c_i(x^*)$ is zero.
- That is, $\lambda_i^* > 0$ for each $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.

Comments

The collection of first-order necessary conditions is more commonly known as the Karush–Kuhn–Tucker, or KKT, conditions. These conditions are central to nonlinear optimization. They refine the idea of optimality by combining feasibility, stationarity, and complementarity in one framework.

The stationarity condition states that the gradient of the Lagrangian vanishes when evaluated at the solution. Feasibility requires that all equality constraints are exactly satisfied and all inequality constraints are respected. But the most distinctive part is complementarity. This condition links multipliers and constraints: for each inequality constraint, either the multiplier is strictly positive and the constraint is active, or the multiplier is zero when the constraint is inactive. Both being positive at the same time is impossible.

A stronger version is known as strict complementarity. It asserts that for every active inequality constraint, the corresponding multiplier is not only nonnegative but strictly positive. In other words, if a constraint touches the solution boundary, it contributes actively to defining the geometry, and the multiplier reflects that influence. Strict complementarity is not guaranteed in all problems, but when it holds, algorithms often perform more reliably, because it clarifies which constraints are truly active.

Another important refinement is that in the stationarity condition, the sum over constraints can be restricted to the active set. This means only those constraints that “bite” at the solution matter in forming the balance of gradients. This interpretation reveals the beauty of the KKT system: the optimizer balances the gradient of the objective function against a weighted combination of gradients of only the active constraints.

Example 6: KKT Conditions for a Box Constraint

F-O Necessary condition
KKT
Tangent Cone & Feasible Directions
Necessary condition
Second-Order Necessary condition
Sufficient Conditions

Example 6

Consider the feasible region Ω defined by the four constraints:

$$x_1 + x_2 \leq 1, \quad x_1 - x_2 \leq 1, \quad -x_1 + x_2 \leq 1, \quad -x_1 - x_2 \leq 1.$$

Restating the problem in standard form with an objective:

$$\min_x \left(x_1 - \tfrac{3}{2}\right)^2 + \left(x_2 - \tfrac{1}{2}\right)^4 \quad \text{s.t.} \quad \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0.$$

The solution is $x^* = (1,\ 0)^T$, where the first two constraints are active. Denoting active ones as $c_1,\ c_2$, and inactive as $c_3,\ c_4$, we compute:

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

$$\lambda^* = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}^T \quad \Rightarrow \quad \text{KKT conditions are satisfied.}$$

Comments

To see how these abstract conditions work in practice, let us consider a concrete example. Imagine a feasible region defined by four linear inequalities: $x_1 + x_2 \leq 1$, $x_1 - x_2 \leq 1$, $-x_1 + x_2 \leq 1$, and $-x_1 - x_2 \leq 1$. Together, these inequalities form a diamond-shaped region centered at the origin.

We now minimize a nonlinear objective: the squared term of $x_1 - \frac{3}{2}$ plus the quartic term of $x_2 - \frac{1}{2}$. Without constraints, the unconstrained minimum would be at $x_1 = \frac{3}{2}$ and $x_2 = \frac{1}{2}$. However, this point lies outside the diamond. Hence, the true solution must occur on the boundary. Careful examination shows that the minimizing point is $x^* = (1, 0)^T$. At this point, the first two constraints are active, while the remaining two are inactive.

We compute the gradient of the objective at this solution, which turns out to be the vector with components -1 and $-\frac{1}{2}$. The gradients of the active constraints are vectors with components $(-1, -1)$ and $(-1, 1)$. The KKT conditions can now be checked directly. By solving for the multipliers, we find $\lambda_1^* = \frac{3}{4}$ and $\lambda_2^* = \frac{1}{4}$, while the remaining multipliers are zero. All conditions are satisfied: stationarity holds, constraints are feasible, nonnegativity of multipliers is respected, and complementarity is fulfilled.

This simple but rich example shows that the KKT framework gives a systematic way to verify optimality and to compute supporting multipliers.

We now relate the tangent cone $T_\Omega(x^*)$ to the set $\mathcal{F}(x^*)$ of first-order feasible directions. This relation requires a constraint qualification (LICQ).

Lemma 5: Tangent Cone and First-Order Feasible Directions

Let x^* be a feasible point. Then:

(i) $T_\Omega(x^*) \subset \mathcal{F}(x^*)$,

(ii) If LICQ holds at x^*, then $T_\Omega(x^*) = \mathcal{F}(x^*)$.

In the proof below we will use the notation

$$A(x^*)^T = [\nabla c_i(x^*)]_{i \in \mathcal{A}(x^*)} \quad \text{(matrix of active constraint gradients)}.$$

Comments

Beyond specific examples, we can deepen our understanding by connecting KKT conditions to geometric concepts. One of the most important is the relationship between the tangent cone at a point and the set of first-order feasible directions. The tangent cone describes all directions in which we can move infinitesimally while staying inside the feasible region. The set of first-order feasible directions, on the other hand, is defined through linear approximations of the constraints.

A general result shows that the tangent cone is always contained within the set of first-order feasible directions. Intuitively, any truly feasible infinitesimal motion must also satisfy the linearized conditions. However, the reverse inclusion does not always hold. In order for the two sets to coincide, a constraint qualification is required. Specifically, when the linear independence constraint qualification is satisfied at the point, the tangent cone and the set of first-order feasible directions are identical.

This equality is crucial because it guarantees that linearization provides a faithful representation of the local geometry. In optimization algorithms, we rely on linearized models to propose steps and directions. If those models overestimate feasible directions, we may encounter false paths. But if LICQ holds, the geometry and algebra align perfectly, making the algorithms reliable.

Formally, the gradients of all active constraints are gathered into a matrix, often denoted $A(x^*)$. This matrix provides a compact way to represent the linearized restrictions, and it becomes a central object in both theory and computation.

Proof: Without loss of generality, assume all constraints $c_i(\cdot)$, $i = 1, 2, \ldots, m$, are active at x^*. (This can be arranged by removing all inactive constraints near x^* and renumbering the active ones.)

To prove (i), let $\{z_k\}$ and $\{t_k\}$ be the sequences for which the following equality holds:

$$\lim_{k\to\infty} \frac{z_k - x^*}{t_k} = d. \qquad (**)$$

Note that $t_k > 0$ for all k. This gives:

$$z_k = x^* + t_k d + o(t_k).$$

Now for any $i \in \mathcal{E}$, applying Taylor's theorem:

$$0 = \frac{1}{t_k} c_i(z_k) = \frac{1}{t_k}\left[c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)\right] =$$
$$= \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}.$$

Taking the limit as $k \to \infty$, the final term vanishes. Thus, we obtain:

$$\nabla c_i(x^*)^T d = 0.$$

Comments

At this point, we begin the detailed proof of the lemma. The first step is to simplify the situation without losing generality. If we focus only on those constraints that are active at the candidate solution, then the analysis becomes cleaner. Inactive constraints do not affect the local geometry, so they can be safely removed, and the active ones can be relabeled. Having done this, we consider sequences that approach the point of interest, together with a sequence of positive scalars that tend to zero.

The condition under study is that the normalized difference between the sequence points and the reference point converges to a certain direction. Intuitively, this means we are zooming in closer and closer to the solution and watching how feasible points approach it. When we write the points in the form of the reference point plus the small scalar times the direction plus higher-order terms, we are effectively applying the definition of differentiability.

Next, we expand each active constraint function around the solution using Taylor's theorem. Because each active constraint vanishes at the reference point, the first nonzero contribution comes from its gradient evaluated at the solution, multiplied by the candidate direction. Higher-order terms become negligible when divided by the small scalar.

As a result, the limit condition enforces that the gradient of each equality constraint at the point must be orthogonal to the direction under consideration. This step confirms that every tangent direction must lie in the null space of the active constraint gradients. This characterization will be essential for completing the proof in later steps.

For active inequality constraints $i \in \mathcal{A}(x^*) \cap \mathcal{I}$, we similarly obtain:

$$0 \le \frac{1}{t_k} c_i(z_k) = \frac{1}{t_k}\left[c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)\right] = \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}.$$

Thus, taking the limit as $k \to \infty$, we conclude:

$$\nabla c_i(x^*)^T d \ge 0.$$

To prove (ii), we use the implicit function theorem. Since LICQ holds, the matrix $A(x^*)$ of active constraint gradients has full row rank m. Let $Z \in \mathbb{R}^{n \times (n-m)}$ be a matrix whose columns form a basis for the null space of $A(x^*)$:

$$Z \text{ has full column rank}, \quad A(x^*)Z = 0.$$

Pick any $d \in \mathcal{F}(x^*)$, and let $\{t_k\}_{k=0}^{\infty}$ be any sequence of positive scalars with $\lim\limits_{k\to\infty} t_k = 0$.

Define the system of equations $R : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ as:

$$R(z, t) = \begin{bmatrix} c(z) - tA(x^*)d \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \qquad (***)$$

Comments

After establishing the condition for equality constraints, we now turn to inequality constraints that are active at the solution. The logic is similar, but there is a subtle difference. For an inequality constraint, feasibility only requires that the function value remain nonnegative, not necessarily equal to zero once we move away. By expanding the inequality constraint around the solution using Taylor's theorem, we see that its gradient dotted with the direction cannot be negative; otherwise, the constraint would be violated for small perturbations.

