PART I. OPTIMIZATION: CLASSICAL APPROACHES
(LECTURE 5)

Shpilev Petr Valerievich
Faculty of Mathematics and Mechanics, SPbU

September, 2025

Санкт-Петербургский
государственный
университет

Comments

In this lecture, we will study the theory of constrained optimization, focusing on how restrictions shape the search for optimal solutions. We begin by formulating constrained problems and examining the nature of local solutions through illustrative examples. The lecture introduces the Lagrangian method and first-order optimality conditions for both equality and inequality constraints, highlighting the role of active sets. We then explore geometric intuition via tangent cones, linearized feasible directions, and constraint qualifications such as LICQ. Through a sequence of progressively complex examples, we demonstrate optimal and nonoptimal cases, emphasizing how geometry and constraint structure determine solution behavior.

Minimize $f(x)$ over $x \in \mathbb{R}^n$ subject to

$$\begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I}, \end{cases}$$

where f and c_i are smooth, real-valued functions on a subset of $\mathbb{R}^n$; $\mathcal{E}$ and $\mathcal{I}$ are finite index sets. As before, f is the objective function; c_i, $i \in \mathcal{E}$ are equality constraints; c_i, $i \in \mathcal{I}$ are inequality constraints.

The feasible set is defined as $\Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$ and we can rewrite the problem as

$$\min_{x \in \Omega} f(x). \tag{6}$$

For unconstrained problems, optimality conditions are:

- Necessary: Any local minimizer x^* must satisfy $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$ (positive semidefinite Hessian).
- Sufficient: If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$ (positive definite), then x^* is a strict local minimizer.

Comments

Now we are starting a new fascinating topic: "Constrained Optimization". When moving from unconstrained optimization to constrained optimization, we significantly change the nature of the problem. In unconstrained optimization, our only goal is to minimize the function $f(x)$ over all possible values of x in $\mathbb{R}^n$. Now, however, we impose additional restrictions, expressed as equations or inequalities. These restrictions define what we call the feasible set, denoted Ω. More formally, Ω is the set of all vectors x such that each equality constraint $c_i(x) = 0$ for indices in $\mathcal{E}$ is satisfied, and each inequality constraint $c_i(x) \geq 0$ for indices in $\mathcal{I}$ is also satisfied. The optimization problem then becomes minimizing $f(x)$ only over this feasible set.

It is important to notice how constraints modify the geometry of the search space. Instead of being free to move in every direction, we are restricted to remain within, or on the boundary of, the feasible set. This makes the analysis both richer and more complicated. Just as in the unconstrained case, we will later derive necessary and sufficient conditions for optimality, but these conditions will have to take into account the presence of the constraints. Recall that in the unconstrained setting, a necessary condition for a local minimizer was that the gradient at the point vanished, and the Hessian matrix was positive semidefinite. For sufficiency, the Hessian had to be positive definite. In constrained problems, the situation is similar in spirit but requires entirely new tools, since feasible solutions are no longer determined solely by the landscape of $f(x)$, but also by the geometry of the constraints.

Adding constraints can simplify or complicate the search for global minima:

- Constraints may eliminate many local minima, making the global minimum easier to identify.
- But constraints can also introduce new local minima and complicate the landscape.

Example

$$\min (x_2 + 100)^2 + x_1^2 \quad \text{s.t.} \quad x_2 - \cos x_1 \geq 0$$

- Without constraint: Unique solution at $(0, -100)^T$
- With constraint: Many local solutions near

$$x^{(k)} = (k\pi, -1)^T, \quad k = \pm 1, \pm 3, \dots$$

Comments

Adding constraints has a dual effect on optimization problems. On one hand, constraints can simplify the problem. By eliminating infeasible regions, they may also eliminate extraneous local minima, leaving fewer candidate points to consider, and sometimes making it easier to locate the global minimizer. On the other hand, constraints may create additional complexity. They can introduce new local minimizers that would not exist in the unconstrained case, and these may be scattered throughout the feasible region.

The example illustrates this phenomenon clearly. Consider the function $(x_2+100)^2 + x_1^2$. Without constraints, the global minimizer is unique and easy to identify: the point $(0, -100)$. However, when we impose the inequality constraint $x_2 - \cos x_1 \geq 0$, the feasible region changes drastically. The unconstrained minimizer no longer belongs to the feasible set, and instead, new local minimizers emerge near the points $(k\pi, -1)$, where k is any odd integer. This situation shows how constraints can fundamentally alter the optimization landscape. What was once a single smooth valley with one bottom point becomes a landscape with multiple valleys and many feasible local solutions. From a practical point of view, this means constrained optimization can sometimes be more challenging, since one must distinguish between multiple local solutions and identify which, if any, is globally optimal.

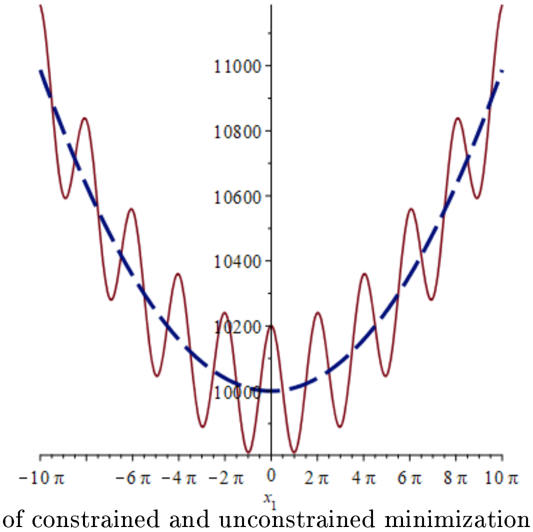

Figure: Comparison of constrained and unconstrained minimization: for each x_1, the unconstrained minimizer lies at $x_2^* = -100$ (blue dash line is shifted by 100^2), while the constraint $x_2 \geq \cos x_1$ shifts the minimizer to the boundary $x_2^* = \cos x_1$ (brown solid line).

Comments

The visual comparison between constrained and unconstrained cases highlights the role of feasibility. In the unconstrained problem, for any value of x_1, the minimizer in the vertical direction occurs at $x_2 = -100$. This corresponds to a flat horizontal line of minima in the (x_1, x_2)-plane. Once the constraint $x_2 \geq \cos x_1$ is introduced, the picture changes completely. The feasible minimizer for each value of x_1 can no longer lie at negative one hundred, because that would violate the constraint. Instead, the minimizer is forced onto the boundary curve $x_2 = \cos x_1$.

This example illustrates an essential principle of constrained optimization: feasible solutions are determined not just by the objective function, but by the interaction between the objective and the feasible set. Often, constraints “push” the minimizer to lie on the boundary rather than inside the region. Thus, constrained optimization problems are frequently about finding the correct balance between the tendency of the objective function to pull the solution toward its unconstrained minimizer, and the restriction of the feasible set, which may prevent this point from being attainable. This boundary interaction is central to the development of optimality conditions, and later we will see how concepts like Lagrange multipliers and Karush-Kuhn-Tucker conditions formalize this balance.

Defenition: local solution

A vector $x^* \in \Omega$ is a local solution of the problem (6) if there exists a neighborhood $\mathcal{N}$ of x^* such that

$$f(x) \geq f(x^*) \quad \text{for all } x \in \mathcal{N} \cap \Omega.$$

Defenition: strict local solution

A vector $x^* \in \Omega$ is a strict (or strong) local solution if there exists a neighborhood $\mathcal{N}$ of x^* such that

$$f(x) > f(x^*) \quad \text{for all } x \in \mathcal{N} \cap \Omega,\ x \neq x^*.$$

Defenition: isolated local solution

A point $x^* \in \Omega$ is an isolated local solution if there exists a neighborhood $\mathcal{N}$ of x^* such that x^* is the only local solution in $\mathcal{N} \cap \Omega$.

Note: Isolated local solutions are always strict, but not every strict local solution is isolated.

