PART I. OPTIMIZATION: CLASSICAL APPROACHES
(LECTURE 7)

Shpilev Petr Valerievich

Faculty of Mathematics and Mechanics, SPbU

September, 2025

Санкт-Петербургский государственный университет

Comments

In this lecture, we continue the study of optimality conditions and move toward duality theory in constrained optimization. We start with examples illustrating second-order sufficient conditions, unbounded objectives, and the role of constraint qualifications such as LICQ and MFCQ in guaranteeing validity of optimality results. Geometric perspectives involving normal cones and necessary conditions are introduced to deepen intuition. We then examine sensitivity analysis through Lagrange multipliers and distinguish between strongly and weakly active constraints. The second part of the lecture develops duality theory, beginning with the Lagrangian dual, weak duality, and the link to KKT conditions, before exploring primal–dual relationships and the Wolfe dual. Finally, we apply these ideas to linear and convex quadratic programming, covering standard form transformations, KKT conditions, dual formulation, and the proof of strong duality, thereby connecting theory with fundamental optimization models.

Example 8 (revisiting Example 2)

Objective: $f(x) = x_1 + x_2$,
Constraint: $c_1(x) = 2 - x_1^2 - x_2^2$
Index sets: $\mathcal{E} = \emptyset$, $\mathcal{I} = \{1\}$
Lagrangian: $\mathcal{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2)$
At $x^* = (-1, -1)^\mathsf{T}$, $\lambda_1^* = \frac{1}{2}$ satisfies KKT.

Lagrangian Hessian:

$$\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ is positive definite,}$$

so it certainly satisfies the conditions of Theorem Second-Order Sufficient Conditions (Theorem 24) $\Rightarrow x^*$ is a strict local solution (and in fact global).

Example 9

Now let's consider a more complex problem

$$\min -0.1(x_1 - 4)^2 + x_2^2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 \geq 0,$$

in which we seek to minimize a nonconvex function over the exterior of the unit circle.

Comments

Having established the second-order sufficient conditions for optimality, it is natural to verify how they operate in practice. The theory by itself is abstract, but the true power of such results becomes visible when we apply them to explicit problems. Let us therefore revisit a familiar optimization setting and check whether the conditions can indeed confirm the optimality of a candidate solution. This process not only consolidates our understanding of the theorem but also illustrates how local analysis based on derivatives can provide global insights.

In the first case, we minimize a simple linear objective subject to a quadratic inequality. The candidate solution at hand satisfies the Karush–Kuhn–Tucker conditions with an associated multiplier, and when we form the Hessian of the Lagrangian, it turns out to be positive definite. This immediately triggers the sufficient conditions: the point must be a strict local minimizer. In fact, because of the geometry of the feasible set and the simplicity of the linear objective, the solution is not just local but also global.

With this straightforward example confirming the theorem, we turn next to a richer problem where the objective is nonconvex. Specifically, the function contains a negative quadratic term, making it unbounded below if unconstrained. The feasible set is also unusual: optimization takes place not inside but outside the unit circle. This setting forces us to rely on the machinery of optimality conditions, as intuition alone may not be enough to determine whether local solutions exist and how they behave.

The objective function is not bounded below on the feasible region: e.g., consider the feasible sequence

$$\begin{bmatrix}10\\0\end{bmatrix}, \begin{bmatrix}20\\0\end{bmatrix}, \begin{bmatrix}30\\0\end{bmatrix}, \begin{bmatrix}40\\0\end{bmatrix}, \dots \Rightarrow f(x) \to -\infty$$

$\Rightarrow$ No global solution exists.

Still: A strict local solution on the constraint boundary may exist.

Approach: Use KKT conditions and second-order conditions (Theorem 24). Lagrangian derivatives:

$$\nabla_x \mathcal{L}(x,\lambda) = \begin{bmatrix}-0.2(x_1-4) - 2\lambda_1 x_1\\ 2x_2 - 2\lambda_1 x_2\end{bmatrix},$$

$$\nabla^2_{xx} \mathcal{L}(x,\lambda) = \begin{bmatrix}-0.2 - 2\lambda_1 & 0\\ 0 & 2-2\lambda_1\end{bmatrix}$$

Comments

The new example presents an interesting challenge: the objective function is nonconvex and, in fact, not bounded below over the feasible region. If we move far enough along the horizontal axis while remaining feasible, the function tends to minus infinity. This observation immediately tells us that a global solution cannot exist. Yet, the absence of a global minimizer does not exclude the possibility of meaningful local solutions. Indeed, optimization theory teaches us that local minima may persist even in nonconvex and unbounded settings.

To identify such candidates, we turn again to the Karush–Kuhn–Tucker framework. By forming the Lagrangian and computing its first and second derivatives, we prepare to test both first- and second-order conditions. The gradient of the Lagrangian gives necessary equations linking the primal variables with the multiplier, while the Hessian describes the curvature of the Lagrangian in the neighborhood of feasible points. At this stage, the analysis is less about guessing and more about systematically applying the theory to filter out points that can legitimately serve as local minimizers.

This example is particularly instructive because it shows that even when the objective "escapes" to negative infinity, local minima constrained to the boundary of the feasible set may still appear. Such solutions are inherently different from global ones, but they are crucial in applications: algorithms for nonlinear optimization typically converge to local solutions, and the theory of second-order conditions gives us tools to certify them.

Point $x^* = (1,0)^\mathsf{T}$ satisfies KKT conditions with $\lambda_1^* = 0.3$, active set $\mathcal{A}(x^*) = \{1\}$.

Compute critical cone:

$$\nabla c_1(x^*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \Rightarrow \mathcal{C}(x^*, \lambda^*) = \left\{(0, w_2)^\mathsf{T} \mid w_2 \in \mathbb{R}\right\}$$

Substitute into Hessian, we have for any $w \in \mathcal{C}(x^*, \lambda^*)$ with $w_2 \neq 0$:

$$\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} -0.4 & 0 \\ 0 & 1.4 \end{bmatrix} \quad \Rightarrow \quad w^\mathsf{T}\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w = 1.4w_2^2 > 0$$

$\Rightarrow$ Second-order sufficient conditions hold $\Rightarrow$ by Theorem 24 $x^* = (1,0)^\mathsf{T}$ is a strict local solution.

Comments

Carrying out the analysis, we arrive at a specific candidate point on the boundary of the feasible set. This point satisfies the Karush–Kuhn–Tucker conditions with a positive multiplier, which already signals that it deserves closer attention. To proceed, we construct the critical cone, representing the feasible directions that remain tangent to the active constraint. This step is essential because second-order sufficient conditions require positivity of the Lagrangian Hessian only along those directions, not in the entire space.

When we substitute the cone directions into the Hessian, the quadratic form turns out to be strictly positive for all nonzero feasible directions. This observation confirms the sufficient condition: the candidate point is indeed a strict local minimizer. Although no global minimum exists, this result demonstrates that the theory successfully identifies local structure even in an unbounded and nonconvex problem.

From a conceptual perspective, the lesson here is significant. Optimization is not always about global best solutions; often it is about characterizing and certifying local optima that arise naturally from constraints and curvature. The second-order sufficient conditions serve exactly this role: they bridge first-order stationarity with genuine optimality. By applying them, we gain both a rigorous test and a deeper understanding of how local minima emerge in constrained optimization.

Example: S-O Suf. Cond.
Constraint Qualifications
Sensitivity Analysis
Duality Theory
Wolfe Dual
LP
KKT Conditions
Dual Problem

Constraint qualifications ensure that the linearized approximation to the feasible set Ω captures its essential shape near x^*.

- A key case occurs when all *active constraints* are linear:
$$c_i(x) = a_i^T x + b_i, \quad \text{for some } a_i \in \mathbb{R}^n,\ b_i \in \mathbb{R}.$$
- In this case, the linearized feasible direction set $\mathcal{F}(x^*)$ accurately represents the feasible set Ω.