Taking the limit, the higher-order terms vanish, and we obtain a nonnegativity condition for the directional derivative. This shows that feasible tangent directions must not decrease active inequalities at the point. Having handled both equality and inequality constraints, the proof advances to its second part. Here, we use a powerful tool from analysis, the implicit function theorem. Because the linear independence constraint qualification holds, the active constraint gradients are linearly independent, which ensures that the associated matrix has full row rank.

This property is crucial: it allows us to define a complementary basis for the space of directions, represented by a matrix whose columns span the null space of the constraints. With this construction, we design a system of equations that couples the original constraints with additional linear conditions involving the null-space basis.

The idea is that solving this system will generate feasible points that move away from the solution in a controlled way along the chosen direction. This prepares the ground for the final argument, where we verify that such constructed sequences indeed realize the definition of tangent vectors.

Proof of Lemma 5 (cont.)

We claim that the solutions $z = z_k$ of this system for small $t = t_k > 0$ give a feasible sequence that approaches x^* and satisfies the definition of tangent vector.

At $t = 0$, $z = x^*$, and the Jacobian of R at this point is

$$\nabla_z R(x^*, 0) = \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix},$$

which is nonsingular by construction of Z.

Thus, by the implicit function theorem, the system $(***)$ has a unique solution z_k for small $t_k > 0$, and from $(***)$ and the definition of the set of linearized feasible directions $(\mathcal{F}(x))$:

$$i \in \mathcal{E} \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d = 0,$$
$$i \in \mathcal{A}(x^*) \cap \mathcal{I} \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d \geq 0,$$

so z_k is feasible.

Comments

The next step is to argue why the constructed system of equations produces feasible points close to the solution. At the specific case where the perturbation parameter equals zero, the only solution is the reference point itself.

To study what happens for small positive values of the parameter, we examine the Jacobian of the system with respect to the variable of interest. This Jacobian turns out to be a block matrix consisting of the active constraint gradients and the transpose of the null-space basis. By design, this matrix is nonsingular, which means we can apply the implicit function theorem. The theorem guarantees that for sufficiently small parameter values, there exists a unique smooth solution to the system.

This solution defines a sequence of points converging back to the original point as the parameter tends to zero. What remains is to check that these points indeed satisfy feasibility conditions. By inspecting the definitions in the system, we see that equality constraints are preserved exactly, while inequality constraints are satisfied in a nondecreasing manner.

This ensures that the constructed points are not only mathematically consistent but also feasible with respect to the optimization problem. Conceptually, this confirms that we can approximate feasible paths by smooth curves that start at the solution and extend in the direction of interest. This bridge between algebraic conditions and geometric intuition is the key contribution of this part of the proof.

It remains to verify that $(**)$ holds for this choice of $\{z_k\}$. Since $R(z_k, t_k) = 0$ and by Taylor's theorem:

$$\begin{aligned} 0 = R(z_k, t_k) &= \begin{bmatrix} c(z_k) - t_k A(x^*) d \\ Z^T (z_k - x^* - t_k d) \end{bmatrix} \\ &= \begin{bmatrix} A(x^*)(z_k - x^*) + o(\|z_k - x^*\|) - t_k A(x^*) d \\ Z^T (z_k - x^* - t_k d) \end{bmatrix} \\ &= \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + o(\|z_k - x^*\|). \end{aligned}$$

Dividing by t_k and using nonsingularity of the coefficient matrix, we get:

$$\frac{z_k - x^*}{t_k} = d + o\left(\frac{\|z_k - x^*\|}{t_k} \right).$$

Thus, $(**)$ is satisfied (for $x = x^*$), and $d \in T_\Omega(x^*)$ for arbitrary $d \in \mathcal{F}(x^*)$, completing the proof. □

Comments

The final part of the proof focuses on verifying the convergence property of the constructed sequence. We return to the condition that defines tangent vectors, which requires that the normalized difference between sequence points and the reference point converges to the candidate direction.

By substituting the system representation and applying Taylor expansion, we can express the residual as a combination of the active constraint gradients and the null-space basis applied to the deviation from the reference point. The higher-order terms again vanish in the limit. The coefficient matrix in front of the deviation is nonsingular, and this allows us to conclude that the deviation, once scaled by the parameter, must approach the chosen direction.

In other words, the constructed sequence of feasible points not only stays within the constraint set but also exhibits the precise directional behavior required by the definition of tangent vectors. This completes the argument by showing that any feasible direction indeed corresponds to a valid tangent direction at the solution. The lemma thus establishes an equivalence between linearized feasible directions and the tangent cone.

This result plays a central role in optimization theory, as it provides the rigorous link between local linear approximations and actual feasible displacements. By completing this proof, we confirm that the geometric concept of tangent cones is perfectly aligned with the analytic characterization obtained through constraint gradients.

$$\min_{x\in\Omega} f(x), \text{ where } \Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\} \qquad (*)$$

Theorem 22

If x^* is a local solution of $(*)$, then we have

$$\nabla f(x^*)^T d \geq 0, \quad \text{for all } d \in T_\Omega(x^*).$$

Proof: Suppose for contradiction there exists d with $\nabla f(x^*)^T d < 0$. Let $\{z_k\}$ and $\{t_k\}$ be the sequences satisfying the Definition of tangent to Ω for d. Then (since $z_k = x^* + t_k d + o(t_k)$) we have

$$\begin{aligned} f(z_k) &= f(x^*) + (z_k - x^*)^T \nabla f(x^*) + o(\|z_k - x^*\|) \\ &= f(x^*) + t_k d^T \nabla f(x^*) + o(t_k). \end{aligned}$$

Since $d^T \nabla f(x^*) < 0$, for sufficiently large k

$$f(z_k) < f(x^*) + \tfrac{1}{2} t_k d^T \nabla f(x^*) < f(x^*),$$

contradicting local optimality of x^*. □

Comments

The necessary condition for local optimality can be expressed in terms of the gradient of the objective function and the feasible directions at a point. Suppose we are analyzing a candidate point that might be a local minimizer. The key observation is that if this point is indeed locally optimal, then moving in any feasible direction cannot decrease the value of the function, at least to first order. Formally, this means that the inner product between the gradient at the point and any tangent direction must be nonnegative.

The gradient represents the direction of steepest increase, and its inner product with a feasible direction measures how the function changes along that direction. If a feasible direction produced a strictly negative value for this product, it would indicate the existence of a descent direction, contradicting local minimality. The proof relies on constructing sequences of feasible points that approximate the tangent vector.

Proof by contradiction. Suppose that there is a direction whose inner product with the gradient is strictly negative. By expanding the objective function around the candidate point using Taylor's theorem, we see that the first-order term dominates small perturbations. If the inner product is negative, we can always find feasible perturbations that reduce the function value, no matter how close we stay to the candidate point.

This logical contradiction shows that for a true local solution, every feasible direction must satisfy the nonnegativity condition. This result forms the foundation for developing stronger optimality conditions and motivates the search for examples where the condition holds but local minimality fails.

The converse does not hold

The converse of Theorem 22 does not hold: it is possible that

$$\nabla f(x^*)^T d \geq 0 \quad \text{for all } d \in \mathcal{T}_\Omega(x^*),$$

yet x^* is not a local minimizer.

Consider the problem:

$$\min x_2 \text{ subject to } x_2 \geq -x_1^2.$$

This problem is unbounded, but at $x^* = (0, 0)^T$, we find that all feasible directions satisfy $d_2 \geq 0$, so $\nabla f(x^*)^T d = d_2 \geq 0$.
However, x^* is not a local minimizer, since $(\alpha, -\alpha^2)^T$ has lower x_2-value and approaches x^* as $\alpha \to 0$.

Comments

Although the nonnegativity condition involving the gradient and feasible directions is essential, it does not fully characterize local minimizers. There exist cases where this condition is satisfied, yet the candidate point is not actually optimal.

A simple yet instructive example is given by minimizing the second coordinate of a vector subject to the constraint that the second coordinate is greater than or equal to the negative square of the first coordinate. Geometrically, this feasible region is the set of points lying above a parabola opening downward. Consider the origin as a candidate solution. At this point, the feasible tangent cone consists of all directions where the second component is nonnegative.

The gradient of the objective function is simply the vector pointing upward, and its inner product with any feasible direction is indeed nonnegative. This satisfies the necessary condition. However, the origin is not a minimizer. One can approach it along points lying on the parabola, which have strictly smaller values of the second coordinate. By choosing a small parameter, these points come arbitrarily close to the origin while providing better function values.

This counterexample illustrates a crucial point: the necessary condition is not sufficient. In practice, this means that checking the gradient condition alone is not enough; further analysis is required to distinguish genuine local solutions from deceptive candidates. The example also highlights the subtle role of curvature in the feasible set. Even when directions seem harmless at first order, second-order geometry can reveal directions that lead to a decrease in the objective, disqualifying the point from being a local solution. This insight motivates the development of second-order optimality conditions.