Comments

To proceed with constrained optimization, we need precise definitions of what constitutes a local solution. A point x^* in the feasible set Ω is called a local solution if there exists a neighborhood around it in which the objective function does not decrease. More formally, for all feasible points near x^*, the value of $f(x)$ is greater than or equal to $f(x^*)$. If the inequality is strict for all neighboring feasible points distinct from x^*, then x^* is called a strict local solution, or sometimes a strong local solution. The strictness ensures that the minimizer is truly unique in its neighborhood in terms of objective value, and not just one of many points with the same value.

Another refinement is the notion of an isolated local solution. This is a stricter concept: not only is the point a minimizer in its neighborhood, but it is the only such minimizer there. It is important to note that every isolated local solution is necessarily strict, but the reverse is not true. There may be strict local solutions that are not isolated—for instance, when there exists a continuum of minimizers forming a curve or a surface. These distinctions will be crucial later when we discuss stability, uniqueness, and algorithms designed to identify or approximate such solutions.

Smoothness and Reformulation

- Smoothness of f and constraints is essential: it ensures predictable behavior and enables effective optimization algorithms.
- Feasible regions often have nonsmooth boundaries (e.g., "kinks" or "jumps"), but can still be described using smooth constraint functions.
- Example: $\|x\|_1 = |x_1| + |x_2| \le 1$ is nonsmooth, but equivalent to four linear constraints: $x_1 + x_2 \le 1,\ x_1 - x_2 \le 1,\ -x_1 + x_2 \le 1,\ -x_1 - x_2 \le 1.$
- Nonsmooth unconstrained problems can sometimes be reformulated as smooth constrained ones.

Example

Minimize $f(x) = \max(x^2, x)$ Reformulated as:

$$\min t \quad \text{s.t.} \quad t \ge x, \quad t \ge x^2$$

Note: Any inequality constraints can be rewritten in the standard form $c_i(x) \ge 0$ by rearranging terms.

Comments

In optimization, smoothness of both the objective function and the constraints plays a crucial role. A smooth function is differentiable, and this property enables us to rely on gradients and higher-order information to guide optimization algorithms. Without smoothness, the behavior of the function can be unpredictable, making convergence difficult or even impossible for many standard methods. At the same time, it is important to understand that the feasible region itself does not necessarily need to look smooth. For instance, the boundary of a set may contain corners, edges, or discontinuities. Nevertheless, by using suitable reformulations, we can often describe these regions with smooth functions. This ability to transform nonsmooth structures into smooth ones is one of the cornerstones of constrained optimization.

A clear example is the ℓ_1-ball, defined by the condition that the sum of absolute values of coordinates is less than or equal to one. This condition is nonsmooth because the absolute value function is not differentiable at zero. However, the same region can be equivalently represented by a finite set of linear inequalities, each of which is smooth. This reformulation opens the door to using powerful smooth optimization techniques. Another instructive example is a minimization problem involving the maximum of two functions. Such a problem is naturally nonsmooth, because the maximum introduces a "kink." But by introducing an auxiliary variable and adding constraints, we can reformulate the task into a smooth constrained problem. This strategy highlights the flexibility of optimization: even when the original problem looks irregular, with a careful reformulation we can bring it into a setting where smooth mathematical tools apply effectively.

Active Set and Example 1: Equality Constraint

Definition: Active Set

The active set $\mathcal{A}(x)$ at any feasible x consists of the equality constraint indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}.$$

Definition: Active and Inactive Constraints

At a feasible point x, the inequality constraint $i \in \mathcal{I}$ is said to be active if $c_i(x) = 0$ and inactive if the strict inequality $c_i(x) > 0$ is satisfied.

Example 1: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$

In the language of the original task we have $f(x) = x_1 + x_2$, $\mathcal{I} = \emptyset$, $\mathcal{E} = \{1\}$, and $c_1(x) = x_1^2 + x_2^2 - 2$.

Comments

When analyzing constrained optimization problems, a central concept is the so-called active set. At any feasible point, the active set collects all constraints that are currently binding. Binding means that the constraint holds exactly as an equality and leaves no "slack." For equality constraints, this is always the case, since they must hold exactly. For inequality constraints, only those that are tight, meaning satisfied as equalities, are considered active. The others, which are satisfied with strict inequality, are inactive and do not directly affect the local behavior of the solution.

This distinction between active and inactive constraints is fundamental because optimization near the boundary of feasibility is governed by the active set. From a geometric point of view, the active constraints describe the local shape of the feasible region. They can be seen as the walls that confine possible movements. Understanding which constraints are active is therefore equivalent to identifying which walls are currently blocking further progress.

To illustrate, consider a simple problem: minimize the sum of two variables subject to the condition that their squared lengths add up to two. This condition describes a circle of radius $\sqrt{2}$. The feasible set is thus exactly the points lying on this circle. Since this is an equality constraint, it is active everywhere along the boundary. In this case, there are no inequality constraints, so the active set coincides with the set of equality indices. This example shows how the concept of the active set gives us a precise language to describe which restrictions matter at a given point and sets the stage for developing optimality conditions.

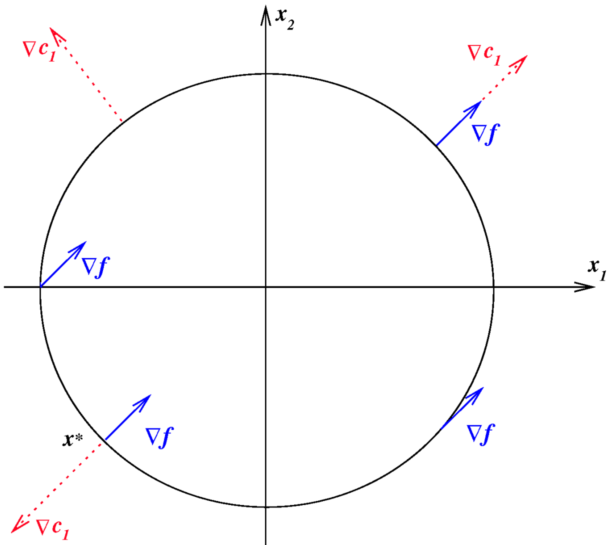

Figure: Problem from Example 1, showing constraint and function gradients at various feasible points.

Comments

The geometric interpretation of constrained optimization provides powerful intuition. Consider again the problem of minimizing the sum of two variables subject to a circular constraint. In this setup, the objective function has gradient vectors that point in the direction of increasing values of the sum. These gradient vectors are constant, since the objective is linear. On the other hand, the constraint corresponds to a circle, and its gradient vectors point radially outward, perpendicular to the boundary.

At feasible points along the circle, we can compare the direction of the objective's gradient with that of the constraint's gradient. If the gradient of the objective points outward the feasible region, then movement against that direction would increase the objective, but it is blocked by the constraint boundary. Conversely, if the gradient points into the feasible region, then movement against it would leave the feasible set. The interesting case arises when the gradient of the target is antiparallel to the gradient of the constraint. At such points, any attempt to move in a direction that maintains feasibility fails to produce descent in the objective. This signals that an optimal solution has been reached.

This simple picture captures a general principle: at an optimum under constraints, the direction of steepest descent cannot be realized because it conflicts with the geometry of feasibility. Instead, the best the optimizer can do is to stop at a point where the objective gradient is balanced against the constraint gradient. This equilibrium is the essence of optimality under equality constraints.

Example 1: Geometry of Optimality

We observe from the figure that at the solution x^* there exists a scalar λ_1^* (in this case, $\lambda_1^* = -\frac{1}{2}$) such that

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).$$

We derive the optimality condition via first-order Taylor expansions:

- Feasibility condition: To stay feasible under $c_1(x) = 0$, we require a small(but nonzero) step s must satisfy that $c_1(x+s) = 0$; that is,

$$c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s = \nabla c_1(x)^T s = 0.$$

- Descent condition: To ensure decrease in $f(x)$, we require

$$f(x+s) - f(x) \approx \nabla f(x)^T s < 0.$$

These two conditions cannot be fulfilled simultaneously, only if $\nabla f(x)$ is collinear to $\nabla c_1(x)$.