Lemma 8: Linear Constraints

Suppose that at some $x^* \in \Omega$, all active constraints $c_i(\cdot)$, $i \in \mathcal{A}(x^*)$, are linear functions. Then:
$$\mathcal{F}(x^*) = \mathcal{T}_\Omega(x^*).$$

Proof: From a previous result (Lemma 5), $\mathcal{T}_\Omega(x^*) \subset \mathcal{F}(x^*)$. To show $\mathcal{F}(x^*) \subset \mathcal{T}_\Omega(x^*)$, take any $w \in \mathcal{F}(x^*)$ and prove $w \in \mathcal{T}_\Omega(x^*)$. By the definition of $\mathcal{F}(x^*)$ and the linear constraint form $c_i(x) = a_i^T x + b_i$, we have:
$$\mathcal{F}(x^*) = \left\{ d \;\middle|\; \begin{array}{ll} a_i^T d = 0, & \text{for all } i \in \mathcal{E}, \\ a_i^T d \geq 0, & \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{array} \right\}.$$

Comments

When studying constraint qualifications, an especially transparent situation arises when all active constraints at the candidate point are linear. In this case, the linearized feasible direction set is not merely an approximation but in fact coincides exactly with the tangent cone of the feasible set. To see why, recall that each active constraint has the form $a_i^T x + b_i$. Because these are linear expressions, the first-order expansion does not introduce any curvature. Therefore, the feasible directions obtained by linearization are already the true feasible directions of the original problem. This observation gives us a powerful simplification: the feasible cone defined by conditions $a_i^T d = 0$ for equality constraints and $a_i^T d \geq 0$ for active inequalities is precisely the tangent cone at x^*. The proof of this result relies on constructing feasible sequences that approach x^* in directions belonging to the linearized cone. Intuitively, this means that if we take any direction consistent with the linear active constraints, then by moving a small positive distance along this direction we stay inside the feasible region. Thus, the linearized model loses no information about the true geometry of the set. This lemma is an important foundation, because in many optimization problems active constraints do turn out to be linear, such as in linear programming. In those cases, verifying the condition is immediate, and the tangent cone characterization becomes exact. Hence, the gap between approximation and reality disappears, making analysis considerably easier.

There exists a positive scalar $\bar{t}$ such that inactive constraints remain inactive at $x^* + tw$ for $t \in [0, \bar{t}]$:

$$c_i(x^* + tw) > 0, \quad \text{for all } i \in \mathcal{I} \setminus \mathcal{A}(x^*),\ t \in [0, \bar{t}].$$

Define the feasible sequence:

$$z_k = x^* + \frac{\bar{t}}{k} w, \quad k = 1, 2, \dots$$

For active inequality constraints $i \in \mathcal{I} \cap \mathcal{A}(x^*)$, since $a_i^T w \geq 0$:

$$c_i(z_k) = c_i(z_k) - \underbrace{c_i(x^*)}_{=0} = a_i^T (z_k - x^*) = \frac{\bar{t}}{k} a_i^T w \geq 0.$$

Thus, z_k is feasible for active inequality constraints. By choice of $\bar{t}$, z_k is feasible for inactive inequality constraints $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$. For equality constraints $i \in \mathcal{E}$:

$$c_i(z_k) = a_i^T z_k + b_i = a_i^T \left(x^* + \frac{\bar{t}}{k} w \right) + b_i = a_i^T x^* + b_i = c_i(x^*) = 0.$$

Hence, z_k is feasible for each $k = 1, 2, \dots$. In addition, we have that

$$\frac{z_k - x^*}{\bar{t}/k} = \frac{(\bar{t}/k) w}{\bar{t}/k} = w,$$

so that indeed w is the limiting direction of $\{z_k\}$. Hence, $w \in \mathcal{T}_\Omega(x^*)$. □

Comments

Let's prove this lemma. The key idea is to show that any direction allowed by the linearized system can be realized as an actual limiting feasible direction. Continuity yields that for each inactive constraint we can find a small positive scalar, let us call it $\bar{t}$, such that moving from x^* by any step t in the range between zero and $\bar{t}$ keeps those constraints inactive. We then define a sequence $z_k = x^* + \frac{\bar{t}}{k} w$. As k grows, these points move closer and closer to x^*, but they always remain feasible. For the active inequality constraints, feasibility is preserved because the expression $a_i^T w$ is nonnegative, and hence $c_i(z_k)$ remains nonnegative. For the equality constraints, feasibility holds exactly, because the linear structure ensures that $c_i(z_k) = c_i(x^*)$, which is zero. Consequently, every element of the sequence z_k lies inside the feasible set. Finally, we compute the limit of the normalized difference between z_k and x^*, and we recover exactly the direction w. This shows that w is indeed a tangent direction, so it belongs to the tangent cone. The conclusion is that the linearized cone and the tangent cone coincide. This constructive approach highlights not just the logic of the proof but also its geometric meaning: feasible directions can be followed by feasible curves, so the geometry of the feasible set is faithfully captured by the linear approximation whenever constraints are linear.

Constraint Qualifications: MFCQ

Linear constraints and LICQ are constraint qualifications ensuring $\mathcal{F}(x^*)$ captures the geometry of Ω near x^*.

- Linear active constraints imply $\mathcal{F}(x^*) = \mathcal{T}_\Omega(x^*)$, a constraint qualification neither weaker nor stronger than LICQ.

Definition: MFCQ

Mangasarian-Fromovitz Constraint Qualification (MFCQ) holds if there exists $w \in \mathbb{R}^n$ such that:

$$\begin{aligned} \nabla c_i(x^*)^T w &> 0, \quad \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I}, \\ \nabla c_i(x^*)^T w &= 0, \quad \text{for all } i \in \mathcal{E}, \end{aligned}$$

and the set $\{\nabla c_i(x^*), i \in \mathcal{E}\}$ is linearly independent.

- Note the ***strict inequality*** for active inequality constraints.
- MFCQ is weaker than LICQ:
  - If LICQ holds, the system

$$\begin{aligned} \nabla c_i(x^*)^T w &= 1, \quad \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I}, \\ \nabla c_i(x^*)^T w &= 0, \quad \text{for all } i \in \mathcal{E}, \end{aligned}$$

    has a solution w, satisfying MFCQ.
  - MFCQ can hold when LICQ does not.

Comments

Beyond the special case of linear constraints, optimization theory introduces broader conditions known as constraint qualifications. These conditions ensure that the linearized approximation reflects the true geometry of the feasible set, even when nonlinear constraints are present. Among the most important is the Mangasarian–Fromovitz Constraint Qualification, or MFCQ. It requires the existence of a vector w such that for every active inequality constraint, the gradient of the constraint at x^*, transposed and multiplied by w, is strictly positive. For equality constraints, the same expression must be zero. Additionally, the set of gradients of equality constraints must be linearly independent. The strict inequality condition is crucial: it demands that we can find a feasible direction pointing strictly into the interior of the feasible set, thereby preventing degeneracy at the boundary. Compared with the Linear Independence Constraint Qualification, or LICQ, MFCQ is weaker. If LICQ holds, then MFCQ automatically follows, because one can solve a simple linear system to find a suitable w. However, the converse is not true: MFCQ may hold even when LICQ fails, for example when gradients of active inequalities are linearly dependent. This flexibility makes MFCQ especially useful in nonlinear problems, where strict linear independence may be too demanding. In practice, MFCQ guarantees that first-order optimality conditions, such as the Karush–Kuhn–Tucker system, remain valid and meaningful. Thus, it forms a cornerstone of constrained optimization theory.

MFCQ can hold when LICQ fails if active constraint gradients are linearly dependent but a vector w satisfies MFCQ conditions.

Example

Consider the feasible set $\Omega = \{x \in \mathbb{R}^2 \mid c_1(x) = -x_1 \geq 0,\ c_2(x) = -x_1 - x_2^2 \geq 0\}$ at $x^* = (0,0)^T$.

- Active constraints: $\mathcal{A}(x^*) = \{1,2\}$, with $c_1(x^*) = 0$, $c_2(x^*) = 0$.
- Gradients: $\nabla c_1(x^*) = [-1,0]^T$, $\nabla c_2(x^*) = [-1,0]^T$.
- LICQ fails: $\{\nabla c_1(x^*), \nabla c_2(x^*)\}$ is linearly dependent (identical vectors).
- MFCQ holds: Choose $w = [1,0]^T$. Then:

$$\nabla c_1(x^*)^T w = [-1,0][1,0]^T = -1 < 0,$$
$$\nabla c_2(x^*)^T w = [-1,0][1,0]^T = -1 < 0.$$

  Adjust $w = [-1,0]^T$:

$$\nabla c_1(x^*)^T w = [-1,0][-1,0]^T = 1 > 0, \quad \nabla c_2(x^*)^T w = 1 > 0.$$

  No equality constraints ($\mathcal{E} = \emptyset$), so MFCQ is satisfied.

- MFCQ holds since $w = [-1,0]^T$ satisfies strict inequalities for active

Comments

An instructive illustration shows that MFCQ may hold even when LICQ fails. Consider a feasible set in two dimensions defined by two inequality constraints: $c_1(x) = -x_1 \geq 0$, and $c_2(x) = -x_1 - x_2^2 \geq 0$. At the point $x^* = (0,0)^T$, both constraints are active. The gradients at this point are identical: both equal to the vector $(-1,0)^T$. Because of this, the set of active gradients is linearly dependent, and hence LICQ is violated. Nevertheless, we can still satisfy MFCQ. By choosing the direction $w = (-1,0)^T$, we compute the gradient transpose times w for both constraints and obtain the value 1, which is strictly positive. Thus, all active inequalities admit a feasible inward direction. Since no equality constraints are present, the remaining conditions of MFCQ are trivially satisfied. This example makes clear that MFCQ is more flexible than LICQ: linear dependence among gradients of active inequalities does not necessarily preclude the existence of a vector that strictly points inward. Geometrically, what matters is not independence but the ability to find a genuine feasible direction that pushes away from the boundary. Therefore, MFCQ provides a robust foundation for analyzing nonlinear constrained problems, ensuring that optimality conditions can still be applied in the cases when LICQ fails.