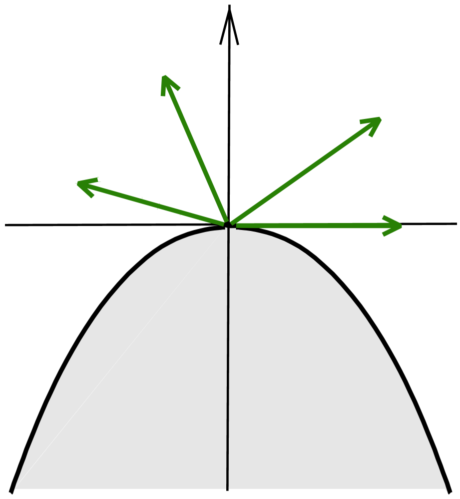

Figure: Problem $\min x_2$ subject to $x_2 \geq -x_1^2$. showing various limiting directions of feasible sequences at the point $(0,0)^T$

Comments

The graphical interpretation of this example offers additional clarity. At the origin, the set of feasible directions is restricted to those pointing upward or horizontally, since moving downward would immediately violate the inequality constraint. This explains why the gradient condition appears to be satisfied: the gradient points upward, and so its inner product with any feasible direction is nonnegative. However, the picture also reveals the weakness of relying solely on first-order analysis. The feasible set has a curved boundary, given by the parabola, and along this boundary one can find sequences of feasible points that approach the origin from below.

These sequences effectively demonstrate that the origin is not a minimizer, despite satisfying the necessary condition. The limiting directions observed in the figure emphasize the subtle interplay between tangent cones and feasible sequences. While the tangent cone captures only the immediate, first-order directions, it does not fully describe how the feasible set behaves nearby.

In our example, the tangent cone at the origin misses the fact that the boundary curves downward, permitting feasible sequences that dip below the candidate point. This mismatch between first-order information and actual local geometry is the heart of why the necessary condition fails to guarantee optimality. By studying these visualizations, one gains intuition about why stronger conditions, such as those involving second derivatives or curvature information, are needed. They provide a more complete characterization of how the objective interacts with the feasible set, ensuring that candidate solutions are not undermined by subtle geometric features invisible to first-order analysis.

Farkas' Lemma: A Classical Theorem of the Alternative

Consider the cone

$$K = \{By + Cw \mid y \geq 0\},$$

where

- $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{n \times p}$,
- $y \in \mathbb{R}^m$, $w \in \mathbb{R}^p$,

and $y \geq 0$ componentwise.

Farkas' Lemma states exactly one of the following holds for a given vector $g \in \mathbb{R}^n$:

$$\text{either } g \in K, \text{ or there exists } d \in \mathbb{R}^n \text{ such that}$$

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0.$$

The two cases are illustrated in the next Figure for the case of B with three columns, C null, and $n = 2$.

Note: In the second case, d defines a hyperplane separating g from the cone K.

Comments

To move beyond these limitations, optimization theory employs powerful results from convex analysis. One of the most fundamental is Farkas' Lemma, a classical result of the alternative. This lemma deals with cones generated by linear combinations of certain columns, where some coefficients must be nonnegative. The cone thus represents a structured subset of space that reflects feasible directions in linear constraint systems.

The lemma asserts a striking dichotomy: given a vector, either it belongs to the cone, or there exists a separating vector that certifies its exclusion. This separating vector has three defining properties: its inner product with the given vector is strictly negative, its inner product with the generating columns subject to nonnegativity is nonnegative, and it is orthogonal to the unconstrained components. Geometrically, this separating vector defines a hyperplane that strictly divides the candidate vector from the cone.

The power of Farkas' Lemma lies in providing a certificate of infeasibility in linear systems. If a candidate vector does not lie in the cone, the lemma guarantees the existence of a simple, linear witness to this fact. This principle becomes the backbone of many optimality conditions in constrained optimization. By reformulating feasibility and optimality questions in terms of cones and separating hyperplanes, one can translate nonlinear optimization problems into settings where classical linear results apply.

In particular, the lemma plays a central role in deriving first-order necessary conditions for constrained problems, bridging the gap between geometric intuition and rigorous analytic criteria. It illustrates the deep connection between linear algebra, convex geometry, and optimization theory.

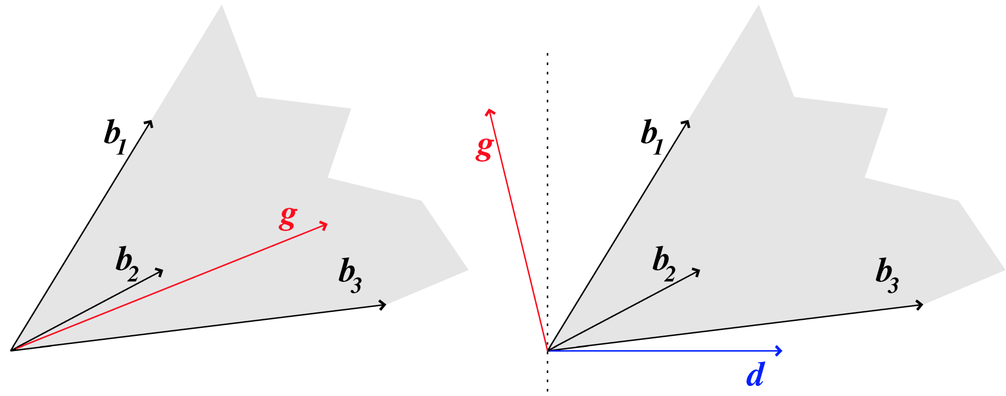

Figure: Farkas' Lemma: Either $g \in K$ (left) or there is a separating hyperplane (right).

Comments

This figure provides a simple two–dimensional illustration of the general problem of separating a vector from a cone. The cone is constructed by taking several generating rays and closing them under nonnegative linear combinations. In two dimensions, this cone looks like a wedge spanned by two or more rays. Every vector lying inside this wedge can be represented as a nonnegative combination of the generators. For example, the vector g in the left panel belongs to the cone, so it can be expressed through those generating rays with nonnegative coefficients.

The situation changes when the vector does not lie inside the cone. In the right panel, g points outside the wedge. In this case, it becomes possible to draw a separating line, sometimes called a separating hyperplane in higher dimensions. This line divides the plane into two half-spaces: all points of the cone remain on one side, while the vector g lies strictly on the other side. Geometrically, the line touches or supports the cone without cutting into it, and at the same time it excludes the vector g.

This very simple two–dimensional picture already captures the essential dichotomy: either the vector lies inside the cone and is representable as a nonnegative combination of the generators, or it lies outside and can be separated from the cone by a line. In the next step, this geometric idea will be extended far beyond the plane. What we will see is that the same reasoning works in any dimension, and the formal statement of this fact is known as Farkas' Lemma.

Lemma 6 (Farkas')

Let the cone $K = \{By + Cw \mid y \geq 0\}$ be defined as before. Given any vector $g \in \mathbb{R}^n$, we have either that $g \in K$ or that there exists $d \in \mathbb{R}^n$ satisfying the conditions:

$$g^T d < 0, \; B^T d \geq 0, \; C^T d = 0. \tag{i}$$

Proof: We first show that the two alternatives cannot hold simultaneously. If $g \in K$, then there exist vectors $y \geq 0$ and w such that

$$g = By + Cw.$$

If there also exists d such that

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0,$$

then taking inner products yields

$$0 > d^T g = d^T By + d^T Cw = (B^T d)^T y + (C^T d)^T w \geq 0,$$

where the last inequality follows from $B^T d \geq 0$, $y \geq 0$, and $C^T d = 0$. Thus, both alternatives cannot hold at the same time.

Comments

Farkas' Lemma is one of the most fundamental results in optimization and convex analysis, because it describes a precise dichotomy about the relationship between a vector and a cone. To understand its meaning, recall that a cone is a set closed under multiplication by positive scalars. In our case, the cone is defined as all combinations of the form $By + Cw$, where y is restricted to be nonnegative, while w is unrestricted. Now, given any vector g, the lemma states that one and only one of two situations must occur. Either the vector g belongs to the cone, or there exists a vector d that separates g from the cone by means of inequalities.

This second possibility is expressed by three properties: the inner product of g and d is strictly negative; the transpose of B times d is nonnegative; and the transpose of C times d equals zero. The key intuition is that g either lies inside the cone, in which case no separating vector can exist, or it lies outside, in which case we can construct such a separator.

The proof starts by showing that both alternatives cannot happen at the same time. If g is inside the cone, then it can be written as $By + Cw$, with y nonnegative. But then, if there also existed a separating vector d, we would derive a contradiction: the inner product would be strictly negative and simultaneously greater than or equal to zero. This is impossible. Hence, the two cases are mutually exclusive. The next step will be to show that at least one of them always holds.