Note: No step s can satisfy both conditions if $\nabla f(x)$ and $\nabla c_1(x)$ are collinear to each other, that is, if there exists a scalar λ such that $\nabla f(x) = \lambda \nabla c_1(x)$.

Comments

The key geometric fact is that, at the constrained minimizer, the objective gradient and the constraint normal are parallel. Precisely: $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$, where λ_1^* is strictly negative; in this example it equals $-\frac{1}{2}$. This statement is the first-order optimality balance for a single equality constraint.

To see why, consider small feasible steps. Feasibility with respect to $c_1(x) = 0$ requires that any infinitesimal step s lie in the tangent space of the constraint. In first-order terms, this is $\nabla c_1(x)^T s = 0$. A step that decreases the objective must also satisfy the descent inequality $\nabla f(x)^T s < 0$. If there exists a direction d with $\nabla c_1(x)^T d = 0$ and $\nabla f(x)^T d < 0$, then we can reduce the objective while staying feasible.

Now, such a feasible descent direction exists unless the two gradients are collinear. Equivalently, if the projection of the negative objective gradient onto the constraint's tangent space is nonzero, we can move along that projection and decrease the objective. The only way this procedure fails is when the projection vanishes, which happens exactly when $\nabla f(x) = \lambda \nabla c_1(x)$ for some negative scalar λ. In that collinear case, no first-order feasible descent is possible.

Define the Lagrangian:

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x),$$

with

$$\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x).$$

- First-order condition: At the solution x^*, there exists λ_1^* such that

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0.$$

- λ_1 is called a *Lagrange multiplier*.
- This condition is *necessary*, but not *sufficient*: it may hold at non-optimal points.
- The sign of λ_1^* depends on the form of the constraint.

Comments

The method of Lagrange multipliers is one of the central tools in constrained optimization. The idea is to embed the constraint into the objective function by introducing an auxiliary variable, the so-called Lagrange multiplier. For a problem with a single constraint $c_1(x) = 0$, we define the Lagrangian function as $\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x)$. Taking the gradient with respect to x, we obtain $\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x)$.

At an optimal solution x^*, there must exist some multiplier λ_1^* such that this gradient vanishes, meaning the condition $\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0$ is satisfied. This expresses the fact that at the optimum, the gradient of the objective is aligned with the gradient of the constraint surface. The multiplier itself has an important interpretation: it represents the shadow price or marginal cost of relaxing the constraint.

However, this first-order condition is necessary but not sufficient. In other words, the condition may hold at points that are not actual optima—for instance, at saddle points or at local maxima. Moreover, the sign of the multiplier depends on the way the constraint is written. If we multiply the constraint by minus one, the feasible set does not change, but the multiplier flips sign.

Example 2: A Single Unequality Constraint

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0.$$

- Feasible region: disk defined by the constraint.
- Solution remains $x^* = (-1, -1)^T$.
- At the boundary, the condition

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*), \quad \lambda_1^* = \tfrac{1}{2}$$

  holds, as in the equality case.
- Key difference: For inequality constraints, the sign of λ_1^* matters.

First-order conditions:

- Descent: $\nabla f(x)^T s < 0$.
- Feasibility (to first order): $c_1(x) + \nabla c_1(x)^T s \geq 0$.

Comments

We now extend the framework to problems with inequality constraints. Consider the example of minimizing $x_1 + x_2$ subject to the condition $2 - x_1^2 - x_2^2 \geq 0$. The feasible region is the closed disk defined by the inequality, which includes both the circle and its interior. The optimal solution remains the point $(-1, -1)^T$, exactly as in the equality-constrained case.

At this solution, the gradient of the objective is equal to $\frac{1}{2}$ times the gradient of the constraint, that is, $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$ with $\lambda_1^* = 1/2$. This is structurally the same as in the equality case. The critical difference, however, is that for inequality constraints the multiplier must satisfy a sign condition. Specifically, it must be nonnegative if the constraint is written in the form $c_1(x) \geq 0$. This requirement prevents us from obtaining spurious solutions that would be inconsistent with feasibility.

To test whether a point is optimal, we use two first-order criteria. First, a descent condition: for a candidate step s, the inner product $\nabla f(x)^T s$ must be strictly negative to reduce the objective. Second, a feasibility condition: moving along s must not violate the inequality, which to first order requires $c_1(x) + \nabla c_1(x)^T s \geq 0$. Thus, a feasible direction must simultaneously decrease the function and preserve the constraint. The interplay between these two requirements is at the heart of the Karush–Kuhn–Tucker conditions that generalize the Lagrangian method to inequality constraints.

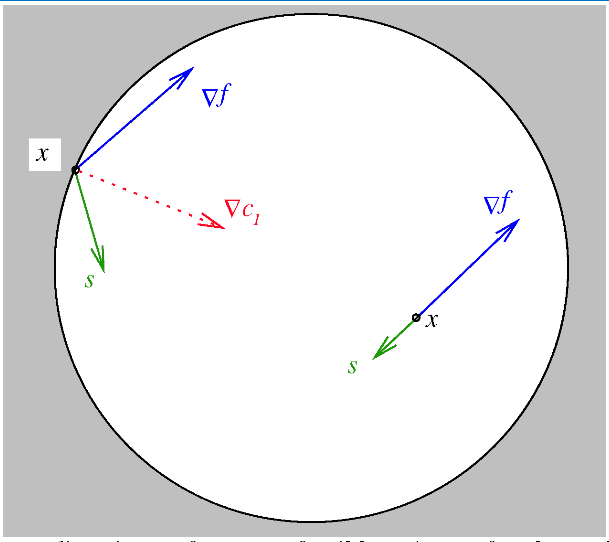

Figure: Improvement directions s from two feasible points x for the problem from Example 2 at which the constraint is active and inactive, respectively.

Comments

The geometry of inequality constraints is well illustrated by considering feasible directions at different types of points. Inside the disk defined by our inequality, the constraint is inactive: small perturbations in any direction remain feasible. At the boundary, however, the constraint becomes active, and movement must be carefully aligned with the feasible region.

To visualize this, imagine two feasible points. At the interior point, every sufficiently small step is allowed, so the optimizer is free to follow the negative gradient direction and reduce the objective. At the boundary point, in contrast, only certain directions preserve feasibility. Specifically, those that do not cross outside the circle. This restriction is equivalent to requiring that the step s satisfies $\nabla c_1(x)^T s \geq 0$.

Therefore, the interaction between the gradient of the objective and the constraint's normal vector determines whether descent is possible. If they are not aligned, then there exist feasible descent directions. If they are aligned, the feasible set blocks further improvement. This intuition explains why the Lagrange multiplier condition emerges naturally: when the gradients coincide up to a nonnegative factor, no descent direction exists, and the point is stationary with respect to the constrained problem.

This picture makes clear the fundamental difference between equality and inequality constraints. Equalities always restrict us to a surface, while inequalities can either be inactive, allowing free movement, or active, restricting feasible directions. This dichotomy lies at the core of constrained optimization theory.

Case I: Interior Point ($c_1(x) > 0$)

- Any sufficiently small step s preserves feasibility:

$$c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s > 0.$$

- If $\nabla f(x) \neq 0$, we can take $s = -\alpha \nabla f(x)$ for small $\alpha > 0$.
- This step yields a first-order decrease in f ($\nabla f(x)^T s < 0$) and keeps $x + s$ feasible.
- If $\nabla f(x) = 0$, no first-order improvement is possible.

Case II: Boundary Point ($c_1(x) = 0$)

- Conditions for improvement and feasibility:

$$\nabla f(x)^T s < 0, \quad \nabla c_1(x)^T s \geq 0.$$

- These define an open and a closed half-space.
- Their intersection is empty if and only if the gradients are positively aligned:

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \quad \lambda_1 \geq 0.$$

Comments

Let us formalize the two possible situations.