Constraint qualifications are sufficient but not necessary for $\mathcal{F}(x^*)$ to capture the geometry of Ω near x^*.

Example

Consider the feasible set $\Omega = \{x \in \mathbb{R}^2 \mid x_2 \geq -x_1^2,\ x_2 \leq x_1^2\}$ at $x^* = (0, 0)^T$.

- No constraint qualifications (e.g., LICQ, MFCQ) are satisfied.
- Linear approximation: $\mathcal{F}(x^*) = \{(w_1, 0)^T \mid w_1 \in \mathbb{R}\}$.
- This accurately reflects the geometry of Ω near x^*.

- Geometric optimality condition for:

$$\min f(x) \quad \text{subject to } x \in \Omega.$$

Depends only on the geometry of Ω, not its algebraic description.

Definition: Normal Cone

The normal cone to Ω at $x \in \Omega$ is:

$$\mathcal{N}_\Omega(x) = \{v \mid v^T w \leq 0 \text{ for all } w \in \mathcal{T}_\Omega(x)\},$$

where $\mathcal{T}_\Omega(x)$ is the tangent cone. Each $v \in \mathcal{N}_\Omega(x)$ is a normal vector.

Comments

When we talk about optimality under constraints, an important subtlety arises. The conditions we often impose, such as the Linear Independence Constraint Qualification or the Mangasarian–Fromovitz Constraint Qualification, are designed to guarantee that the tangent cone captures the geometry of the feasible region near a point. But in fact, these conditions are not always necessary. Sometimes, even when no constraint qualification is satisfied, the tangent cone still gives an accurate description of the local geometry.

Consider the example of the feasible region defined by the inequalities $x_2 \geq -x_1^2$ and $x_2 \leq x_1^2$. At the origin, both constraints “touch” in such a way that no standard qualification holds. Yet, if we approximate the feasible set linearly, we obtain the set of all vectors of the form $(w_1, 0)^T$. This simple one-dimensional subspace perfectly matches the geometry of the feasible set at that point. This teaches us that the essential feature is not whether the algebraic conditions are satisfied, but whether the tangent cone aligns with the true local geometry.

This brings us naturally to the notion of the normal cone. Just as the tangent cone describes directions that remain feasible, the normal cone describes directions that are “orthogonal” in a generalized sense. Formally, a vector belongs to the normal cone at a point if its inner product with every tangent direction is nonpositive. Intuitively, we can think of these vectors as those that “push against” the feasible set, restricting motion. They represent generalized outward normals, even in nonsmooth or nonconvex settings.

This geometric perspective shifts our attention away from the algebraic form of constraints toward the underlying shape of the feasible region. Optimality conditions then become conditions on the relationship between the gradient of the objective and the normal cone, rather than on the specific constraint equations.

A geometric first-order condition relies on the normal cone $\mathcal{N}_\Omega(x^*)$ to characterize local minimizers.

Theorem 25: Geometric Necessary Condition

Suppose x^* is a local minimizer of f in Ω. Then:

$$-\nabla f(x^*) \in \mathcal{N}_\Omega(x^*).$$

Proof:

- For any $d \in \mathcal{T}_\Omega(x^*)$, there exist $\{t_k\}$, $\{z_k\}$ (by the definition) such that:
$$z_k \in \Omega, \quad z_k = x^* + t_k d + o(t_k).$$
- Since x^* is a local minimizer, $f(z_k) \geq f(x^*)$ for all k sufficiently large.
- By Taylor's theorem, with f continuously differentiable:
$$f(z_k) - f(x^*) = t_k \nabla f(x^*)^T d + o(t_k) \geq 0.$$
- Divide by t_k and taking limits as $k \to \infty$, we have:
$$\nabla f(x^*)^T d \geq 0.$$
- Thus, $-\nabla f(x^*)^T d \leq 0$ for all $d \in \mathcal{T}_\Omega(x^*)$.
- By the normal cone definition, $-\nabla f(x^*) \in \mathcal{N}_\Omega(x^*)$. □

Comments

Once we have the normal cone, we can state the fundamental geometric necessary condition for constrained optimization. Suppose we are given a differentiable objective function and a feasible region. If a point is a local minimizer of the function over the feasible set, then its gradient must interact with the geometry of the set in a very specific way. Concretely, the negative gradient at the minimizer must belong to the normal cone of the feasible region at that point. To understand this condition, recall that the gradient of the function indicates the steepest ascent direction. Its negative, therefore, points toward steepest descent. For the point to be a minimizer subject to constraints, there can be no feasible descent direction available. This is precisely captured by the inclusion of the negative gradient in the normal cone: it ensures that the inner product between the gradient and any tangent direction is nonnegative.

The proof of this condition relies on the definition of the tangent cone. If we take any feasible tangent direction, we can approximate feasible points along that direction by constructing sequences that remain inside the feasible set. Because the candidate point is assumed to be a minimizer, the function values at those nearby feasible points cannot fall below the value at the minimizer. By Taylor expansion, the first-order term governing this change is the step size multiplied by the dot product of the gradient with the tangent direction. Since the difference in function values is nonnegative, this dot product must also be nonnegative for all feasible directions. Rearranging the inequality, we obtain that the negative gradient produces nonpositive dot products with every tangent direction. This is exactly the definition of membership in the normal cone. Thus, the geometric condition concisely encodes the idea that at a constrained minimizer, descent directions are blocked by the geometry of the feasible region.

Under LICQ, the normal cone $\mathcal{N}_\Omega(x^*)$ is directly related to the cone of active constraint gradients.

Lemma 9: Normal Cone under LICQ

Suppose the LICQ holds at x^*. Then the normal cone is:
$$\mathcal{N}_\Omega(x^*) = -\mathcal{N},$$
where $\mathcal{N}$ is the cone of active constraint gradients.

Proof:

- From Farkas' Lemma, for $g \in \mathcal{N}$:
$$g^T d \geq 0 \text{ for all } d \in \mathcal{F}(x^*).$$
- Since LICQ holds, $\mathcal{F}(x^*) = \mathcal{T}_\Omega(x^*)$ (by Lemma 5).
- Thus, for $g \in -\mathcal{N}$:
$$g^T d \leq 0 \text{ for all } d \in \mathcal{T}_\Omega(x^*).$$
- By the normal cone definition, $\mathcal{N}_\Omega(x^*) = \{v \mid v^T d \leq 0 \text{ for all } d \in \mathcal{T}_\Omega(x^*)\}$.
- Hence, $\mathcal{N}_\Omega(x^*) = -\mathcal{N}$. □

Comments

The concept of the normal cone becomes especially powerful when combined with constraint qualifications. Under the Linear Independence Constraint Qualification, or LICQ, the geometry of the feasible set aligns cleanly with the gradients of the active constraints. In this setting, the normal cone at a point can be described directly in terms of those gradients. Specifically, it is equal to the negative of the cone generated by the active constraint gradients. Intuitively, each active constraint has a gradient vector pointing into the feasible region. Taking nonnegative combinations of these gradients describes all directions that push against the feasible set. By including the negative sign, we obtain the set of vectors that oppose feasible tangent directions, which is exactly what the normal cone represents.

The formal justification uses Farkas' Lemma. This result connects feasibility of linear inequalities with the existence of certain nonnegative multipliers. Applying it here, one shows that if a vector belongs to the cone generated by active gradients, then its dot product with every feasible direction is nonnegative. Under LICQ, the linearized feasible cone coincides with the tangent cone, so this reasoning extends precisely to the geometry of the set. Reversing the sign, the negative of that cone characterizes all vectors that satisfy the defining inequality of the normal cone. The result is both elegant and practical: instead of reasoning directly with tangent cones, which can be abstract, one can rely on a concrete description in terms of constraint gradients.

Lagrange Multipliers: Sensitivity Analysis

Lagrange multipliers λ_i^* measure the sensitivity of $f(x^*)$ to perturbations in constraint c_i.