We now show that one of the alternatives holds. To be precise, we show how to construct d with the properties

$$\mathrm{g}^{\mathsf{T}}\mathrm{d} < 0, \quad \mathrm{B}^{\mathsf{T}}\mathrm{d} \geq 0, \quad \mathrm{C}^{\mathsf{T}}\mathrm{d} = 0$$

in the case that $\mathrm{g} \notin \mathrm{K}$.
It can be shown that K is a closed set. Let $\hat{\mathrm{s}}$ be the vector in K that is closest to g in the sense of the Euclidean norm. Then $\hat{\mathrm{s}}$ is well defined and is given by the solution of the following optimization problem:

$$\min \|\mathrm{s} - \mathrm{g}\|_2^2 \quad \text{subject to } \mathrm{s} \in \mathrm{K}.$$

Since $\hat{\mathrm{s}} \in \mathrm{K}$, we have from the fact that K is a cone that $\alpha\hat{\mathrm{s}} \in \mathrm{K}$ for all scalars $\alpha \geq 0$. Since $\|\alpha\hat{\mathrm{s}} - \mathrm{g}\|_2^2$ is minimized by $\alpha = 1$, we have by simple calculus that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\|\alpha\hat{\mathrm{s}} - \mathrm{g}\|_2^2\Big|_{\alpha=1} = 0 \Rightarrow \left(-2\hat{\mathrm{s}}^{\mathsf{T}}\mathrm{g} + 2\alpha\hat{\mathrm{s}}^{\mathsf{T}}\hat{\mathrm{s}}\right)\Big|_{\alpha=1} = 0 \Rightarrow \hat{\mathrm{s}}^{\mathsf{T}}(\hat{\mathrm{s}} - \mathrm{g}) = 0.$$

Comments

We now turn to the constructive part of the proof, where we must show that whenever the vector g does not belong to the cone, we can explicitly build a vector d with the desired separating properties. The strategy is geometric: we look for the point in the cone that is closest to g in terms of Euclidean distance. This point, denoted by $\hat{\mathrm{s}}$, always exists because the cone is a closed set, meaning that it contains its boundary and has no "holes."

By definition, $\hat{\mathrm{s}}$ solves an optimization problem: minimize the squared distance between s and g subject to s being inside the cone. Since the cone is a cone in the mathematical sense, any positive multiple of $\hat{\mathrm{s}}$ also belongs to the cone. Now, calculus provides an important condition. If we vary the scalar multiplier α and minimize the distance between $\alpha\hat{\mathrm{s}}$ and g, the minimizing value occurs at $\alpha = 1$. Differentiating the squared norm with respect to α and evaluating at one gives a simple equation: the inner product of $\hat{\mathrm{s}}$ with the difference between $\hat{\mathrm{s}}$ and g equals zero.

This orthogonality condition is crucial. It says that the line connecting g and $\hat{\mathrm{s}}$ is perpendicular to $\hat{\mathrm{s}}$ itself. Intuitively, this reflects the fact that $\hat{\mathrm{s}}$ is the closest point in the cone to g. In geometry, whenever we project a point onto a convex set, the vector connecting the point to its projection is orthogonal to the set at that projection point. This observation will serve as the foundation for constructing the separating vector d.

Now, let s be any other vector in K. Since K is convex, we have by the minimizing property of $\hat{s}$ that

$$\|\hat{s} + \theta(s - \hat{s}) - g\|_2^2 \geq \|\hat{s} - g\|_2^2 \quad \text{for all } \theta \in [0, 1],$$

and hence

$$2\theta(s - \hat{s})^{\mathsf{T}}(\hat{s} - g) + \theta^2\|s - \hat{s}\|_2^2 \geq 0.$$

By dividing this expression by θ and taking the limit as $\theta \downarrow 0$, we have

$$(s - \hat{s})^{\mathsf{T}}(\hat{s} - g) \geq 0.$$

Therefore, because of $\hat{s}^{\mathsf{T}}(\hat{s} - g) = 0$ we have

$$s^{\mathsf{T}}(\hat{s} - g) \geq 0, \quad \text{for all } s \in K.$$

We claim now that the vector

$$d = \hat{s} - g$$

satisfies the conditions (i):

$$g^{\mathsf{T}}d < 0, \quad B^{\mathsf{T}}d \geq 0, \quad C^{\mathsf{T}}d = 0.$$

Comments

Having established the orthogonality relation, we continue by exploring how this property extends to all other points in the cone. Take any vector s that lies in the cone. Because the cone is convex, we can form convex combinations between $\hat{s}$ and any other s. In other words, the point $\hat{s} + \theta(s - \hat{s})$ remains in the cone for all θ between zero and one. By the minimizing property of $\hat{s}$, the distance from this new point to g cannot be smaller than the distance from $\hat{s}$ to g.

Expanding this inequality and performing a simple limit argument shows that the inner product of the difference $s - \hat{s}$ with the vector $\hat{s} - g$ is nonnegative. Combining this with the orthogonality condition, we arrive at a general inequality: the inner product of any s in the cone with the vector $\hat{s} - g$ is nonnegative. This is an important result, because it suggests a natural candidate for the separating vector. We can now define d as $\hat{s} - g$.

By construction, d is nonzero whenever g does not belong to the cone. Furthermore, this inequality shows that d forms a nonnegative inner product with every element of the cone. Thus, d satisfies one part of the alternative conditions we are seeking. The remaining step is to check directly that d also satisfies the other two requirements involving g and C, namely that the transpose of g times d is negative and the transpose of C times d equals zero. These verifications will confirm that we have indeed constructed the separating hyperplane.

Proof of Farkas' Lemma (conclusion)

Note that $d \neq 0$ because $g \notin K$. From the equality $\hat{s}^T(\hat{s} - g) = 0$ we have

$$d^T g = d^T(\hat{s} - d) = (\hat{s} - g)^T \hat{s} - d^T d = -\|d\|_2^2 < 0,$$

so that d satisfies the first property in

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0.$$

Since $d^T s \geq 0$ for all $s \in K$, we have

$$d^T(By + Cw) \geq 0 \quad \text{for all } y \geq 0,\ w.$$

- Fix $y = 0$: then $(C^T d)^T w \geq 0$ for all w implies $C^T d = 0$.
- Fix $w = 0$: then $(B^T d)^T y \geq 0$ for all $y \geq 0$ implies $B^T d \geq 0$.

Thus, d satisfies all properties in the alternative and the proof is complete. □

F-O Necessary condition

KKT

Tangent Cone & Feasible Directions

Necessary condition

Second-Order Necessary condition

Sufficient Conditions

Comments

To complete the argument, let us carefully verify the properties of the vector d. Recall that d is defined as $\hat{s} - g$, and g is not in the cone, so d cannot be the zero vector. We can easily deduce from the orthogonality condition that the inner product of d with g is strictly negative.

More precisely, substituting back, we obtain that this inner product equals the negative of the squared norm of d, which is strictly less than zero. Thus, the first required property is satisfied. Next, because d has a nonnegative inner product with every element of the cone, we can test this condition on particular elements. If we take s to be equal to Cw, where $y = 0$, we conclude that the transpose of C times d must vanish; otherwise, the inequality would not hold for all possible w.

Similarly, if we take s equal to By with $w = 0$, the condition reduces to the requirement that the transpose of B times d is nonnegative for all $y \geq 0$. This is only possible if $B^T d$ itself is a nonnegative vector. Therefore, d satisfies all three properties simultaneously: the inner product with g is negative, $B^T d$ is nonnegative, and $C^T d$ equals zero. Together, these facts complete the proof of Farkas' Lemma.

The lemma guarantees that for any given vector, either it is in the cone or there is a separating hyperplane defined by d. This result is elegant because it transforms a membership problem into a dual characterization involving linear inequalities, forming a cornerstone of duality theory in optimization.

Closedness of the Cone K

Let's now return to our claim that the set

$$\mathrm{K} = \{\mathrm{B}y + \mathrm{C}w \mid y \geq 0\},$$

(where B and C are matrices of dimensions $n \times m$ and $n \times p$, respectively, and y, w are vectors of appropriate dimensions) is a closed set. This fact was used in the proof of Farkas' Lemma.

Lemma 7

The set K is closed.

Proof: By splitting w into positive and negative parts, we write

$$\mathrm{K} = \left\{ \begin{bmatrix} \mathrm{B} & \mathrm{C} & -\mathrm{C} \end{bmatrix} \begin{bmatrix} y \\ w^+ \\ w^- \end{bmatrix} \;\middle|\; \begin{bmatrix} y \\ w^+ \\ w^- \end{bmatrix} \geq 0 \right\}.$$

Hence, we may assume w.l.o.g. that $\mathrm{K} = \{\mathrm{B}y \mid y \geq 0\}$. Suppose $\mathrm{B} \in \mathbb{R}^{n \times m}$.

Comments

When studying convex cones in linear algebra, one important property is whether such a set is closed. Recall that the cone we are working with is defined as all vectors of the form $\mathrm{B}y + \mathrm{C}w$, where y is constrained to be nonnegative. This structure is essential, because it appears directly in the reasoning behind separation results. To check closedness, we begin by reformulating the set so that all variables are explicitly nonnegative.

This is done by decomposing the vector w into its positive and negative components, which ensures that the cone is expressed only through nonnegative combinations of columns. After this transformation, the cone can be written simply as all vectors $\mathrm{B}y$ with $y \geq 0$. This reduction is helpful because it removes unnecessary complications and allows us to focus on the basic mechanism: a cone generated by a finite set of vectors.

The question now is: if we take a convergent sequence of points inside this cone, does the limit point also belong to the cone? Intuitively, this means we want to know whether the cone is robust under limits, or whether points can "slip out" as we pass to the boundary. Proving that the cone is closed is important, because when proving Farkas' lemma we rely on the fact that limits of feasible points remain feasible. Thus, our immediate goal is to build the argument step by step, first establishing a minimal representation of elements in the cone, and then using that to demonstrate closedness rigorously.