Case I: Interior point, where $c_1(x) > 0$. In this case, feasibility is easy to preserve. For any sufficiently small step s, the first-order approximation tells us that $c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s$ remains strictly positive. Thus, the inequality is satisfied automatically. If the gradient of the objective is nonzero, we can safely take $s = -\alpha \nabla f(x)$ for some small positive α. This yields a strict decrease in the function since $\nabla f(x)^T s < 0$. If instead the gradient vanishes, no first-order improvement is possible, and the point is already stationary.

Case II: Boundary point, where $c_1(x) = 0$. Here, the situation is subtler. Any candidate step must both decrease the objective and avoid violating the constraint. This means that $\nabla f(x)^T s < 0$ and $\nabla c_1(x)^T s \geq 0$ must hold simultaneously. Geometrically, the first condition defines an open half-space, while the second defines a closed half-space. Feasible descent directions exist if and only if these two regions overlap.

The intersection becomes empty precisely when the two gradients point in the same direction, that is, when $\nabla f(x) = \lambda_1 \nabla c_1(x)$ for some nonnegative λ_1. In this case, no feasible descent step exists, and the point is a candidate optimum under inequality constraints.

This characterization captures the essential difference from equality-constrained problems. For inequalities, the multiplier is constrained to be nonnegative, ensuring consistency with the feasible region. These ideas generalize naturally to multiple constraints, forming the basis of the Karush–Kuhn–Tucker conditions.

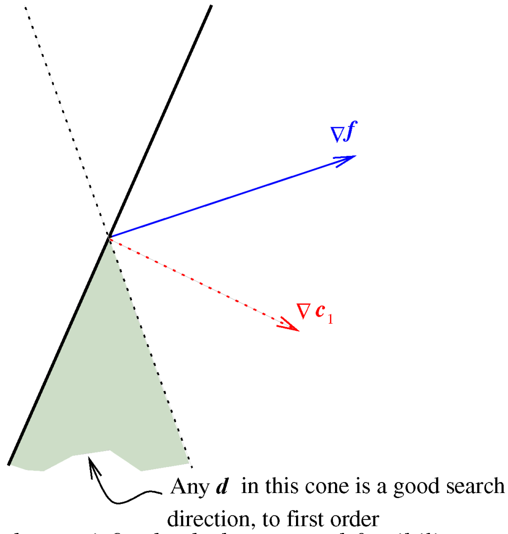

Figure: A direction d that satisfies both descent and feasibility conditions lies in the intersection of a closed half-plane and an open half-plane (light green area).

Comments

When dealing with inequality constraints, the concept of a good search direction becomes more subtle than in unconstrained optimization. A candidate direction must satisfy two simultaneous requirements. First, it should decrease the objective function, which means that the inner product of the gradient of the objective with the step direction must be strictly negative. Second, the direction must preserve feasibility, which requires that the step does not violate the inequality constraint when approximated to first order. Geometrically, these two conditions define two half-planes: one open and one closed. The feasible descent directions are exactly those that belong to the intersection of these two regions. This intersection can be a wedge-shaped sector that narrows or widens depending on how the constraint's gradient aligns with the objective's gradient. If the wedge is nonempty, a descent step exists that simultaneously reduces the function and respects the constraint. If the wedge is empty, then no such direction exists, and we are forced to conclude that the current point is stationary with respect to the constrained problem. This geometric perspective is helpful because it illustrates how feasibility interacts with optimality. In particular, it shows that unlike the unconstrained case, where any negative gradient is a valid descent direction, inequality constraints cut down the available choices. Thus, a "good" search direction is not only about improving the function but also about respecting the boundary geometry of the feasible set.

- The sign of the multiplier matters: if $\lambda_1 < 0$, then $\nabla f(x)$ and $\nabla c_1(x)$ point in opposite directions (if the equality $\nabla f(x) = \lambda_1 \nabla c_1(x)$ holds).
- In this case, descent directions satisfying both

$$\nabla f(x)^T s < 0, \quad \nabla c_1(x)^T s \geq 0$$

would occupy an entire open half-plane.

- When no feasible descent direction exists, optimality is characterized by:

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \quad \lambda_1^* \geq 0,$$

$$\lambda_1^* c_1(x^*) = 0.$$

- The last condition is the complementarity condition, requiring the Lagrange multiplier $\lambda_1^* > 0$ only if $c_1(x^*) = 0$ (constraint is active).
- Case I (interior): $c_1(x^*) > 0 \Rightarrow \lambda_1^* = 0 \Rightarrow \nabla f(x^*) = 0$.
- Case II (boundary): $c_1(x^*) = 0 \Rightarrow \lambda_1^* \geq 0$ allowed, so

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).$$

Comments

The framework can be formalized through the Lagrangian approach, which elegantly incorporates inequality constraints. Here the sign of the multiplier plays a decisive role. If the multiplier were negative, then the gradient of the objective and the gradient of the constraint would point in opposite directions. In that case, feasible descent directions would fill an entire half-plane, contradicting the absence of improvement. Therefore, only nonnegative multipliers are admissible. Optimality is then characterized by three conditions. First, the gradient of the Lagrangian with respect to the decision variables vanishes at the candidate point. Second, the multiplier itself must be nonnegative. Third, we impose the complementarity condition: the product of the multiplier and the constraint value equals zero. This last requirement is crucial, as it enforces that the multiplier can be strictly positive only when the constraint is active, that is, exactly satisfied. Two distinct situations follow. In the interior case, where the inequality is strictly satisfied, the multiplier must vanish, and thus the usual condition that the gradient of the objective is zero is recovered. At the boundary, where the inequality holds with equality, the multiplier may take a nonnegative value, and the stationarity condition requires the gradient of the objective to be a multiple of the gradient of the active constraint. These rules together provide a unified and precise characterization of optimality under inequality constraints.

Example 3: Two Active Constraints

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0$$

- The feasible region is a half-disk. The solution is at $x^* = (-\sqrt{2}, 0)^T$, where both constraints are active.
- As in earlier examples, we look for a direction d such that:

$$\nabla c_i(x)^T d \geq 0, \quad i \in \mathcal{I} = \{1, 2\}, \quad \nabla f(x)^T d < 0$$

- However, no such direction exists at x^*.The conditions $\nabla c_i(x^*)^T d \geq 0$, $i = 1, 2$, are both satisfied only if d lies in the quadrant defined by $\nabla c_1(x^*)$ and $\nabla c_2(x^*)$.
- All directions d in this quadrant satisfy $\nabla f(x^*)^T d \geq 0$, so no first-order descent is possible.

Comments

Let us now examine the case with more than one active constraint. Consider the problem of minimizing the sum of the first variable and the second variable, subject to two conditions: first, that the point lies within the disk of radius $\sqrt{2}$, and second, that the second variable is nonnegative. The feasible region is therefore a half-disk. The optimal solution is at the boundary point where the first variable equals $-\sqrt{2}$ and the second variable equals zero. At this point, both constraints are active simultaneously. To check optimality, we consider possible descent directions. A valid direction must make a nonnegative inner product with the gradient of each constraint, while making a strictly negative inner product with the gradient of the objective. However, these requirements are mutually incompatible at the solution. The gradients of the two constraints define a quadrant, and only directions lying in this quadrant satisfy feasibility. Yet, in this entire quadrant, the inner product with the objective's gradient is nonnegative, meaning that no feasible descent is possible. This situation illustrates the power of the Lagrangian framework with multiple constraints: the conditions exclude any descent direction and thereby certify optimality. It also highlights how the presence of several active constraints can restrict feasible motion much more severely than a single constraint.