- For inactive constraints ($c_i(x^*) > 0$, $i \notin \mathcal{A}(x^*)$):
  - x^* and $f(x^*)$ are unaffected by small perturbations.
  - $\lambda_i^* = 0$ (KKT conditions) indicates no significance.
- For active constraint $i \in \mathcal{A}(x^*)$, perturb $c_i(x) \geq 0$ to $c_i(x) \geq -\epsilon\|\nabla c_i(x^*)\|$:
$$-\epsilon\|\nabla c_i(x^*)\| \approx (x^*(\epsilon) - x^*)^T \nabla c_i(x^*),$$
$$0 \approx (x^*(\epsilon) - x^*)^T \nabla c_j(x^*), \quad j \in \mathcal{A}(x^*),\ \forall\, j \neq i.$$
- Objective change:
$$f(x^*(\epsilon)) - f(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla f(x^*) =$$
$$= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\epsilon) - x^*)^T \nabla c_j(x^*) \approx -\epsilon\|\nabla c_i(x^*)\|\lambda_i^*.$$
- Sensitivity:
$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_i^*\|\nabla c_i(x^*)\|.$$
- Large $\lambda_i^*\|\nabla c_i(x^*)\|$: High sensitivity to c_i.
- $\lambda_i^* = 0$ for active constraint: negligible first-order effect.

Comments

Having established the link between the normal cone and active constraint gradients, we now turn to the interpretation of Lagrange multipliers. Beyond their role in optimality conditions, multipliers have a powerful sensitivity interpretation. Each multiplier associated with an active constraint measures how strongly the objective value responds to perturbations of that constraint. If a constraint is inactive at the solution, then loosening or tightening it slightly has no effect on the optimal point or its function value. In this case, the multiplier is zero, which aligns with the Karush–Kuhn–Tucker conditions.

For an active constraint, the situation is more subtle. Imagine perturbing the constraint by shifting its right-hand side outward by a small parameter ϵ, scaled by the norm of its gradient. This perturbation moves the feasible boundary slightly. The corresponding optimal point shifts accordingly, and the first-order change in the objective function can be computed. It turns out that this change is approximately $-\epsilon$ times the multiplier associated with that constraint, multiplied by the gradient's norm. Thus, the multiplier quantifies how sensitive the optimal value is to small relaxations of the constraint. A large multiplier signals that even a tiny relaxation of the constraint yields a significant improvement in the objective. Conversely, a zero multiplier for an active constraint indicates that relaxing it has negligible effect at first order.

This interpretation is crucial in applications. In economics, multipliers are often called shadow prices, because they measure the marginal value of resources represented by constraints. In engineering, they indicate which constraints are truly binding and which can be relaxed without much impact. In both settings, sensitivity analysis guided by multipliers provides insight far beyond the feasibility of optimality conditions—it highlights the practical importance of individual constraints.

Lagrange multipliers distinguish strongly and weakly active constraints and maintain sensitivity under constraint scaling.

Definition: Strongly and Weakly Active Constraints

For a solution x^* of $\min f(x)$ s.t. $c_i(x) = 0$, $i \in \mathcal{E}$, $c_i(x) \geq 0$, $i \in \mathcal{I}$, with KKT conditions satisfied we say that an inequality constraint c_i is

- *strongly active* (or *binding*) if $i \in \mathcal{A}(x^*)$, $\lambda_i^* > 0$ for some λ^* satisfying KKT conditions;
- *weakly active*: $i \in \mathcal{A}(x^*)$, $\lambda_i^* = 0$ for all λ^* satisfying KKT conditions.

- Sensitivity analysis is invariant to constraint scaling:
  - Replace c_i by $10c_i$: Same feasible set, $\lambda_i^* \to \lambda_i^*/10$, $\|\nabla c_i(x^*)\| \to 10\|\nabla c_i(x^*)\|$.
  - Product $\lambda_i^*\|\nabla c_i(x^*)\|$ unchanged.
- Scaling objective f by 10f:
  - $\lambda_i^* \to 10\lambda_i^*$, increasing sensitivity $\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_i^*\|\nabla c_i(x^*)\|$ by 10.

Comments

When we examine inequality constraints in optimization, it is important to distinguish between those that truly shape the optimal solution and those that, although satisfied with equality, exert no real "force" on the problem. This leads us to the notions of strongly and weakly active constraints. A constraint is called strongly active (or binding) if, at the optimal point, its corresponding Lagrange multiplier is strictly positive. This means that the constraint plays a crucial role: relaxing it would improve the objective, while tightening it would make the problem infeasible. In contrast, a constraint is weakly active if its multiplier is zero for all multiplier choices that satisfy the Karush–Kuhn–Tucker conditions. In such a case, the constraint touches the solution without actually influencing it. An important property here is the invariance of sensitivity analysis under scaling of constraints. If we multiply a constraint function by a constant factor, the feasible region remains unchanged, and the multiplier rescales accordingly, while the product of the multiplier and the gradient norm stays the same. This shows that the dual information is well-behaved and does not depend on arbitrary choices of constraint scaling. On the other hand, scaling the objective function by a factor directly scales the multipliers and therefore alters sensitivity results. This distinction highlights that multipliers reflect relative importance with respect to the objective, not absolute magnitudes of the constraints themselves.

Duality theory constructs an alternative problem to provide insights, bounds, or algorithms for the original problem.

- Duality motivates algorithms (e.g., augmented Lagrangian) and applies to convex and discrete optimization.
- Dual problem may be easier to solve or provide a lower bound on the primal objective.

Consider the problem with no equality constraints, convex f, and $-c_i$:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad c_i(x) \geq 0, \quad i = 1, 2, \ldots, m.$$

Constraint vector:

$$c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \\ \vdots \\ c_m(x) \end{bmatrix}.$$

Simplified form:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad c(x) \geq 0.$$

Comments

The analysis of constraint qualifications naturally leads us toward a deeper structural concept in optimization — the theory of duality. Duality theory is one of the central ideas in optimization. The main idea is that instead of solving the original, or primal, problem directly, one can formulate a related but different problem called the dual problem. The dual problem provides valuable insights, such as lower bounds on the primal objective, conditions for optimality, and in some cases computational simplification. In convex optimization, the duality framework is especially powerful, since strong theoretical guarantees often hold. For instance, under certain regularity conditions, the optimal values of the primal and dual problems coincide. But duality is not limited to convex optimization: ideas of dual formulations appear even in discrete and combinatorial problems, where they serve as a basis for relaxation techniques and approximation methods. A key motivation for developing duality theory comes from algorithms. Many modern optimization methods, such as the augmented Lagrangian method or interior-point methods, make essential use of dual variables and dual functions. To illustrate the setup, let us focus on inequality-constrained convex problems. We can write such a problem in compact form by collecting all constraints into a single vector inequality. This representation simplifies notation and sets the stage for constructing the Lagrangian, which will serve as the bridge between the primal and dual perspectives. In doing so, we begin to see how constraints can be encoded into the objective function with multipliers, thereby opening a path toward the dual problem.

The dual problem maximizes a convex Lagrangian-based objective, often simplifying computation.

- Lagrangian for $\lambda \in \mathbb{R}^m$:

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x).$$

- Dual objective:

$$q(\lambda) = \inf_x \mathcal{L}(x, \lambda).$$

- Computing $\inf_x \mathcal{L}(x, \lambda)$ is often difficult.
- If f, $-c_i$ convex and $\lambda \geq 0$, $\mathcal{L}(\cdot, \lambda)$ is convex, so local minimizers are global, making $q(\lambda)$ practical.
- Domain of q:

$$\mathcal{D} = \{\lambda \mid q(\lambda) > -\infty\}.$$

- Dual problem:

$$\max_{\lambda \in \mathbb{R}^m} q(\lambda) \quad \text{s.t.} \quad \lambda \geq 0.$$

Comments

The core tool for moving from the primal problem to the dual is the Lagrangian function. By combining the objective function with the constraints weighted by nonnegative multipliers, we construct a function that encodes both feasibility and optimality information. For a given multiplier vector, minimizing the Lagrangian over the decision variables defines the dual function. This dual function gives, for each multiplier choice, a bound on the primal objective value. Because it is formed as an infimum of affine functions in the multipliers, the dual function is always concave, even when the primal problem is not convex. This universal concavity makes the dual problem a maximization problem: we seek the best lower bound on the primal by maximizing the dual function over nonnegative multipliers. An essential practical consideration is the domain of the dual function. If the Lagrangian can be driven to negative infinity, then the corresponding multipliers are not meaningful and must be excluded. Therefore, the effective domain of the dual function is defined by all multipliers for which the infimum is finite. When both the objective and constraints are convex, minimizing the Lagrangian with respect to the primal variables is tractable, and the dual function can often be evaluated effectively. The resulting dual problem then provides a powerful tool: it is convex by construction, and its solution supplies both a bound on the primal and insights into the role of constraints through the values of the multipliers.

The dual problem simplifies optimization by maximizing a derived objective.