Step 1: Minimal Representation in Cone K

- First, we show that for any $s \in K$, we can write $s = B_I y_I$ with $y_I > 0$, where $I \subset \{1, \dots, m\}$, B_I is the submatrix of B with columns indexed by I having full column rank, and I has minimal cardinality.
- Assume for contradiction that $\mathcal{N} \subset \{1, \dots, m\}$ is a minimal index set such that $s = B_{\mathcal{N}} y_{\mathcal{N}}$, $y_{\mathcal{N}} > 0$ (since $\mathcal{N}$ is minimal, all components of $y_{\mathcal{N}}$ are strictly positive.), but the columns of $B_{\mathcal{N}}$ are linearly dependent.
- Then, there exists a nonzero vector t with $B_{\mathcal{N}} t = 0$. For any scalar τ, we have:

$$s = B_{\mathcal{N}}(y_{\mathcal{N}} + \tau t).$$

- We adjust τ so that some components of $y_{\mathcal{N}} + \tau t$ become zero, while others remain positive. Let $\hat{\mathcal{N}}$ be the indices with positive components, and let $\bar{y}_{\hat{\mathcal{N}}}$ be the corresponding vector.
- Then $s = B_{\hat{\mathcal{N}}} \bar{y}_{\hat{\mathcal{N}}}$ with $\bar{y}_{\hat{\mathcal{N}}} > 0$, contradicting the minimality of $\mathcal{N}$.

Comments

The first step toward proving closedness is to understand how an arbitrary element of the cone can be represented. Take any vector s belonging to the cone. By construction, it can be written as B times some nonnegative vector y. But among all possible such representations, we want one that uses the fewest columns of B. That means choosing an index set I with minimal cardinality such that $s = B_I y_I$, with $y_I > 0$, and with the chosen columns linearly independent.

Why is this reduction important? Because if the chosen columns were linearly dependent, then we could adjust the coefficients and remove one of them, producing a more economical representation. The key contradiction argument works like this: assume we have already chosen a minimal set of indices, yet the corresponding columns are linearly dependent. Then there exists a nonzero vector t such that multiplying $B_{\mathcal{N}}$ by t gives zero. If we add any scalar multiple of t to our coefficient vector, the product with B remains unchanged. This allows us to adjust coefficients so that at least one component becomes zero while the others stay positive.

In this way, we construct a new representation of s using fewer columns, contradicting the assumption that the index set was minimal. Therefore, in a true minimal representation, the selected columns must necessarily be linearly independent. This structural property guarantees that every element of the cone can be written in a way that avoids redundancy. Later, this will play a central role in handling sequences of points, since linear independence allows us to manipulate limits more effectively.

Step 2: Closedness of Cone K

- Let $\{s^j\}$ be a sequence with $s^j \in K$ for all j and $s^j \to s$. We aim to prove that $s \in K$.
- By the previous claim, for each j we can write $s^j = B_{I_j} y^j_{I_j}$ with $y^j_{I_j} > 0$, I_j minimal, and B_{I_j} has linearly independent columns.
- Since there are only finitely many possible choices of index set I_k, at least one index set occurs infinitely often in the sequence. By choosing such an index set I, we can take a subsequence if necessary and assume without loss of generality that $I_k = I$ for all k.
- Then $s^j = B_I y^j_I$ with $y^j_I > 0$ and B_I of full column rank. So $B_I^T B_I$ is invertible, and
$$y^j_I = (B_I^T B_I)^{-1} B_I^T s^j.$$
- Taking the limit as $j \to \infty$ and using $s^j \to s$, we obtain
$$y^j_I \to y_I := (B_I^T B_I)^{-1} B_I^T s \geq 0.$$
- Thus, $s = B_I y_I$ with $y_I \geq 0$, so $s \in K$.

□

Comments

Now we are ready to establish closedness formally. Consider a sequence of points inside the cone that converges to some vector s. The question is whether s also lies in the cone. Thanks to the minimal representation result, each point in the sequence can be expressed using a linearly independent set of columns. Since the total number of columns in B is finite, there are only finitely many possible index sets.

By the pigeonhole principle, at least one index set must appear infinitely often. That means we can extract a subsequence where every element uses the same set of linearly independent columns. For this fixed index set I, we then write each element of the subsequence as B_I times a nonnegative vector y_I. Because the columns of B_I are independent, the matrix transpose of B_I times B_I is invertible. This allows us to solve uniquely for y_I in terms of the point s. Now, since our subsequence converges to s, the corresponding sequence of y vectors also converges, and importantly, the limit remains nonnegative.

This is crucial: nonnegativity is preserved under limits when the coefficients are obtained through a continuous linear transformation. Thus, the limit vector s can still be written as B_I times a nonnegative vector, meaning it belongs to the cone. This proves that the cone is closed. Conceptually, what we have shown is that the cone is stable under convergence: if you approach a boundary point by feasible points, that boundary point is still feasible. This stability under limits ensures the reliability of the cone in optimization and separation theorems.

By applying Farkas' Lemma to the cone

$$N = \left\{ \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) \,\middle|\, \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I} \right\},$$

with $g = \nabla f(x^*)$, we get the alternative:

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I},$$

or there exists d with $d^T \nabla f(x^*) < 0$ and $d \in \mathcal{F}(x^*)$.

Proof of Theorem 21 (First-Order Necessary Conditions):
Combining Lemmas 5 and 6, we derive the KKT conditions. Suppose $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds. Then if x^* is a local solution for $(*)$, there exists $\lambda^* \in \mathbb{R}^m$ such that the KKT conditions are satisfied.

Comments

We are now ready to begin the proof of the theorem 21, which postulates the necessary conditions for a point to be a local minimum. The crucial tool that allows us to proceed is Farkas' Lemma. Recall that this lemma provides a sharp alternative: either a given vector belongs to a certain cone, or there exists a separating direction that violates the nonnegativity property. In our context, the cone is formed by linear combinations of the gradients of the active constraints at the candidate point, with nonnegative coefficients for the inequality constraints.

The vector of interest is the gradient of the objective function at x^*. If this gradient belongs to the cone, then we can represent it as a weighted combination of the active gradients, exactly the structure required for the KKT system. On the other hand, if the gradient does not belong to this cone, then Farkas' Lemma guarantees the existence of a feasible direction d such that the inner product of d with the gradient is strictly negative. But such a direction would allow us to decrease the objective function while remaining feasible, contradicting the assumption that x^* is a local minimizer.

Therefore, the second alternative is impossible, and the gradient must lie in the cone. This establishes the foundation of the proof: we have shown that if x^* is indeed a local solution, then the gradient of the objective can be expressed as a combination of active gradients with nonnegative coefficients. This is the central step that connects the geometry of feasible directions with the algebraic representation required for the multipliers.

Existence of Multipliers and Construction of λ^*

We first show that there exist multipliers λ_i, $i \in \mathcal{A}(x^*)$, satisfying

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I}.$$

From Theorem 22, $d^T \nabla f(x^*) \geq 0$ for all $d \in \mathcal{T}_\Omega(x^*)$. Since LICQ holds, Lemma 5 gives $\mathcal{T}_\Omega(x^*) = \mathcal{F}(x^*)$, so we have

$$d^T \nabla f(x^*) \geq 0 \ \forall d \in \mathcal{F}(x^*).$$

Then by Farkas' Lemma, the desired multipliers exist.
We now define:

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathcal{A}(x^*), \\ 0, & i \in \mathcal{I} \setminus \mathcal{A}(x^*), \end{cases}$$

and show that λ^* and x^* satisfy the KKT conditions.

We check these conditions in turn.

Comments

Having established that the gradient lies within the cone of active constraints, the next step is to demonstrate the existence and explicit construction of the multipliers. We start from the fact that for any feasible direction d, the inner product of d with the gradient of the objective at x^* is greater than or equal to zero. This follows from Theorem 22: a local minimizer cannot admit a feasible descent direction. Under the Linear Independence Constraint Qualification, the tangent cone coincides with the feasible direction cone. Therefore, the condition of nonnegativity holds for all feasible directions.

By applying Farkas' Lemma once again, we deduce that the gradient can indeed be represented as a linear combination of the active gradients with nonnegative coefficients. This provides us with the set of multipliers associated with active constraints. To extend this into a complete multiplier vector, we simply assign zero to the components corresponding to inactive inequality constraints. Thus, we obtain a full multiplier vector, which we denote λ^*.

This construction is essential, because it ensures that every constraint is accounted for consistently: active constraints receive the multipliers provided by the cone representation, while inactive constraints are given zero weight. At this point, we have built the mathematical object that should satisfy the KKT conditions, and our task reduces to verifying that each condition in the theorem holds for the pair consisting of x^* and λ^*.

- Condition (7a) ($\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$) follows immediately from

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I},$$

and from the definition of the Lagrangian function,

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x),$$

as well as the previous definition of the multipliers, $\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathcal{A}(x^*), \\ 0, & i \in \mathcal{I} \setminus \mathcal{A}(x^*). \end{cases}$

- Since x^* is feasible, the conditions (7b) and (7c) are satisfied:

$$c_i(x^*) = 0 \quad \text{for all } i \in \mathcal{E}, \qquad c_i(x^*) \geq 0 \quad \text{for all } i \in \mathcal{I}.$$

- From the active combination condition, $\lambda_i^* \geq 0$ for $i \in \mathcal{A}(x^*) \cap \mathcal{I}$, and from the definition of λ^*, $\lambda_i^* = 0$ for $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$. Hence, $\lambda_i^* \geq 0$ for all $i \in \mathcal{I}$, so that (7d) holds.
- For $i \in \mathcal{A}(x^*) \cap \mathcal{I}$, we have $c_i(x^*) = 0$, while for $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$, we have $\lambda_i^* = 0$. Therefore, $\lambda_i^* c_i(x^*) = 0 \quad$ for all $i \in \mathcal{I}$. so that (7e) is satisfied as well.