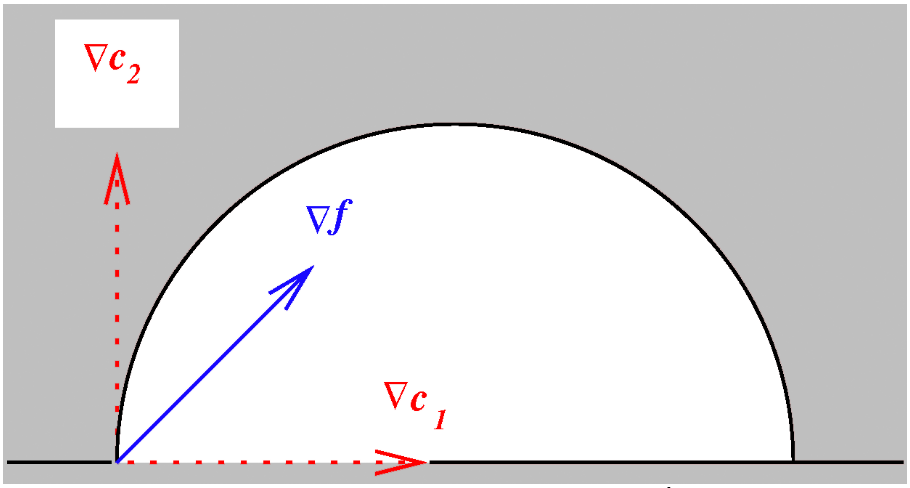

Figure: The problem in Example 3, illustrating the gradients of the active constraints and objective at the solution.

Comments

The geometric picture of this example helps consolidate the key ideas. At the solution point, we can visualize the gradients of both active constraints and of the objective. The two constraint gradients span a cone that restricts all feasible directions. Because the objective's gradient lies within this cone, every feasible step fails to reduce the objective value. This observation confirms stationarity under the Karush–Kuhn–Tucker framework. More importantly, it provides an intuitive understanding of why the complementarity condition and nonnegativity of multipliers are necessary. Each active constraint introduces a boundary, and the associated multiplier adjusts the balance between the objective and the constraint's influence. When multiple constraints are active, their gradients combine to form a feasible cone, and the multipliers weigh how these constraints prevent descent. If the objective gradient can be written as a nonnegative linear combination of the constraint gradients, then no feasible descent exists, and the point is optimal. This geometric interpretation shows that constrained optimization is not only about algebraic conditions but also about the geometry of feasible sets. By visualizing how gradients and feasible directions interact, one gains deeper insight into why the Karush–Kuhn–Tucker conditions provide a complete description of optimality in inequality-constrained problems.

The Lagrangian for the problem in Example 3 is:

$$L(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x),$$

where $\lambda = (\lambda_1, \lambda_2)^T$ is the vector of Lagrange multipliers.

First-order optimality condition:

$$\nabla_x L(x^*, \lambda^*) = 0, \quad \text{for some } \lambda^* \geq 0.$$

Complementarity conditions:

$$\lambda_1^* c_1(x^*) = 0, \quad \lambda_2^* c_2(x^*) = 0.$$

At the point $x^* = (-\sqrt{2}, 0)^T$:

$$\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Choosing:

$$\lambda^* = \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} \\ 1 \end{bmatrix}$$

satisfies the optimality and complementarity conditions.

Comments

At this stage, we extend the method of Lagrange multipliers to the case of two inequality constraints, as posed in Example 3. The Lagrangian is constructed by subtracting each constraint function, multiplied by its corresponding multiplier, from the objective. Thus, the multipliers act as weights that penalize infeasibility, while also shaping the first-order conditions for optimality. Importantly, the multipliers are not arbitrary; they must be nonnegative. This reflects the intuition that constraints only restrict, they never provide "negative pressure."

The first-order stationarity condition states that the gradient of the Lagrangian with respect to the decision variables must vanish at the solution. In other words, the direction of steepest increase of the objective is exactly balanced by a combination of the constraint gradients, scaled by the multipliers. This balancing ensures that no feasible direction exists that can further improve the objective.

Additionally, the complementarity conditions are essential. For each constraint, either the multiplier is zero, meaning the constraint is inactive, or the constraint is active, meaning it binds at the solution and its multiplier can be positive. At the candidate point $(-\sqrt{2}, 0)^T$, the gradients of the objective and the two constraints can be explicitly computed. Choosing multipliers equal to $\frac{1}{2\sqrt{2}}$ and 1, respectively, satisfies both the stationarity and complementarity conditions. Since both multipliers are nonnegative, all conditions for optimality are met.

This point illustrates the fundamental mechanism of constrained optimization: optimality is achieved not by the objective alone, but by a delicate equilibrium between the objective gradient and the constraint gradients.

Nonoptimal Point: Two Active Constraints

We now consider feasible points that are *not* optimal for problem in Example 3, and examine the behavior of the Lagrangian and its gradient.

Point: $x = (\sqrt{2}, 0)^T$

- Both constraints are active at this point.
- A feasible descent direction exists:

$$d = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

- For this x, the Lagrangian gradient condition is satisfied only if

$$\lambda = \begin{bmatrix} -\frac{1}{2\sqrt{2}} \\ 1 \end{bmatrix}$$

- But this violates the condition $\lambda \geq 0$.

Therefore, the Lagrangian gradient condition is not satisfied at $x = (\sqrt{2}, 0)^T$, even though both constraints are active.

Comments

Having established the optimal point, we now examine a different feasible point to see why it fails. Consider the point $(\sqrt{2}, 0)^T$. At this location, both constraints are active, meaning they hold with equality. At first glance, one might expect this to indicate a potential solution, since active constraints often characterize optimal boundaries. However, the analysis shows otherwise.

The existence of a feasible descent direction is crucial. At this point, the vector $(-1, 0)^T$ serves as such a direction: moving in that direction maintains feasibility while strictly reducing the objective. This fact alone suggests that the point cannot be optimal, because improvement is still possible.

From the Lagrangian perspective, the gradient condition can only be satisfied if the multipliers take specific values: $-\frac{1}{2\sqrt{2}}$ and 1. While this combination balances the gradients, the first multiplier is negative. This violates the nonnegativity requirement of the Karush–Kuhn–Tucker conditions. Therefore, the formal optimality conditions are not met.

The lesson here is that merely having multiple active constraints does not guarantee optimality. The key lies in whether the objective gradient can be expressed as a non-negative combination of the active constraint gradients. If not, the geometry indicates that a feasible descent direction must exist, allowing us to move toward better values.

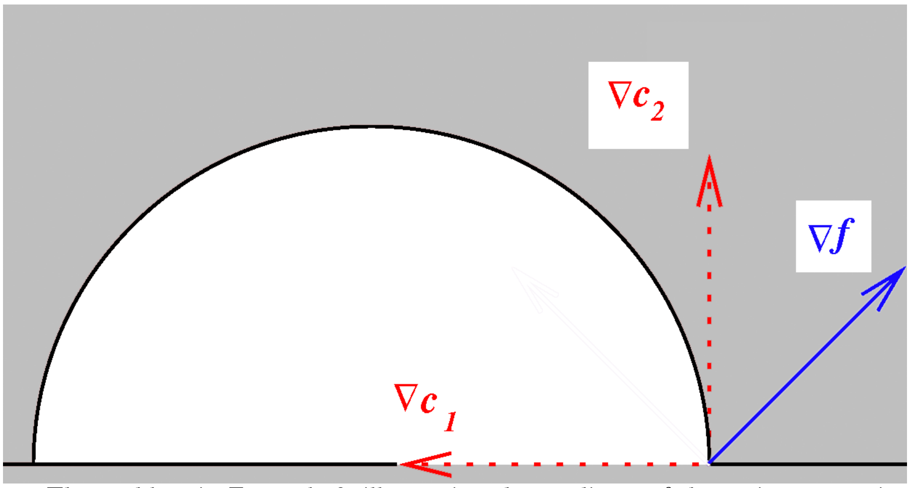

Figure: The problem in Example 3, illustrating the gradients of the active constraints and objective at a nonoptimal point.

Comments

The geometric picture provides additional insight into why the nonoptimal point fails. At the point with coordinates $(\sqrt{2}, 0)^{\mathrm{T}}$, the gradients of the two active constraints are orthogonal, forming directions that restrict movement into the feasible region.

At the point under discussion, the two constraint gradients span a cone of feasible directions. The question is whether the negative gradient of the objective, representing steepest descent, lies inside this cone. If it does, no feasible improvement is possible, and the point could be optimal. If it does not, then a feasible descent direction exists.