Example

Consider the problem:

$$\min_{x_1,x_2} 0.5(x_1^2 + x_2^2) \quad \text{s.t.} \quad x_1 - 1 \geq 0.$$

- Lagrangian: $\mathcal{L}(x_1, x_2, \lambda_1) = 0.5(x_1^2 + x_2^2) - \lambda_1(x_1 - 1)$.
- Convex in (x_1, x_2), so infimum at $\partial\mathcal{L}/\partial x_1 = 0$, $\partial\mathcal{L}/\partial x_2 = 0$:

$$x_1 - \lambda_1 = 0, \quad x_2 = 0.$$

- Dual objective: $q(\lambda_1) = -0.5\lambda_1^2 + \lambda_1$.
- Dual problem:

$$\max_{\lambda_1 \geq 0} -0.5\lambda_1^2 + \lambda_1.$$

- Solution: $\lambda_1 = 1$.

Comments

To clarify the ideas considered, let us examine a simple quadratic programming example with one inequality constraint. The primal problem asks us to minimize a quadratic objective subject to a linear inequality. The Lagrangian combines the quadratic term with the constraint, scaled by a nonnegative multiplier. Because the quadratic objective is convex and the constraint linear, the Lagrangian is convex in the primal variables, making the search for its infimum straightforward. By setting the derivatives with respect to the primal variables equal to zero, we obtain explicit expressions for the minimizers as functions of the multiplier. Substituting these back into the Lagrangian yields the dual function, which in this case is a concave quadratic in the multiplier. The dual problem then becomes a one-dimensional concave maximization subject to nonnegativity. Solving it produces an explicit optimal multiplier. This example is illuminating because it demonstrates all the key steps of duality in a transparent way: construction of the Lagrangian, computation of the dual function, and formulation of the dual maximization problem. Moreover, it highlights the power of duality: although the primal is already simple, the dual provides an alternative lens and shows how constraints influence the solution through multipliers. In more complex problems, this same process allows us to transform a difficult constrained minimization into a potentially simpler maximization, often revealing structural properties that are otherwise hidden in the primal form.

The dual objective q is concave, and its optimal value bounds the primal objective.

Theorem 26: Concavity of Dual Objective

The dual objective $q(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \inf_x [f(x) - \lambda^T c(x)]$ is concave, and its domain $\mathcal{D} = \{\lambda \mid q(\lambda) > -\infty\}$ is convex.

Proof: For any λ^0, λ^1 in $\mathbb{R}^m$, any $x \in \mathbb{R}^n$, and any $\alpha \in [0, 1]$, we have:

$$\mathcal{L}(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) = (1-\alpha)\mathcal{L}(x, \lambda^0) + \alpha\mathcal{L}(x, \lambda^1).$$

Taking infimum of both sides in this expression over x:

$$q((1-\alpha)\lambda^0 + \alpha\lambda^1) \geq (1-\alpha)q(\lambda^0) + \alpha q(\lambda^1).$$

Confirms q is concave. If both λ^0 and λ^1 belong to $\mathcal{D}$, this inequality implies that $q((1-\alpha)\lambda^0 + \alpha\lambda^1) \geq -\infty$ also, and therefore $(1-\alpha)\lambda^0 + \alpha\lambda^1 \in \mathcal{D}$, verifying convexity of $\mathcal{D}$. □

Comments

An essential property of the dual function is its concavity. To understand why this holds, recall that the dual function is defined as the infimum over the decision variables of the Lagrangian. Explicitly, the dual objective, written as $q(\lambda) = \inf_x[f(x) - \lambda^T c(x)]$, takes the lowest possible value of the Lagrangian for a fixed multiplier vector. Because the Lagrangian is affine in the multiplier variable, the resulting function is concave in λ. The proof uses a simple convex combination argument: if we take any two multipliers, call them λ^0 and λ^1, and combine them with a weight α between zero and one, then the value of the Lagrangian at this mixture is the same mixture of the two original Lagrangians. When we then take the infimum over all x, the inequality reverses into the definition of concavity. This establishes that q is concave. Furthermore, the effective domain of q, meaning the set of multipliers for which the infimum is not negative infinity, is a convex set. This is because the infimum respects convex combinations in exactly the same way. Concavity of the dual is a universal result: it does not require convexity of the primal problem, and it guarantees that the dual problem always takes the form of a maximization over a concave function. This structural fact is one of the reasons duality theory is so robust. Even in problems where the primal objective is nonconvex and difficult to analyze, the dual problem inherits a convex structure automatically. Therefore, the dual provides not only theoretical insights but also computational leverage, since optimization of a concave function is well-posed and tractable.

Theorem 27: Weak Duality

For any feasible $\bar{x}$ in $\min_x f(x)$ s.t. $c(x) \geq 0$ and $\bar{\lambda} \geq 0$, $q(\bar{\lambda}) \leq f(\bar{x})$.

Proof: $q(\bar{\lambda}) = \inf_x [f(x) - \bar{\lambda}^T c(x)] \leq f(\bar{x}) - \bar{\lambda}^T c(\bar{x})$ where the final inequality follows from $\bar{\lambda} \geq 0$ and $c(\bar{x}) \geq 0$. □

- KKT conditions for the primal problem:

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0, \tag{9a}$$

$$c(\bar{x}) \geq 0, \tag{9b}$$

$$\bar{\lambda} \geq 0, \tag{9c}$$

$$\bar{\lambda}_i c_i(\bar{x}) = 0, \quad i = 1, 2, \ldots, m, \tag{9d}$$

where $\nabla c(x)$ is the $n \times m$ matrix defined by $\nabla c(x) = [\nabla c_1(x), \ldots, \nabla c_m(x)]$.

Theorem 28: Solutions of the Dual Problem

Suppose that $\bar{x}$ is a solution of $\min_{x \in \mathbb{R}^n} f(x) \quad$ s.t. $c(x) \geq 0$ and that f and $-c_i$, $i = 1, 2, \ldots, m$ are convex functions on $\mathbb{R}^n$ that are differentiable at $\bar{x}$. Then any $\bar{\lambda}$ for which $(\bar{x}, \bar{\lambda})$ satisfies the KKT conditions (9) is a solution of the dual problem $\max_{\lambda \in \mathbb{R}^m} q(\lambda)$ s.t. $\lambda \geq 0$.

Comments

The next central result is the principle of weak duality. According to this principle for any feasible point of the primal problem and any nonnegative multiplier vector, the value of the dual function does not exceed the value of the primal objective. The proof is straightforward: substituting $\bar{x}$ into the definition of q yields $f(\bar{x}) - \bar{\lambda}^T c(\bar{x})$. Because feasibility requires that $c(\bar{x})$ is greater than or equal to zero, and because $\bar{\lambda}$ is also nonnegative, the term $\bar{\lambda}^T c(\bar{x})$ is nonnegative. Thus, $q(\bar{\lambda})$ is bounded above by $f(\bar{x})$. Weak duality has profound consequences: it guarantees that the dual problem always provides a lower bound on the primal objective. This bound may or may not be tight, but it gives us a systematic way to certify optimality. The link between primal and dual solutions is formalized by the Karush–Kuhn–Tucker conditions. These conditions include stationarity, which requires that the gradient of f at $\bar{x}$ equals the gradient of c at $\bar{x}$ multiplied by $\bar{\lambda}$; primal feasibility, which requires $c(\bar{x})$ to be nonnegative; dual feasibility, which requires $\bar{\lambda}$ to be nonnegative; and complementary slackness, which states that $\bar{\lambda}_i$ times $c_i(\bar{x})$ equals zero for every constraint index i. Together these four conditions provide a necessary system that any primal–dual optimal pair must satisfy. When convexity holds, they are also sufficient. This interplay of weak duality and KKT conditions is the foundation of modern optimization, guiding both theoretical understanding and algorithm design.

The Karush–Kuhn–Tucker conditions not only characterize optimal solutions but also reveal the deep link between primal and dual problems. Suppose a pair consisting of $\bar{x}$ and $\bar{\lambda}$ satisfies the KKT system. Then, under convexity assumptions, we can show that the value of the primal objective at $\bar{x}$ equals the value of the dual function at $\bar{\lambda}$. This equality, often called strong duality, means that both problems achieve the same optimal value.

KKT conditions and strict convexity link primal and dual problem solutions.

Proof:

Suppose $(\bar{x}, \bar{\lambda})$ satisfies KKT conditions:

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0, \ c(\bar{x}) \geq 0, \ \bar{\lambda} \geq 0, \ \bar{\lambda}_i c_i(\bar{x}) = 0, \ i = 1, 2, \dots, m.$$

- Since $\bar{\lambda} \geq 0$, $\mathcal{L}(\cdot, \bar{\lambda})$ is convex, so for any x, $\mathcal{L}(x, \bar{\lambda}) \geq \mathcal{L}(\bar{x}, \bar{\lambda}) +$
$$+\nabla_x \mathcal{L}(\bar{x}, \bar{\lambda})^T (x - \bar{x}) = \mathcal{L}(\bar{x}, \bar{\lambda}) + (\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda})^T (x - \bar{x}) = \mathcal{L}(\bar{x}, \bar{\lambda}).$$
- Thus, $q(\bar{\lambda}) = \inf_x \mathcal{L}(x, \bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) = f(\bar{x})$.
- By weak duality (Theorem 27), $q(\lambda) \leq f(\bar{x})$ for all $\lambda \geq 0$, so $q(\bar{\lambda}) = f(\bar{x})$ implies $\bar{\lambda}$ solves $\max\limits_{\lambda \geq 0} q(\lambda)$. □

If f, c_i are continuously differentiable and a constraint qualification such as LICQ holds at $\bar{x}$, then an optimal Lagrange multiplier is guaranteed to exist.