And the proof is complete. □

Comments

The final step of the proof is to check each of the KKT conditions systematically. We begin with the stationarity condition, which requires that the gradient of the Lagrangian function at x^* equals the transpose of the Jacobian of the constraints at x^* multiplied by λ^*. This holds directly by construction, since the multipliers were chosen to represent the gradient in this precise form. Next, the feasibility conditions must be verified.

Because x^* is assumed to be a feasible point, all equality constraints vanish and all inequality constraints are satisfied with nonnegative values. The nonnegativity condition on the multipliers follows immediately: multipliers for active inequalities are nonnegative, and those for inactive inequalities were set to zero. Finally, complementary slackness must be checked.

For every inequality constraint, either it is active, in which case its multiplier may be positive but the constraint itself is zero, or it is inactive, in which case the multiplier is zero while the constraint is strictly positive. In both cases, the product of the multiplier with the constraint value is zero, as required. Having confirmed stationarity, feasibility, nonnegativity, and complementary slackness, all conditions of the theorem are satisfied.

This completes the proof of the KKT theorem: any local solution under the Linear Independence Constraint Qualification admits a corresponding multiplier vector such that the KKT system holds. Thus, the argument is closed, and the result stands as a cornerstone of constrained optimization.

Second-Order Optimality Conditions: Motivation

First-order (KKT) conditions describe how ∇f and gradients of active c_i behave at a solution x^*.

- When these conditions hold, a move along any $w \in \mathcal{F}(x^*)$ either:
  - increases f to first order: $w^T \nabla f(x^*) > 0$, or
  - leaves f unchanged to first order: $w^T \nabla f(x^*) = 0$.
- Second-order conditions analyze such "undecided" directions.

Key Idea: In directions where $w^T \nabla f(x^*) = 0$, the second-order terms in the Taylor expansions of f and c_i are examined to determine whether f increases or decreases. This analysis concerns the curvature of the Lagrangian in those directions.

Assumption: Further, until it is said otherwise, all f and c_i, $i \in \mathcal{E} \cup \mathcal{I}$, are assumed to be twice continuously differentiable.

Comments

Up to now, we have mainly discussed first-order conditions, also known as the Karush–Kuhn–Tucker conditions. These describe how the gradient of the objective function relates to the gradients of the active constraints at a candidate solution. They allow us to understand whether moving in a certain feasible direction will tend to increase the objective function or leave it unchanged, at least to first order.

However, there are directions where first-order information is silent: directions in which the inner product between the feasible direction and the gradient of the objective function is zero. Along such directions, the first-order expansion provides no clear guidance. This is precisely where second-order conditions become essential. By examining the second derivatives of both the objective and the constraints, we are able to capture the curvature properties of the problem.

In other words, second-order terms act as a kind of "tiebreaker," deciding whether a feasible move will ultimately increase or decrease the objective. The curvature is analyzed through the Hessian of the Lagrangian function, which integrates the objective and constraints into a single mathematical object. Since curvature considerations are more delicate, we impose stronger smoothness assumptions, requiring both the objective and the constraints to be twice continuously differentiable.

This guarantees that the Taylor expansions we rely on behave properly. Altogether, the step from first- to second-order analysis represents a refinement: it moves from directional slopes to local surface shapes, providing deeper insight into the structure of optimal solutions.

Given $\mathcal{F}(x^*)$ (the set of linearized feasible directions) and a Lagrange multiplier vector λ^* satisfying the KKT conditions (7), we define the *critical cone* $\mathcal{C}(x^*, \lambda^*)$ as:

$$\mathcal{C}(x^*, \lambda^*) = \left\{ w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T w = 0, \text{ for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \right\}$$

Equivalently:

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \text{for all } i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w = 0, & \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0, \\ \nabla c_i(x^*)^T w \geq 0, & \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0. \end{cases} \tag{8}$$

Key property: From the definition (8) and the fact that $\lambda_i^* = 0$ for all inactive constraints $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$, it follows that

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.$$

From the first KKT condition (7a) and the definition of the Lagrangian function, we have that

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow w^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0.$$

Hence the critical cone $\mathcal{C}(x^*, \lambda^*)$ contains directions from $\mathcal{F}(x^*)$ for which it is not clear from first derivative information alone whether f will increase or decrease.

Comments

To deepen our understanding of second-order conditions, we need a precise way of describing the feasible directions that are truly “critical.” These are not just any directions that satisfy the linearized constraints; instead, they are directions for which the first-order information fails to distinguish whether the objective will improve or worsen. The mathematical object that captures this subtle set of directions is called the critical cone.

Formally, the critical cone consists of feasible directions that are tangent to all equality constraints and to those active inequality constraints with strictly positive Lagrange multipliers. For constraints with zero multipliers, the feasible directions may allow movement that does not immediately violate feasibility, so a weak inequality suffices. This careful definition ensures that the cone isolates exactly the cases where the first-order expansion leaves us undecided.

An important property follows immediately: for any direction in the critical cone, the linear term of the Lagrangian expansion vanishes, meaning that the directional derivative of the objective is zero. Consequently, if we want to decide about local optimality, we must look at second-order terms in precisely these directions. The introduction of the critical cone thus provides the right framework: it trims away irrelevant directions, leaving us with the delicate set where higher-order analysis is truly necessary.

Example 7

Consider the problem:

$$\min x_1 \quad \text{subject to } x_2 \geq 0, \quad 1 - (x_1 - 1)^2 - x_2^2 \geq 0.$$

The solution is $x^* = (0,0)^\mathsf{T}$ with active set $\mathcal{A}(x^*) = \{1,2\}$ and a unique optimal Lagrange multiplier:

$$\lambda^* = (0, 0.5)^\mathsf{T}.$$

Since the gradients of active constraints at x^* are $(0,1)^\mathsf{T}$ and $(2,0)^\mathsf{T}$, the LICQ holds, ensuring uniqueness of λ^*.

Linearized feasible set:

$$\mathcal{F}(x^*) = \{d \mid d \geq 0\}.$$

Critical cone:

$$\mathcal{C}(x^*, \lambda^*) = \left\{(0, w_2)^\mathsf{T} \mid w_2 \geq 0\right\}.$$

Key observation: The only direction in the critical cone is vertical and nonnegative, due to the fact that only the second constraint has a positive multiplier.

Comments

Let us now turn to an example that illustrates the definition of the critical cone. Consider an optimization problem with two inequality constraints: one requires a variable to be nonnegative, while the other describes the interior of a circle. At the solution, both constraints are active. Computing the gradients of the constraints shows that the linear independence constraint qualification is satisfied, ensuring a unique multiplier vector.

Interestingly, only one of the active constraints has a strictly positive multiplier. This detail is decisive, because it dictates which directions belong to the critical cone. Specifically, the linearized feasible set allows only nonnegative movements in both coordinates. But once we impose the definition of the critical cone, we see that only vertical, nonnegative directions remain. In other words, the problem structure and the positivity of the multiplier eliminate all other feasible tangents from critical consideration.

This example demonstrates how the cone is typically much smaller than the entire feasible set: it isolates the directions along which first-order conditions provide no information. Through this concrete case, we see the importance of multipliers in shaping the geometry of the critical cone. The example also hints at the deeper role of second-order analysis: once we know which directions are critical, we can check whether the curvature of the Lagrangian guarantees nonnegative behavior there, a condition that is central to second-order optimality.

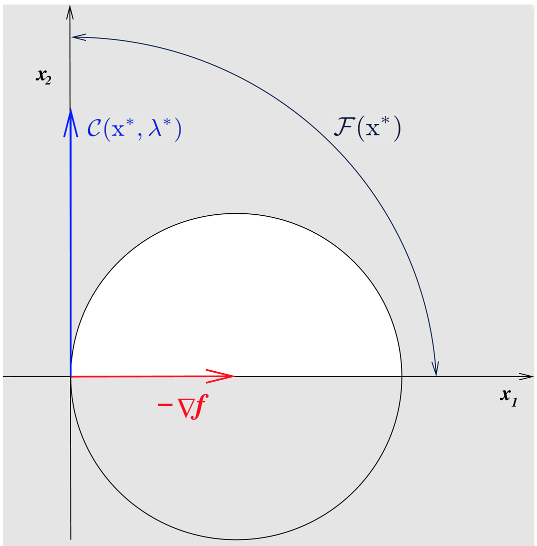

Figure: Problem from Example 7, showing $\mathcal{F}(x^*)$ and $\mathcal{C}(x^*, \lambda^*)$.

Comments

A geometric picture makes these concepts far clearer. In the example just discussed, the feasible set is the intersection of a half-space and a disk. At the solution point, located on the boundary of both sets, we can visualize the directions allowed by the linearized constraints. These directions form a cone in the first quadrant. Within this cone, however, the critical cone is only a narrow slice: the vertical ray extending upward. The visualization thus makes concrete the distinction between feasible directions and critical directions.