Here, the illustration makes clear that the objective gradient cannot be written as a nonnegative combination of the two constraint gradients. This mismatch means that the descent direction avoids the infeasible region, and hence, improvement is possible. Visually, we see that the optimality conditions are geometric constraints on how the objective aligns with the active constraint surfaces.

This graphical interpretation reinforces the algebraic conclusion: although both constraints bind, the point cannot be a solution because the geometry still allows descent. The figure thus bridges the intuition between the Lagrangian equations and the underlying feasible set. It highlights the importance of checking not only which constraints are active, but also how their gradients combine with the objective. With this understanding, we can proceed to examine situations where only one constraint is active.

Nonoptimal Point: One Active Constraint

Point: $x = (1,0)^T$, where only constraint c_2 is active.

- Since any small step s away from this point will continue to satisfy $c_1(x+s) > 0$, only c_2 affects feasible descent directions.
- These directions must satisfy:

$$\nabla c_2(x)^T d \geq 0, \quad \nabla f(x)^T d < 0$$

- With:

$$\nabla f(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_2(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- A feasible descent direction is:

$$d = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{4} \end{bmatrix}$$

To satisfy $\nabla_x L(x, \lambda) = 0$, we must set $\lambda_1 = 0$ (since $c_1(x) > 0$). But no value of λ_2 can satisfy:

$$\nabla f(x) - \lambda_2 \nabla c_2(x) = 0$$

So this point fails to satisfy the optimality conditions.

Comments

At this stage we examine a boundary point where only one of the two inequality constraints is active. The chosen point has the first constraint strictly satisfied, so it plays no role locally. The second constraint, however, lies exactly on the boundary, which means it alone governs the feasible directions. Geometrically, the feasible region at this point resembles a half-plane bounded by a straight line determined by the active inequality. Any admissible step must respect this line and remain on the feasible side.

The gradient of the objective at the point is a vector pointing diagonally upward. The gradient of the active constraint points vertically upward. To determine if improvement is possible, we check whether there exists a direction that both respects the constraint and decreases the objective. By constructing such a vector explicitly, one observes that it indeed exists: a small movement with a negative horizontal component and a slightly positive vertical component keeps feasibility intact while reducing the objective. This demonstrates that the point cannot be optimal, because one can still descend along this feasible direction.

A complementary way to confirm nonoptimality is to attempt balancing the gradients through the method of multipliers. If the point were optimal, one could assign a nonnegative multiplier to the active constraint such that the gradient of the objective becomes exactly offset by this weighted constraint gradient. But here this balancing fails: no value of the multiplier can make the two vectors align. This mismatch underlines that the point lacks the necessary conditions for optimality.

Thus, from both the geometric picture and the algebraic check, we conclude that this boundary point is not optimal.

We examine whether a linearized model gives meaningful information near a feasible point $x^* \in \Omega$.

- The linearized problem is only helpful if it accurately reflects the local geometry of the feasible region near x^*.
- This requires that the feasible set and its linear approximation be "similar" near x^*.
- This similarity is ensured by constraint qualifications.

Definition: feasible sequence

Given a feasible point x, we call $\{z_k\}$ a *feasible sequence* approaching x if

$$z_k \in \Omega \text{ for all large } k, \ \ z_k \to x^*.$$

A tangent direction is a limiting direction of a feasible sequence.
We define the tangent cone $T_\Omega(x^*)$ as the set of all such directions.

Comments

When analyzing constrained optimization problems, we often ask whether the linearized model around a feasible point gives reliable information about the original nonlinear system. This question is subtle, because linearization is useful only if it preserves the essential local geometry of the feasible region. If the linearized approximation is too different from the actual feasible set, then it may predict directions of movement that are impossible in reality. To address this, we introduce the notion of constraint qualifications. These are conditions that guarantee a meaningful relationship between the feasible set and its linearized counterpart.

To formalize the discussion, we define what it means for a sequence to be feasible. A feasible sequence is simply a sequence of points, all lying within the feasible set, that converges to the point under consideration. From this perspective, tangent directions can be understood as limiting directions of such feasible sequences. If you imagine zooming in on the feasible region around the point, tangent directions describe all possible directions in which you can infinitesimally move while remaining feasible.

The set of all these tangent directions is called the tangent cone. The terminology reflects the fact that the collection of tangent directions often resembles a cone emanating from the point. This cone provides the geometric foundation for analyzing optimality conditions. If the tangent cone is aligned in a way that prevents descent of the objective function, the point may be optimal. If instead there exist tangent directions that allow descent, the point cannot be optimal.

Constraint qualifications are thus crucial because they ensure that the tangent cone, which is defined geometrically, corresponds well to the algebraic description provided by the constraint functions. Without such assumptions, the linearized feasible directions may fail to represent the true geometry of the feasible set, leading to misleading conclusions about optimality.

Definition: tangent cone

The vector d is said to be a *tangent* (or tangent vector) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d.$$

The set of all tangents to Ω at x is called the *tangent cone* and is denoted by $T_\Omega(x)$.

Note: the tangent cone $T_\Omega(x)$ is indeed a cone: $d \in T_\Omega(x) \Rightarrow \alpha d \in T_\Omega(x)$, $\forall \alpha > 0$ and $0 \in T_\Omega(x)$ trivially, using $z_k \equiv x$.

Definition: set of linearized feasible directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, *the set of linearized feasible directions* $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \mid d^T \nabla c_i(x) = 0, \text{ for all } i \in \mathcal{E}, \quad d^T \nabla c_i(x) \geq 0, \text{ for all } i \in \mathcal{A}(x) \cap \mathcal{I} \right\}.$$

Comments

To give a more rigorous description of tangent directions, let us define them formally. A direction vector is called tangent at a feasible point if there exists a feasible sequence approaching that point, together with a sequence of positive step lengths tending to zero, such that the difference between each sequence element and the point, scaled by the step length, converges to the direction. In other words, if you zoom in infinitely close to the point, the sequence of steps approaches a straight-line movement along this direction.

The set of all such tangent vectors is called the tangent cone. This terminology emphasizes two important properties. First, if a direction belongs to the tangent cone, then any positive multiple of that direction also belongs to it. This reflects the conic nature of the set. Second, the zero vector is trivially included, since a constant sequence equal to the point itself defines a tangent of zero.

Alongside this purely geometric construction, we introduce the concept of linearized feasible directions. Here, we return to the algebraic formulation of the problem, where constraints are described by equality and inequality functions. Given a feasible point, the active constraints are those that are exactly satisfied at that point. The set of linearized feasible directions consists of all directions that satisfy certain linear conditions: they must be orthogonal to the gradients of equality constraints, and they must not point into the infeasible side of the active inequalities.

This distinction is essential. While the tangent cone depends only on the geometry of the feasible set, the linearized feasible directions depend on the algebraic representation of the constraints. When constraint qualifications hold, these two sets coincide locally. When they do not, the linearized feasible directions may give an inaccurate picture, highlighting why constraint qualifications are fundamental in optimization theory.

Feasible Sequence Approaching a Nonoptimal Point

We return to the first example:

$$\min x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$

Point: $x = (-\sqrt{2}, 0)^T$

- Consider a feasible sequence approaching x:

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}$$

- Let $t_k = \|z_k - x\|$. Then:

$$d = \lim_{k\to\infty} \frac{z_k - x}{t_k} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \text{a tangent direction}$$

- The objective function increases along the sequence:

$$f(z_{k+1}) > f(z_k), \quad \text{for } k = 2, 3, \ldots$$

- Thus, $f(z_k) < f(x)$ for all large k

Since $f(z_k) < f(x)$ for all k, the point $x = (-\sqrt{2}, 0)^T$ cannot be optimal.

Comments

Let us now consider an explicit example that illustrates these definitions. Suppose the objective is to minimize the sum of the two variables, subject to the equality constraint that the sum of their squares equals two. This feasible set is a circle of radius $\sqrt{2}$. Focus on the point with coordinates $(-\sqrt{2}, 0)^T$. This point lies on the circle and is therefore feasible.