Dual-to-primal: If $\mathcal{L}(\cdot, \bar{\lambda})$ is strictly convex (e.g., f strictly convex or c_i strictly convex with $\bar{\lambda}_i > 0$), dual solutions can yield primal solutions.

Comments

The proof proceeds as follows. Since $\bar{\lambda}$ is nonnegative, the Lagrangian with multiplier $\bar{\lambda}$ is convex in the primal variables. By the definition of convexity, the Lagrangian at any x is greater than or equal to the Lagrangian at $\bar{x}$ plus the gradient term. But the stationarity condition ensures that this gradient term vanishes, leaving us with the inequality that the Lagrangian at any x is at least as large as the Lagrangian at $\bar{x}$. Thus, the infimum over x is achieved at $\bar{x}$, and $q(\bar{\lambda})$ equals the Lagrangian of $\bar{x}$, $\bar{\lambda}$. By complementary slackness, this equals $f(\bar{x})$. Therefore, $q(\bar{\lambda}) = f(\bar{x})$. Weak duality then guarantees that $\bar{\lambda}$ is optimal for the dual. Importantly, if the functions are continuously differentiable and a regularity condition such as the linear independence constraint qualification holds, then an optimal multiplier is guaranteed to exist. Moreover, if the Lagrangian is strictly convex in the primal variables, for example if the objective is strictly convex or if certain constraints with positive multipliers are strictly convex, then the dual solution can be used to recover the primal solution uniquely. This dual-to-primal recovery is central in optimization algorithms, where solving the dual efficiently can directly yield the primal solution.

Theorem 29: The uniqueness of a Solution

Suppose that f and $-c_i$, $i = 1, 2, \dots, m$ are convex and continuously differentiable on $\mathbb{R}^n$. Suppose that $\bar{x}$ is a solution of $\min_x f(x)$ s.t. $c(x) \geq 0$ at which LICQ holds. Suppose that $\hat{\lambda}$ solves the dual problem $\max_{\lambda \geq 0} q(\lambda)$, $\inf_x \mathcal{L}(x, \hat{\lambda})$ and that the infimum in $\inf_x \mathcal{L}(x, \hat{\lambda})$ is attained at $\hat{x}$. Assume further that $\mathcal{L}(\cdot, \hat{\lambda})$ is a strictly convex function. Then $\bar{x} = \hat{x}$ (that is, $\hat{x}$ is the unique solution), and $f(\bar{x}) = \mathcal{L}(\hat{x}, \hat{\lambda})$.

Proof: Assume $\bar{x} \neq \hat{x}$. By Theorem 21 (because of the LICQ assumption), there exists $\bar{\lambda}$ satisfying KKT conditions.

- Theorem 28 and KKT conditions implies $\mathcal{L}(\bar{x}, \bar{\lambda}) = q(\bar{\lambda}) = q(\hat{\lambda}) = \mathcal{L}(\hat{x}, \hat{\lambda})$.
- Since $\hat{x} = \arg\min_x \mathcal{L}(x, \hat{\lambda})$, we have that $\nabla_x \mathcal{L}(\hat{x}, \hat{\lambda}) = 0$. By strict convexity:

$$\mathcal{L}(\bar{x}, \hat{\lambda}) > \mathcal{L}(\hat{x}, \hat{\lambda}) + \nabla_x \mathcal{L}(\hat{x}, \hat{\lambda})^T (\bar{x} - \hat{x}) = \mathcal{L}(\hat{x}, \hat{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}).$$

- Thus, $-\hat{\lambda}^T c(\bar{x}) > -\bar{\lambda}^T c(\bar{x}) = 0$, contradicting $\hat{\lambda} \geq 0$, $c(\bar{x}) \geq 0$. Hence, $\bar{x} = \hat{x}$. □

Comments

A natural question arises: when is the primal solution uniquely determined by the dual? The answer is given by a converse result under conditions of strict convexity. Suppose the primal problem has a solution $\bar{x}$ that satisfies the linear independence constraint qualification, and suppose the dual problem has a solution $\hat{\lambda}$ for which the Lagrangian is strictly convex in the primal variables. Then the point that minimizes the Lagrangian for $\hat{\lambda}$, call it $\hat{x}$, must coincide with the primal solution $\bar{x}$. To prove it, assume the opposite, that $\bar{x}$ and $\hat{x}$ are different. The KKT theorem ensures that there exists a multiplier $\bar{\lambda}$ such that the pair $\bar{x}$ and $\bar{\lambda}$ satisfy the conditions. By dual optimality, we then have that the Lagrangian evaluated at $\bar{x}$, $\bar{\lambda}$ equals the dual value $q(\bar{\lambda})$, which in turn equals $q(\hat{\lambda})$, which equals the Lagrangian at $\hat{x}$, $\hat{\lambda}$. But since $\hat{x}$ is the unique minimizer of the strictly convex Lagrangian for $\hat{\lambda}$, any other point such as $\bar{x}$ must give a strictly larger value. This creates a contradiction, because the equalities we obtained earlier show that both values should be the same. The contradiction arises from the assumption that $\bar{x}$ and $\hat{x}$ differ. Therefore, they must in fact be equal, and uniqueness is established. This result is powerful: it tells us that under convexity and strict convexity assumptions, solving the dual not only provides the optimal value but actually identifies the unique primal solution. Thus, duality does not just bound the problem but can, in favorable cases, completely resolve it.

Theorem 30: Wolfe Dual

If f, $-c_i$ $(i = 1, \dots, m)$ are convex, continuously differentiable on $\mathbb{R}^n$, and $(\bar{x}, \bar{\lambda})$ solves $\min_x f(x)$ s.t. $c(x) \geq 0$ with LICQ, then $(\bar{x}, \bar{\lambda})$ solves

$$\max_{x,\lambda} \mathcal{L}(x, \lambda) \text{ s.t. } \nabla_x \mathcal{L}(x, \lambda) = 0,\ \lambda \geq 0.$$

Proof: KKT conditions give $\nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) = 0$, $\lambda \geq 0$, $\mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x})$.

- For any (x, λ) with $\nabla_x \mathcal{L}(x, \lambda) = 0$, $\lambda \geq 0$:

$$\mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) \geq f(\bar{x}) - \lambda^T c(\bar{x}) = \mathcal{L}(\bar{x}, \lambda).$$

- Convexity of $\mathcal{L}(\cdot, \lambda)$ implies:

$$\mathcal{L}(\bar{x}, \lambda) \geq \mathcal{L}(x, \lambda) + \nabla_x \mathcal{L}(x, \lambda)^T (\bar{x} - x) = \mathcal{L}(x, \lambda).$$

- Thus, $\mathcal{L}(\bar{x}, \bar{\lambda}) \geq \mathcal{L}(x, \lambda)$, so $(\bar{x}, \bar{\lambda})$ solves the Wolfe dual. □

Example: For $\min_{x_1,x_2} 0.5(x_1^2 + x_2^2)$ s.t. $x_1 - 1 \geq 0$, at $(x_1, x_2, \lambda_1) = (1, 0, 1)$, $\nabla_x \mathcal{L} = (x_1 - 1, x_2) = 0$, $\mathcal{L} = 0.5$. For $\nabla_x \mathcal{L} = (x_1 - \lambda_1, x_2) = 0$, $\lambda_1 \geq 0$, $\mathcal{L} = -\lambda_1^2/2 + \lambda_1 \leq 0.5$.

Comments

The Wolfe dual formulation provides a powerful way to connect the primal problem with its dual counterpart by reformulating the optimization task in terms of the Lagrangian function. When the objective function and the inequality constraints are convex and differentiable, and the Linear Independence Constraint Qualification is satisfied, the KKT conditions not only characterize optimality in the primal problem but also guarantee equivalence with a specific maximization problem involving the Lagrangian. This is known as the Wolfe dual.

The essence of the Wolfe dual lies in requiring two conditions simultaneously: stationarity of the Lagrangian with respect to the primal variables, and nonnegativity of the multipliers. Stationarity enforces that the primal solution is "balanced" with respect to the constraints, while the multipliers represent the strength of each inequality constraint at the solution. Together, they ensure that the pair of primal and dual solutions mutually reinforce one another.