The feasible cone tells us where we can move infinitesimally without immediately violating constraints, but the critical cone further filters this set to the directions where the first-order change in the objective vanishes. The picture shows this filtering vividly: the wide region of feasible movements collapses to a single line of critical interest.

Such illustrations are extremely helpful for building intuition about the geometry of optimality conditions. They emphasize that the interplay between constraints and multipliers is not abstract but manifests directly in the shape of allowable movement near the solution. This geometric perspective provides a bridge between algebraic definitions and practical understanding, preparing us for the more general second-order theory.

Theorem 23: Second-Order Necessary Conditions

Suppose x^* is a local solution of the problem

$$\min_{x\in\Omega} f(x),\ \ \Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\} \qquad (*)$$

and that LICQ holds. Let λ^* be a Lagrange multiplier satisfying the KKT conditions (7). Then:

$$w^T \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w \geq 0 \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*).$$

Key idea: At a local solution x^*, the Hessian of the Lagrangian must have nonnegative curvature along all critical directions (that is, the directions in $\mathcal{C}(x^*, \lambda^*)$).

Proof: Note that for a local solution x^*, any feasible sequence $\{z_k\} \to x^*$ must satisfy $f(z_k) \geq f(x^*)$ for all k sufficiently large.
Since $w \in \mathcal{C}(x^*, \lambda^*) \subset \mathcal{F}(x^*)$, we can use the technique in the proof of Lemma 5, to choose a feasible sequence $\{z_k\} \to x^*$ and positive scalars $\{t_k\}$ such that

$$\lim_{k\to\infty} \frac{z_k - x^*}{t_k} = w, \quad \text{or equivalently,} \quad z_k - x^* = t_k w + o(t_k).$$

Comments

We now turn to the second-order necessary condition. Recall that we already have a candidate point, which we call x^*, and a multiplier vector, which we call λ^*, that satisfy the Karush–Kuhn–Tucker conditions. Our goal is to show that at a local minimizer the Lagrangian must have nonnegative curvature along every critical direction. Concretely, for any direction w in the critical cone, the quadratic form given by $w^T\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w$ must be greater than or equal to zero.

The proof is based on building feasible approximations along w: starting from a critical direction w, we build a feasible sequence of points that approach x^* with steps proportional to w. To do this we use the tangent/feasible construction: choose positive scalars t_k that go to zero and define points $z_k = x^* + t_k w$ plus higher-order terms.

By the definition of critical directions, these z_k can be chosen feasible, so the objective values at z_k cannot be less than $f(x^*)$ for large k. This feasibility and limiting construction are the heart of the argument: they let us compare the objective evaluated at z_k with the Taylor expansion of the Lagrangian around x^*. The rest of the proof quantifies the second-order term in that expansion and shows it cannot be negative without contradicting local optimality. In short, the plan is: 1) build feasible approximations along w, 2) expand the Lagrangian to second order, and 3) use optimality to force the quadratic form to be nonnegative.

Second-Order Necessary Conditions (cont.)

Because of the construction of the feasible sequence $\{z_k\}$, we have:

$$c_i(z_k) = t_k \nabla c_i(x^*)^\top w, \quad \text{for all } i \in \mathcal{A}(x^*).$$

From this equality and the definition of critical cone (see (8)), it follows that:

$$\mathcal{L}(z_k, \lambda^*) = f(z_k) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* c_i(z_k) =$$
$$= f(z_k) - t_k \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*)^\top w = f(z_k).$$

On the other hand, from Taylor's theorem (Theorem 1) and continuity of the Hessians $\nabla^2 f$ and $\nabla^2 c_i, i \in \mathcal{E} \cup \mathcal{I}$, we obtain the expansion:

$$\mathcal{L}(z_k, \lambda^*) = \mathcal{L}(x^*, \lambda^*) + (z_k - x^*)^\top \nabla_x \mathcal{L}(x^*, \lambda^*) =$$
$$+ \frac{1}{2}(z_k - x^*)^\top \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*)(z_k - x^*) + o(\|z_k - x^*\|^2).$$

Comments

Continuing the argument, we exploit two linked observations. First, because the constructed sequence z_k is feasible and because of the linearization of constraints, as we showed in the proof of Lemma 5 the value of each active constraint at z_k equals t_k times the inner product of the corresponding active gradient with w, up to higher orders. Second, by the definition of the critical cone and complementarity, the sum over active indices of λ^* times those linearized constraint values vanishes.

Combining these facts yields an important identity: the Lagrangian evaluated at z_k with multipliers λ^* equals simply $f(z_k)$, because the constraint terms cancel at the relevant order. This equality is extremely useful because it lets us work with the Lagrangian expansion instead of the raw objective. We next apply Taylor's theorem to the Lagrangian about x^*: the Lagrangian at z_k equals the Lagrangian at x^* plus the linear term, which vanishes by stationarity, plus one-half times $(z_k - x^*)^\top \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*)(z_k - x^*)$, plus little-o of the squared norm.

By the complementarity condition ($\lambda_i^* c_i(x^*) = 0$, for all $i \in \mathcal{E} \cup \mathcal{I}$), we have:

$$\mathcal{L}(x^*, \lambda^*) = f(x^*).$$

From the stationarity condition, the linear term in the Taylor expansion vanishes:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0.$$

Using the expansion $z_k - x^* = t_k w + o(t_k)$, we obtain:

$$\mathcal{L}(z_k, \lambda^*) = f(x^*) + \frac{1}{2} t_k^2 w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w + o(t_k^2).$$

Since $\mathcal{L}(z_k, \lambda^*) = f(z_k)$, this yields:

$$f(z_k) = f(x^*) + \frac{1}{2} t_k^2 w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w + o(t_k^2).$$

If $w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w < 0$, then $f(z_k) < f(x^*)$ for all k sufficiently large, contradicting local optimality of x^*. Hence, the condition

$$w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w \geq 0$$

must hold for all $w \in \mathcal{C}(x^*, \lambda^*)$. □

Comments

Replacing $z_k - x^*$ by $t_k w+$ smaller terms produces a second-order term proportional to t_k^2 times the quadratic form $w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w$, plus higher-order remainders. Because Lagrangian at z_k equals $f(z_k)$, we now have a precise asymptotic expansion of $f(z_k)$ in powers of t_k, with the coefficient of t^2 given by one-half of that quadratic form.

We are now in a position to finish the necessary condition. Because x^* is a local minimizer, $f(z_k)$ is at least $f(x^*)$ for all sufficiently large k. Substituting the Taylor expansion derived previously, we obtain that $f(x^*) + \frac{1}{2} t_k^2 w^\mathsf{T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w+$ smaller order terms is greater than or equal to $f(x^*)$ for small t_k. Dividing by t_k^2 and letting $k \to \infty$ forces the quadratic form to be nonnegative. If instead that quadratic form were strictly negative, then for sufficiently small t_k the objective at z_k would be strictly less than $f(x^*)$, contradicting local optimality.

Hence the quadratic form must be greater than or equal to zero for every critical direction w. Conceptually, this result says that along directions where the first derivative gives no information, the second derivative of the Lagrangian must curve upwards or stay flat — it cannot bend downward. This requirement is a natural higher-order analogue of the stationarity condition: stationarity kills linear terms, and the second-order necessary condition controls the sign of the remaining quadratic terms. We will later contrast this necessary statement with a sufficient condition that strengthens “greater than or equal to” to “strictly greater than” and, in return, guarantees strict local optimality.

Key idea: A feasible point satisfying KKT conditions is a strict local solution if the Lagrangian is strictly convex along all nonzero critical directions.

Theorem 24: Second-Order Sufficient Conditions

Suppose that for some feasible point $x^* \in \mathbb{R}^n$, there exists a multiplier λ^* such that the KKT conditions are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0 \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*),\ w \neq 0.$$

Then x^* is a strict local solution of problem $(*)$:

$$\min_{x \in \Omega} f(x),\ \Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\} \qquad (*)$$

- Sufficient conditions guarantee that x^* is a local solution to $(*)$.
- Unlike necessary conditions, they do not require a constraint qualification.
- The inequality is strict: ">" instead of "≥".

Comments

We now state and interpret a second-order sufficient condition. Suppose we have a feasible point x^* and a multiplier vector λ^* satisfying the KKT conditions. If the Hessian of the Lagrangian at x^*, λ^* is strictly positive along every nonzero vector in the critical cone — that is, for every nonzero w in the critical cone the quadratic form $w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w$ is strictly greater than zero — then x^* is a strict local minimizer.

Intuitively, this condition says that not only are the linear terms neutralized by stationarity, but the curvature in all previously undecided directions is strictly upward. Any small feasible perturbation then increases the Lagrangian — and, because the constraint linearization and complementarity arguments carry through, increases the objective.

Two useful comments follow. First, unlike the necessary condition, the sufficient condition does not require a constraint qualification: we only need a KKT pair to exist. Second, the inequality is strict: "greater than" rather than "greater than or equal to". That strictness is what converts nondecrease into a genuine local increase, producing a neighborhood where x^* yields a strictly smaller objective than any other feasible point.

In practice, checking the sufficient condition provides a certificate of strict local optimality and is widely used in nonlinear programming algorithms to verify convergence to a true minimizer.