To analyze whether it could be optimal, we construct a feasible sequence converging to it. One such sequence is defined by setting the first coordinate equal to $-\sqrt{2 - 1/k^2}$, and the second coordinate equal to $-1/k$. As k increases, these points remain feasible, and they converge to our chosen point. We then normalize the step lengths by choosing t_k equal to the distance between z_k and x. In the limit, the normalized difference yields a tangent direction pointing straight downward.

Examining the objective along this sequence, we find that it strictly increases: each successive point yields a higher value than the previous one, and all values are below the objective at the original point. This observation immediately shows that the point cannot be optimal, since we have exhibited feasible perturbations that reduce the objective.

The geometric lesson is that although the point lies on the feasible set, the tangent cone at this point admits directions of descent.

Feasible Sequence from the Opposite Direction

Now consider a different feasible sequence approaching the same point $x = (-\sqrt{2}, 0)^T$ from the opposite direction:

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix}$$

- This sequence also lies on the constraint set $x_1^2 + x_2^2 = 2$
- As $k \to \infty$, $z_k \to x$ and the direction of approach is:

$$d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- The objective $f(x) = x_1 + x_2$ *decreases* along this sequence
- Tangents from both sequences are of the form $d = (0, d_2)^T$

The tangent cone at $x = (-\sqrt{2}, 0)^T$ is given by

$$T_\Omega(x) = \left\{ \begin{bmatrix} 0 \\ d_2 \end{bmatrix} \middle| d_2 \in \mathbb{R} \right\}$$

Comments

The picture becomes even clearer when we consider an alternative feasible sequence approaching the same point from the opposite side. Again, we keep the first coordinate equal to $-\sqrt{2 - 1/k^2}$, but now set the second coordinate equal to $1/k$. These points also lie on the circle defined by the equality constraint and converge to the same boundary point.

The limiting tangent direction obtained from this sequence points upward along the vertical axis. Thus, when taken together, the two feasible sequences demonstrate that the tangent cone at the point is the entire vertical line through it. More formally, the tangent cone consists of all vectors with zero in the first coordinate and any real value in the second coordinate.

From the perspective of the objective, this second sequence has the opposite effect: along it, the value of the objective decreases rather than increases. In fact, the objective strictly improves as k grows. The coexistence of both directions — one yielding improvement and the other worsening — shows that the tangent cone captures all possible infinitesimal approaches.

This example reinforces the power of the tangent cone concept. By characterizing all potential limiting directions, it provides a complete geometric picture of feasible movements. Whether a point is optimal depends on how the objective aligns with these directions. If every tangent direction fails to reduce the objective, the point may be optimal. But if even one direction allows descent, optimality is ruled out.

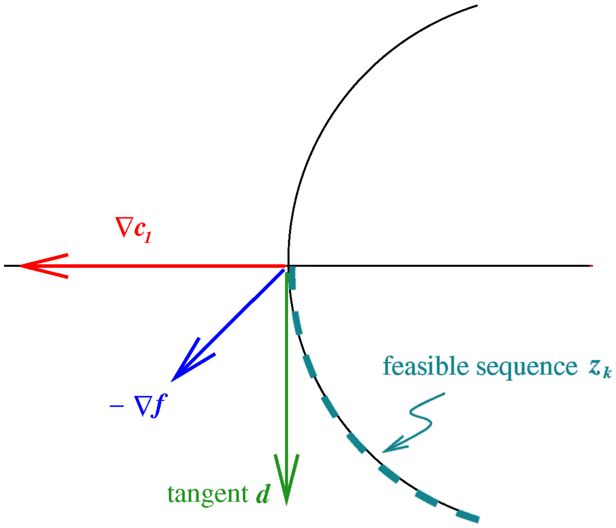

Figure: Constraint normal, objective gradient, and feasible sequence for Example 4.

Comments

We now examine a new example that further illustrates the relationship between constraint geometry and optimality. Consider the circle defined by the equation $x_1^2 + x_2^2 = 2$. The figure shows the boundary of this constraint together with the negative gradient of the objective function and a feasible sequence approaching the boundary point. At the chosen point, with coordinates $(-\sqrt{2}, 0)^T$, the constraint surface is smooth and its normal vector is horizontal, pointing leftward. This normal represents the gradient of the constraint function, and it defines the directions that are infeasible.

The negative objective gradient, on the other hand, points diagonally downward, forming a nonzero angle with the constraint normal. The important observation is that feasible directions lie tangentially along the circle at this point. One can move vertically up or down while remaining feasible. The figure highlights this by sketching a possible feasible sequence, which traces the curve of the circle while respecting the constraint.

This example shows that geometric intuition plays a key role in constrained optimization. The feasible region is not simply any set of directions, but rather those aligned with the tangent space of the constraint surface. The gradient of the objective must be considered relative to this tangent space: if there exists a feasible direction along which the objective decreases, then the point cannot be optimal. In this case, such directions clearly exist, as illustrated by the vertical movement along the circle. This confirms that geometry provides both intuition and rigorous guidance when analyzing constrained problems.

Example 4

We return to the original problem: $x_1^2 + x_2^2 - 2 = 0$
and evaluate the linearized feasible directions at $x = (-\sqrt{2}, 0)^T$:

$$\nabla c_1(x)^T d = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2} d_1 = 0$$

- Hence, $\mathcal{F}(x) = \left\{ \begin{bmatrix} 0 \\ d_2 \end{bmatrix} \middle| d_2 \in \mathbb{R} \right\}$
- In this case: $\mathcal{F}(x) = T_\Omega(x)$

Now suppose that the feasible set is defined instead by the formula

$$\Omega = \{x \mid c_1(x) = 0\}, \quad \text{where } c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0.$$

Then:

$$\nabla c_1(x) = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- This implies $\nabla c_1(x)^T d = 0$ for all d
- So $\mathcal{F}(x) = \mathbb{R}^2$, while $T_\Omega(x)$ remains unchanged

Even though the set Ω is the same, its algebraic form affects $\mathcal{F}(x)$.
In this case: $\mathcal{F}(x) \neq T_\Omega(x)$.

Comments

Let us now analyze the same problem from the perspective of linearized feasible directions. The constraint is given by the equation $x_1^2 + x_2^2 - 2 = 0$. At the point with coordinates $(-\sqrt{2}, 0)^T$, the gradient of the constraint is the vector with components $-2\sqrt{2}$ and 0. The condition for a feasible direction is that the gradient of the constraint, transposed and multiplied by the direction vector, equals zero. Substituting, this becomes $-2\sqrt{2} \cdot d_1 = 0$. Therefore, d_1 must be zero, while d_2 is arbitrary. Hence, the linearized feasible set consists of all vertical directions, which coincides with the tangent cone at this point.

Now suppose we represent the same feasible set in an alternative algebraic form. Instead of writing the constraint as $x_1^2 + x_2^2 - 2 = 0$, we square this expression and write $c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0$. Although the underlying set is identical, its gradient behaves differently. Computing the gradient at the same point gives the zero vector. This means that the linearized condition reduces to $0 = 0$ for all directions, implying that every direction is feasible in the linearized sense. Thus, the linearized feasible set becomes the entire plane, even though the tangent cone remains vertical.

This example is crucial because it shows that the algebraic representation of constraints influences the linearized feasible set. The geometry of the feasible region is unchanged, but its linearized approximation depends on how the constraint is written. In this modified form, the linearized feasible set no longer matches the tangent cone, demonstrating that consistency between algebraic form and geometric reality is essential in optimization theory.

Example 5

We reconsider the problem (from Example 2):

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$

The solution $x = (-1, -1)^T$ is the same as in the equality-constrained version.