Convexity plays a central role here. Because the Lagrangian is convex in the primal variable, we can compare the value at the candidate solution with any other feasible point. The inequalities show that no other feasible primal-dual pair achieves a higher value of the Lagrangian, thereby confirming that the solution is indeed optimal for the dual problem. This symmetry between minimization of the primal objective and maximization of the Lagrangian demonstrates the unifying role of duality theory.

A simple quadratic example illustrates the idea: minimizing a convex quadratic function subject to a linear inequality constraint leads to a dual problem in which the multiplier's optimization reproduces the same optimal value. This provides intuition for why dual problems can sometimes be easier to analyze than their primal counterparts.

Linear programming duals simplify computations in certain cases.

Example: Linear Programming

Primal problem:

$$\min_x c^T x \quad \text{s.t.} \quad Ax - b \geq 0.$$

Dual objective: $q(\lambda) = \inf_x [c^T x - \lambda^T (Ax - b)] = \inf_x [(c - A^T \lambda)^T x + b^T \lambda]$.

- If $c - A^T \lambda \neq 0$, $q(\lambda) = -\infty$.
- If $c - A^T \lambda = 0$, $q(\lambda) = b^T \lambda$.
- Dual problem: $\max_\lambda b^T \lambda$ s.t. $A^T \lambda = c$, $\lambda \geq 0$.

Wolfe dual: $\max_{x,\lambda} c^T x - \lambda^T (Ax - b)$ s.t. $A^T \lambda = c$, $\lambda \geq 0$, reduces to the dual problem.

For some matrices A, the dual problem is computationally easier to solve than the primal.

Comments

Linear programming provides one of the most direct illustrations of duality. In its standard form, the primal problem is to minimize a linear cost subject to linear inequalities. The dual problem emerges naturally from the Lagrangian relaxation, where multipliers are introduced to enforce the inequalities. By carefully analyzing the infimum of the Lagrangian over the primal variables, we can distinguish two possibilities: if the linearity conditions are not satisfied, the Lagrangian can be driven to negative infinity; but if the compatibility condition holds, the infimum collapses into a finite linear function in the multipliers.

This reasoning gives the dual problem: maximize a linear functional of the multipliers subject to an equality constraint that reflects the compatibility condition, and nonnegativity constraints on the multipliers. Importantly, the structure of the dual problem mirrors that of the primal: the objective is linear, and the constraints are also linear. In fact, the dual of a linear program is another linear program.

One striking feature of linear programming duality is computational efficiency. For certain structures of the constraint matrix, solving the dual is substantially easier than solving the primal. For example, when the number of constraints is small compared to the number of variables, the dual may have fewer variables and hence be more efficient to optimize. Conversely, in other situations, the primal may be easier. Thus, duality not only provides theoretical insights into optimality and sensitivity analysis but also has direct implications for numerical computation.

Note, that in this case the Wolfe dual formulation applied to linear programs recovers exactly the standard dual problem.

Quadratic programming duals leverage convexity for explicit solutions.

Example: Convex Quadratic Programming

Primal problem (G symmetric, positive definite):

$$\min_x \frac{1}{2}x^T G x + c^T x \quad \text{s.t.} \quad Ax - b \geq 0.$$

Dual objective: $q(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \inf_x \left[\frac{1}{2}x^T G x + c^T x - \lambda^T(Ax - b)\right]$.

- Since G is positive definite, $\mathcal{L}(\cdot, \lambda)$ is strictly convex.
- Infimum at $\nabla_x \mathcal{L} = Gx + c - A^T\lambda = 0$.
- Substituting $x = G^{-1}(A^T\lambda - c)$ gives:

$$q(\lambda) = -\frac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + b^T\lambda \quad \text{s.t.} \quad \lambda \geq 0$$

Comments

Quadratic programming extends linear programming by allowing a quadratic objective while keeping linear inequality constraints. When the quadratic term is defined by a symmetric positive definite matrix, the primal problem remains convex, which ensures a unique minimizer. Constructing the dual proceeds by evaluating the infimum of the Lagrangian with respect to the primal variables. Because the quadratic form is strictly convex, this infimum exists at a unique point determined by the stationarity condition.

Specifically, setting the gradient of the Lagrangian equal to zero gives an explicit expression for the primal variable in terms of the multipliers. Substituting this expression back into the Lagrangian produces a closed-form dual objective. The resulting dual problem has a concave quadratic objective in the multipliers, along with nonnegativity constraints. Thus, the dual transforms a constrained quadratic minimization into an unconstrained quadratic maximization, up to the inequality conditions on the multipliers.

This reformulation highlights an important computational advantage: sometimes optimizing over the dual variables is easier than directly solving the primal. The dual function automatically incorporates the quadratic curvature of the primal, reducing the complexity of the optimization landscape. Moreover, because the dual objective explicitly involves the inverse of the quadratic matrix, it emphasizes how primal curvature shapes dual feasibility.

Convex Quadratic Programming: Wolfe Dual

The Wolfe dual reformulates the quadratic program with explicit constraints.

Example: Convex Quadratic Programming (continued)

Wolfe dual:

$$\max_{x,\lambda} \frac{1}{2}x^T G x + c^T x - \lambda^T (Ax - b) \quad \text{s.t.} \quad Gx + c - A^T\lambda = 0, \quad \lambda \geq 0.$$

To make it clearer that the objective is concave we can rewrite the dual formulation (using $x^T G x = -(c - A^T\lambda)^T x$) as follows:

$$\max_{x,\lambda} -\frac{1}{2}x^T G x + \lambda^T b \quad \text{s.t.} \quad Gx + c - A^T\lambda = 0, \quad \lambda \geq 0.$$

The Wolfe dual requires only positive semidefiniteness of G.

Comments

The Wolfe dual for convex quadratic programming presents an alternative but equivalent way to express the dual problem. Instead of eliminating the primal variable explicitly, we retain both the primal variable and the multipliers, subject to the stationarity condition. This gives a maximization problem with quadratic terms but linear equality constraints linking the variables.

One of the advantages of the Wolfe dual formulation is that it exposes the concavity of the objective function more directly. By rewriting the quadratic expression, we make it clear that the Wolfe dual is a concave maximization problem, which guarantees the existence of an optimal solution provided the feasible region is nonempty. Furthermore, the Wolfe dual requires only that the quadratic matrix be positive semidefinite, not strictly positive definite. This broadens the range of applicability, since many practical problems involve semidefinite matrices.

From a computational perspective, the Wolfe dual can sometimes be solved more efficiently than the standard dual, depending on the problem structure. The explicit presence of both primal and dual variables can be exploited by specialized algorithms, such as interior-point or decomposition methods. This flexibility makes the Wolfe dual a valuable tool in both theory and practice.

Conceptually, the Wolfe dual reinforces the central theme of duality theory: optimization problems can often be reformulated in multiple equivalent ways, each highlighting different aspects of structure and computational tractability.

Linear programs have a linear objective function and linear constraints (equalities/inequalities). The feasible set is a polytope (convex, connected, with flat polygonal faces).

Note: Solutions may be non-unique (e.g., same $c^T x$ over an edge or face). No solution exists if the feasible set is empty (the infeasible case) or the objective is unbounded below (the unbounded case).

Standard Form of Linear Programs:

$$\min c^T x, \text{ subject to } Ax = b, x \geq 0 \tag{10}$$

where $c, x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and A is an $m \times n$ matrix.

Conversion to Standard Form:

For a problem with inequality constraints:

$$\min c^T x, \text{ subject to } Ax \leq b$$

introduce slack variables z to obtain:

$$\min c^T x, \text{ subject to } Ax + z = b, z \geq 0$$

Comments

Up to this point, we have studied general approaches to optimization with constraints, often involving nonlinear functions and more complex feasible sets. Now we turn to a very special but extremely important class of problems: linear programming. These are optimization problems where both the objective function and the constraints are linear. Despite their apparent simplicity, linear programs form the backbone of many applications in economics, engineering, operations research, and computer science.

A linear program seeks to minimize a linear objective, written as $c^T x$, subject to linear equality and inequality constraints. The set of feasible solutions is always a convex polytope. A polytope can be thought of as a geometric object with flat polygonal faces, and this convexity ensures that if two points are feasible, then every point on the line between them is also feasible. This property is crucial because it implies that local optima are always global optima.

However, several issues may arise. The solution may not be unique: for example, an entire edge or face of the polytope might yield the same objective value. In some cases, the problem has no solution at all. This happens either because the feasible set is empty, meaning the constraints are inconsistent, or because the objective is unbounded. An unbounded case occurs when the feasible region extends infinitely in a direction that keeps decreasing the objective value.