Proof: Let $\tilde{\mathcal{C}} = \{d \in \mathcal{C}(x^*, \lambda^*) \mid \|d\| = 1\}$.
This set is compact, and by the condition

$$d^{\mathsf{T}} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d > 0 \quad \forall d \in \tilde{\mathcal{C}},$$

the minimizer of $d^{\mathsf{T}} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d$ over $\tilde{\mathcal{C}}$ is a strictly positive number, say σ. Since $\mathcal{C}(x^*, \lambda^*)$ is a cone, we have that $(w/\|w\|) \in \tilde{\mathcal{C}}$ if and only if $w \in \mathcal{C}(x^*, \lambda^*)$, $w \neq 0$. Therefore:

$$w^{\mathsf{T}} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq \sigma \|w\|^2 \quad \forall w \in \mathcal{C}(x^*, \lambda^*).$$

Now suppose there exists a feasible sequence $\{z_k\} \to x^*$ such that

$$f(z_k) < f(x^*) + \frac{\sigma}{2} \|z_k - x^*\|^2 \quad \text{for all } k \text{ sufficiently large.}$$

Let's show that this assumption leads to a contradiction.

Comments

To establish the second-order sufficient condition, we begin by focusing on the geometry of the critical cone. Consider the normalized cone, denoted $\tilde{\mathcal{C}}$, which consists of all unit-length directions that lie inside the critical cone. Because this set is compact, and because the quadratic form defined by $d^{\mathsf{T}} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d$ is strictly positive for all such directions, we can conclude that this quadratic form has a minimum value on the set.

Importantly, this minimum value, which we call σ, is strictly greater than zero. The interpretation of this fact is straightforward: the Lagrangian does not merely have positive curvature in critical directions, but in fact has uniformly positive curvature bounded away from zero. Extending from the normalized cone to the full cone, we obtain the inequality that for every direction w in the critical cone, the quadratic form $w^{\mathsf{T}} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq \sigma \|w\|^2$.

This inequality is crucial because it provides a uniform lower bound on curvature in all feasible critical directions. The proof then proceeds by contradiction. Suppose there exists a feasible sequence of points approaching x^* whose function values fall strictly below $f(x^*) + \frac{\sigma}{2} \|z_k - x^*\|^2$. If such a sequence existed, it would mean that the objective could decrease faster than the quadratic lower bound established by the Hessian. Our task is to show that such a situation is impossible, because it would contradict the uniform curvature guaranteed by the second-order sufficient condition.

By taking a subsequence, we get a limit direction:

$$\lim_{k\to\infty} \frac{z_k - x^*}{\|z_k - x^*\|} = d.$$

Then $d \in \mathcal{F}(x^*)$ by properties of feasible directions (Lemma 5).
From KKT conditions and feasibility:

$$\mathcal{L}(z_k, \lambda^*) = f(z_k) - \sum_{i\in\mathcal{A}(x^*)} \lambda_i^* c_i(z_k) \le f(z_k).$$

Taylor expansion of $\mathcal{L}(z_k, \lambda^*)$ from Theorem 23 still holds.
If $d \notin \mathcal{C}(x^*, \lambda^*)$, then there exists some $j \in \mathcal{A}(x^*) \cap \mathcal{I}$ such that

$$\lambda_j^* \nabla c_j(x^*)^T d > 0,$$

and for the remaining indices $i \in \mathcal{A}(x^*)$, we have

$$\lambda_i^* \nabla c_i(x^*)^T d \ge 0.$$

Comments

Let us assume that such a contradictory sequence does exist and examine its limiting behavior. Any sequence approaching x^* can be normalized by dividing by its distance from x^*, yielding a limiting direction, which we call d. By construction, d belongs to the set of feasible directions.

The Karush–Kuhn–Tucker conditions now enter the argument. Since the sequence is feasible, we can express the Lagrangian at each point z_k as $f(z_k) - \sum_{i\in\mathcal{A}(x^*)} \lambda_i^* c_i(z_k)$. Because the multipliers are nonnegative and constraints are satisfied, this Lagrangian value is always less than or equal to the raw objective. At the same time, the Taylor expansion of the Lagrangian around x^* still applies.

Now we distinguish two cases: if the limiting direction d is not in the critical cone, then there must exist an active inequality constraint with positive multiplier such that the gradient of that constraint at x^*, dotted with d, is strictly positive. Meanwhile, for the other active constraints, the corresponding dot products are nonnegative. This separation tells us that the candidate direction d cannot escape the cone without activating a constraint in a way that contributes positively, which becomes incompatible with the assumed inequality for the sequence. In other words, directions outside the cone are eliminated by the structure of the KKT conditions combined with feasibility.

From Taylor's theorem and the limiting direction d:

$$\lambda_j^* c_j(z_k) = \lambda_j^* c_j(x^*) + \lambda_j^* \nabla c_j(x^*)^\mathsf{T}(z_k - x^*) + o(\|z_k - x^*\|) =$$
$$= \|z_k - x^*\| \lambda_j^* \nabla c_j(x^*)^\mathsf{T} d + o(\|z_k - x^*\|).$$

Then,

$$\begin{aligned}\mathcal{L}(z_k, \lambda^*) &= f(z_k) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* c_i(z_k) \\ &\le f(z_k) - \lambda_j^* c_j(z_k) \\ &\le f(z_k) - \|z_k - x^*\| \lambda_j^* \nabla c_j(x^*)^\mathsf{T} d + o(\|z_k - x^*\|).\end{aligned}$$

But Taylor expansion also gives:

$$\mathcal{L}(z_k, \lambda^*) = f(x^*) + O(\|z_k - x^*\|^2),$$

so combining:

$$f(z_k) \ge f(x^*) + \|z_k - x^*\| \lambda_j^* \nabla c_j(x^*)^\mathsf{T} d + o(\|z_k - x^*\|).$$

Comments

To make this exclusion precise, we apply Taylor's theorem to the constraint functions themselves. For the particular active constraint identified, say index j, the value of $\lambda_j^* c_j(z_k)$ expands as the distance from z_k to x^* multiplied by λ_j^* times the gradient dotted with direction d, plus smaller terms. Since $z_k - x^*$ has length equal to the distance from z_k to x^*, this expansion simplifies to that distance multiplied by λ_j^* times the gradient dotted with direction d, plus smaller terms.

Now recall that we previously assumed $\lambda_j^* \nabla c_j(x^*)^\mathsf{T} d$ is strictly positive. Therefore the contribution of this term grows linearly with the distance from x^*. Plugging this estimate back into the Lagrangian expression yields an inequality: the Lagrangian at z_k is bounded above by $f(z_k)-$ a positive multiple of the distance from x^*, up to smaller terms. But on the other hand, Taylor expansion of the Lagrangian about x^* guarantees it equals $f(x^*)+$ terms of order squared distance. Combining both perspectives, we deduce that $f(z_k)$ must be at least $f(x^*)+$ a positive multiple of the distance.

From $\lambda_j^* \nabla c_j(x^*)^\mathsf{T} d > 0$, we get a contradiction with our assumption (that is, that $f(z_k) < f(x^*) + (\sigma/2)\|z_k - x^*\|^2$), so $d \in \mathcal{C}(x^*, \lambda^*)$, and

$$d^\mathsf{T} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d \geq \sigma.$$

Thus, using Taylor expansion and the expression for the limiting direction:

$$\begin{aligned} f(z_k) &\geq f(x^*) + \frac{1}{2}(z_k - x^*)^\mathsf{T} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)(z_k - x^*) + o(\|z_k - x^*\|^2) \\ &= f(x^*) + \frac{1}{2} d^\mathsf{T} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d \|z_k - x^*\|^2 + o(\|z_k - x^*\|^2) \\ &\geq f(x^*) + (\sigma/2)\|z_k - x^*\|^2 + o(\|z_k - x^*\|^2). \end{aligned}$$

This contradicts our assumption, thus we conclude that every feasible sequence $\{z_k\}$ approaching x^* must satisfy $f(z_k) \geq f(x^*) + (\sigma/2)\|z_k - x^*\|^2$, for all k sufficiently large, so x^* is a strict local solution. □

Comments

But this contradicts to our earlier assumption that $f(z_k)$ was strictly less than $f(x^*) + \frac{\sigma}{2}\|z_k - x^*\|^2$. Hence directions d belongs to the critical cone.

In this case, the uniform curvature bound applies: the quadratic form $d^\mathsf{T} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d$ is greater than or equal to σ. Substituting into the Taylor expansion of the Lagrangian, we find that $f(z_k)$ is bounded below by $f(x^*) + \frac{1}{2}$ times this quadratic form times the squared distance, up to higher-order terms.

Since the quadratic form is at least σ, this inequality reduces to $f(z_k)$ being greater than or equal to $f(x^*) + \frac{1}{2}\sigma\|z_k - x^*\|^2$, plus smaller corrections. This clearly contradicts the original assumption that $f(z_k)$ was strictly less than $f(x^*) + \frac{1}{2}\sigma\|z_k - x^*\|^2$. The contradiction shows that no such sequence exists. Therefore, every feasible sequence approaching x^* must eventually satisfy the inequality $f(z_k) \geq f(x^*) + \frac{1}{2}\sigma\|z_k - x^*\|^2$.

This conclusion precisely establishes that x^* is a strict local minimizer: in a neighborhood around x^*, no feasible point yields a smaller objective value, and in fact, the objective increases quadratically away from x^*. This completes the proof of the second-order sufficient condition, which complements the necessary condition studied earlier and provides a robust certificate of strict local optimality.