- There are infinitely many feasible sequences converging to a boundary point $x = (-\sqrt{2}, 0)^T$ along a straight line from the interior of the circle.
- These sequences have the form:

$$z_k = (-\sqrt{2}, 0)^T + \frac{1}{k} w,$$

where w is any vector whose first component is positive $w_1 > 0$.
- The point z_k is feasible provided that $\|z_k\| \leq \sqrt{2}$, that is:

$$\left(-\sqrt{2} + \frac{w_1}{k}\right)^2 + \left(\frac{w_2}{k}\right)^2 \leq 2$$

This inequality holds for $k \geq \dfrac{w_1^2 + w_2^2}{2\sqrt{2} w_1}$

Comments

When we move from equality-constrained optimization problems to inequality-constrained ones, the geometry of feasible sets becomes richer and more complex. Consider the problem of minimizing the sum of two variables, subject to the constraint that the point lies inside or on the boundary of a circle of radius $\sqrt{2}$. At first glance, this problem looks similar to its equality-constrained counterpart, but the difference lies in the nature of the feasible region. Instead of being restricted to the curve of the circle, we now have access to the entire disk.

The point of interest, $(-\sqrt{2}, 0)^T$, lies on the boundary of this disk. To study tangent cones, we examine how feasible sequences approach this boundary point. Importantly, there are infinitely many ways to approach it. One family of sequences takes the form of straight lines from the interior, written as $(-\sqrt{2}, 0)^T + \frac{1}{k} w$, where the first component of w is strictly positive. These sequences remain feasible as long as their norm is less than or equal to $\sqrt{2}$. By analyzing the inequality that ensures feasibility, we find a lower bound on k depending on the components of w.

The key lesson here is that the inequality-constrained case permits a larger set of feasible directions than the equality case. This abundance of feasible sequences provides more candidate tangent directions, enriching the tangent cone. Thus, the transition from equality to inequality not only changes the algebraic description but also deepens the geometric picture of how solutions can be approached.

In addition to straight-line feasible sequences, we can construct infinitely many curved sequences approaching $x = (-\sqrt{2}, 0)^T$ from the interior of the circle.

- These sequences also contribute to the tangent cone
- Therefore, the tangent cone at $x = (-\sqrt{2}, 0)^T$ is:

$$T_\Omega(x) = \{(w_1, w_2)^T \mid w_1 \geq 0\}$$

- For the inequality-constrained definition:

$$\Omega = \{x \mid 2 - x_1^2 - x_2^2 \geq 0\}$$

  we have from the definition of the set of linearized feasible directions:

$$d \in \mathcal{F}(x) \quad \text{if} \quad \nabla c_1(x)^T d \geq 0$$

- At $x = (-\sqrt{2}, 0)^T$, this becomes:

$$\nabla c_1(x)^T d = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 2\sqrt{2} d_1 \geq 0$$

We conclude that $\mathcal{F}(x) = T_\Omega(x)$ for this formulation of the feasible set.

Comments

Building on this, we recognize that straight-line sequences are not the only paths to the boundary point. In fact, curved paths that remain entirely within the disk also converge to the same boundary location. This observation significantly expands the collection of feasible sequences. The consequence is that the tangent cone at $(-\sqrt{2}, 0)^T$ is not restricted to a narrow set of directions but instead includes all vectors whose first component is nonnegative. Put differently, the tangent cone consists of every direction that points outward or along the boundary without moving back into the infeasible region.

To connect this geometric description with the algebraic definition, we recall that feasible directions can also be characterized by linearization. For inequality constraints, a vector d is feasible if the gradient of the constraint at the point, transposed and multiplied by d, is greater than or equal to zero. For our example, the gradient of c_1 at $(-\sqrt{2}, 0)^T$ simplifies to the vector $(2\sqrt{2}, 0)^T$. The feasibility condition then reduces to $2\sqrt{2} \cdot d_1 \geq 0$, which simply means that the first component of d must be nonnegative.

This algebraic condition perfectly matches the earlier geometric interpretation. Therefore, in this case, the set of linearized feasible directions coincides exactly with the tangent cone. This alignment emphasizes that the tangent cone is not an abstract construct but one that faithfully represents both geometry and algebra when the constraints are well-behaved.

Constraint qualifications are conditions under which the linearized feasible set $\mathcal{F}(x)$ accurately reflects the tangent cone $T_\Omega(x)$.

- Typically, constraint qualifications ensure that $\mathcal{F}(x) = T_\Omega(x)$.
- They guarantee that the linearization captures the essential local geometry of Ω near x.

Example: Constraints

$$c_1(x) = 1 - x_1^2 - (x_2 - 1)^2 \geq 0, \quad c_2(x) = -x_2 \geq 0$$

define the feasible set $\Omega = \{(0,0)^T\}$.

- At $x = (0,0)^T$, all feasible sequences approaching x must have $z_k = x = (0,0)^T$ for all k sufficiently large

$$\Rightarrow T_\Omega(x) = \{(0,0)^T\}$$

- However, linearizing gives:

$$\mathcal{F}(x) = \{(d_1, 0)^T \mid d_1 \in \mathbb{R}\}$$

In this case, $\mathcal{F}(x) \neq T_\Omega(x)$ — constraint qualifications are not satisfied.

Comments

We have just considered the example in which the tangent cone and the linearized feasible set agree. However, this is not always guaranteed. To formalize the relationship, we introduce the concept of constraint qualifications. These are conditions that ensure the linearized feasible directions truly capture the local geometry of the feasible region. In many practical problems, constraint qualifications guarantee that the linearized feasible set equals the tangent cone, preserving consistency between algebraic approximation and geometric reality.

To see why these conditions are important, consider an example in which the feasible region collapses to a single point, the origin. This occurs when we impose two inequalities simultaneously: the first describes the interior of a shifted circle, while the second restricts us to nonnegative vertical coordinates. Their intersection reduces the feasible set to the single point $(0,0)^T$. In this setting, the tangent cone is trivial; it contains only the zero vector, since there is no way to move infinitesimally within the feasible region.

Yet, when we attempt to construct the linearized feasible set, the outcome is quite different. Linearization suggests that all horizontal directions are allowed, yielding an entire axis of feasible directions. This mismatch between geometry and algebra reveals that constraint qualifications are not satisfied. The linearized set no longer represents the true local behavior of the feasible region.

Thus, constraint qualifications serve as safeguards. They ensure that the simplified linear models we use in optimization remain faithful to the underlying geometry. Without them, algorithms risk following directions that do not actually preserve feasibility, potentially undermining convergence and accuracy.

Definition: LICQ

Given the point x and the active set $\mathcal{A}(x)$, we say that the *linear independence constraint qualification* (LICQ) holds if the set of active constraint gradients

$$\{\nabla c_i(x),\ i \in \mathcal{A}(x)\}$$

is linearly independent.

- LICQ *fails* for:
  - Example with $\Omega = \{(0,0)^T\}$
  - The alternative formulation in Example 4
- If LICQ holds, then *none* of the active constraint gradients can be zero.

Comments

Among the various constraint qualifications, one of the most fundamental and widely used is the linear independence constraint qualification, often abbreviated LICQ. The principle behind LICQ is straightforward but powerful. At any feasible point, we consider the active constraints—those inequalities that are satisfied exactly at equality. Each of these constraints contributes a gradient vector, and LICQ requires that this collection of gradients be linearly independent.

The intuition is that if the active constraints are independent, they intersect cleanly and define a well-structured feasible region locally. In contrast, if some of these gradients are redundant or linearly dependent, the linearized feasible set may misrepresent the true tangent cone, leading to difficulties in optimization. For example, in the case where the feasible set reduces to a single point, the active constraints do not form an independent set. Consequently, LICQ fails, and the linearized model diverges from geometric reality.

Another important implication is that none of the active gradients can be the zero vector. A zero gradient would provide no directional information, violating the independence requirement. Ensuring LICQ holds is particularly significant in the design of optimization algorithms, because many theoretical guarantees—such as the existence of Lagrange multipliers and the validity of optimality conditions—rely on it.

In summary, LICQ acts as a cornerstone condition in constrained optimization. By demanding linear independence among active constraint gradients, it ensures that the interplay between algebraic linearization and geometric structure remains intact, paving the way for robust and reliable analysis.