To study these problems systematically, it is convenient to use a standard form. In this form, we minimize $c^T x$, subject to $Ax = b$ and $x \geq 0$. Any linear program with inequalities can be converted into this format by introducing auxiliary variables. For instance, inequality constraints can be transformed into equalities by adding slack variables, which are required to remain nonnegative. This process allows us to work within a unified framework for analysis and algorithms.

Variable Splitting:

Split $x = x^+ - x^-$, where $x^+ = \max(x, 0) \geq 0$, $x^- = \max(-x, 0) \geq 0$. Then:

$$\min \begin{bmatrix} c \\ -c \\ 0 \end{bmatrix}^T \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix}, \text{ s.t. } \begin{bmatrix} A & -A & I \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} = b, \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} \geq 0$$

Handling Inequalities:

- $x \leq u \Leftrightarrow x + w = u$, $w \geq 0$ (slack variables).
- $Ax \geq b \Leftrightarrow Ax - y = b$, $y \geq 0$ (surplus variables).

Maximization $\max c^T x$ becomes $\min(-c)^T x$.

Note: The linear program (10) is *infeasible* if the feasible set is empty; *unbounded* if there exists a sequence x^k such that $c^T x^k \to -\infty$.

Comments

Once the standard form has been introduced, the natural question is how to convert an arbitrary linear program into this representation. There are several important steps in this process.

First, we must ensure that all variables are nonnegative. If a variable is unrestricted in sign, we split it into two components: a nonnegative part and a nonnegative counterpart. Specifically, a variable x is written as $x^+ - x^-$, where $x^+ = \max(x, 0)$, and $x^- = \max(-x, 0)$. Both of these are greater than or equal to zero. In effect, this substitution replaces a single unrestricted variable with two nonnegative variables.

Second, inequality constraints must be rewritten as equalities. For example, an upper bound constraint such as $x \leq u$ can be converted by adding a nonnegative slack variable w, so that $x + w = u$. Similarly, a lower bound inequality such as $Ax \geq b$ can be expressed as $Ax - y = b$, where y is a nonnegative surplus variable. These manipulations bring every inequality into an equality format, which matches the requirements of the standard form.

It is also conventional to write maximization problems as minimization problems. For instance, maximizing $c^T x$ is equivalent to minimizing $-c^T x$. This simple transformation ensures that all linear programs can be expressed consistently.

Finally, we must consider the possibility that a program is infeasible or unbounded. If no feasible solution satisfies the constraints, then the problem is infeasible. On the other hand, if there exists a sequence of feasible solutions such that the objective decreases without limit, then the program is unbounded. These two situations represent fundamental difficulties that algorithms must be able to detect.

Lagrangian:

For $\min c^T x$, s.t. $Ax = b$, $x \geq 0$, partition multipliers into $\lambda \in \mathbb{R}^m$ (for the equality constraints $Ax = b$) and $s \in \mathbb{R}^n$ (for the bound constraints $x \geq 0$).

Lagrangian: $\mathcal{L}(x, \lambda, s) = c^T x - \lambda^T (Ax - b) - s^T x$

KKT Conditions:

x^* is a solution if $\exists \lambda, s$ s.t.:

$$A^T \lambda + s = c, \tag{11a}$$

$$Ax = b, \tag{11b}$$

$$x \geq 0, \tag{11c}$$

$$s \geq 0, \tag{11d}$$

$$x_i s_i = 0, \quad i = 1, 2, \ldots, n. \tag{11e}$$

Complementarity : $x_i s_i = 0$ (or $x^T s = 0$) means $x_i = 0$ or $s_i = 0$.

Comments

To analyze optimality in linear programming, we rely on the Karush–Kuhn–Tucker conditions. For a linear program in standard form, the Lagrangian plays a central role. Recall that the objective is to minimize $c^T x$, subject to $Ax = b$ and $x \geq 0$. We introduce multipliers for the constraints. For the equality constraints, we associate a vector of multipliers, usually denoted by λ. For the nonnegativity constraints, we introduce a vector of multipliers s, sometimes called dual variables.

The Lagrangian is then defined as $c^T x - \lambda^T (Ax - b) - s^T x$. The KKT conditions emerge by requiring stationarity, primal feasibility, dual feasibility, and complementarity. More explicitly: first, the gradient condition implies that $A^T \lambda + s = c$. Second, primal feasibility enforces $Ax = b$ and $x \geq 0$. Third, dual feasibility requires $s \geq 0$. Finally, complementarity states that $x_i s_i = 0$ for each index i. In words, for each component, either the variable is zero or the corresponding dual variable is zero, but not both positive simultaneously.

In linear programming, these conditions are not only necessary but also sufficient for optimality. That is, if we can find x, λ, and s satisfying all these equations and inequalities, then x is guaranteed to be a global solution. This strong property is one of the reasons why linear programming is so tractable compared to nonlinear optimization.

Let $\langle x^*, \lambda^*, s^* \rangle$ satisfy KKT conditions:

$$A^T\lambda^* + s^* = c, \quad Ax^* = b, \quad x^* \geq 0, \quad s^* \geq 0, \quad {x^*}^T s^* = 0.$$

Then:

$$c^T x^* = \left(A^T\lambda^* + s^*\right)^T x^* = (Ax^*)^T \lambda^* = b^T\lambda^*,$$

indicating that the primal and dual objectives are equal (since $b^T\lambda$ is the objective function for the dual problem).

- The KKT conditions (11) are sufficient for x^* to be a global solution of (10). Let $\bar{x}$ be any other feasible point (with $A\bar{x} = b$, $\bar{x} \geq 0$):

$$c^T\bar{x} = \left(A^T\lambda^* + s^*\right)^T \bar{x} = b^T\lambda^* + \bar{x}^T s^* \geq b^T\lambda^* = c^T x^*, \text{ since } \bar{x} \geq 0,\, s^* \geq 0.$$

  Thus, no feasible point has a lower objective value than $c^T x^*$.
- $\bar{x}$ is optimal if and only if:

$$\bar{x}^T s^* = 0,$$

  so when $s_i^* > 0$, then $\bar{x}_i = 0$ for all solutions $\bar{x}$.

Comments

An essential feature of linear programming is the deep connection between the primal and dual problems. If we have a solution triple x^*, λ^*, s^* that satisfies the KKT conditions, then strong duality holds: the objective values of the primal and the dual problems are equal. In particular, we can write $c^T x^* = b^T\lambda^*$. This relationship reveals that solving the primal problem automatically provides a solution to the dual, and vice versa.

The equality of objectives allows us to establish optimality. Suppose x^* is feasible and paired with multipliers λ^* and s^* satisfying the KKT system. Then for any other feasible point $\bar{x}$, the objective value $c^T\bar{x}$ can be expressed as $b^T\lambda^* + \bar{x}^T s^*$. Since both $\bar{x}$ and s^* are nonnegative, this expression is always greater than or equal to $b^T\lambda^*$. But that equals $c^T x^*$, the objective at the candidate solution. Therefore, no feasible point can achieve a smaller value. This proves that x^* is indeed optimal.

Moreover, the complementarity condition offers a finer characterization of the solution. If s_i^* is strictly positive, then the corresponding primal variable x_i must be zero in every optimal solution. Conversely, if x_i is strictly positive, then the associated s_i^* must vanish. This interplay provides insight into which constraints are active at optimality and which variables are essential in the solution.

Thus, the primal–dual structure of linear programming not only ensures optimality but also provides powerful tools for interpretation and sensitivity analysis.

Let's consider the problem (10) with the given data c, b, and A. The *dual problem* to the primal problem (10) is:

$$\max b^T\lambda, \text{ subject to } A^T\lambda \leq c. \tag{12}$$

- **Alternative Formulation:** Introduce a vector of dual slack variables s to rewrite the dual problem as:
$$\max b^T\lambda, \quad \text{subject to } A^T\lambda + s = c, \quad s \geq 0.$$
The variables (λ, s) are sometimes collectively called dual variables.
- **Relationship:** The primal and dual problems provide two perspectives on the same data.
- **Standard Form:** To align with standard minimization form, restate the dual problem as:
$$\min -b^T\lambda, \quad \text{subject to } c - A^T\lambda \geq 0.$$
- **Lagrangian:** Using $x \in \mathbb{R}^n$ as Lagrange multipliers for $A^T\lambda \leq c$, the Lagrangian is:
$$\tilde{\mathcal{L}}(\lambda, x) = -b^T\lambda - x^T(c - A^T\lambda).$$

Comments

The construction of the dual problem is one of the most fundamental ideas in linear programming. Given a primal problem with cost vector c, constraint matrix A, and right-hand side b, the dual problem is defined as maximizing $b^T\lambda$, subject to the inequality $A^T\lambda \leq c$. Here, λ is the vector of dual variables, sometimes interpreted as shadow prices, because each component measures the marginal value of relaxing the corresponding primal constraint.

Just as in the case of a primal problem, slack variables are introduced to bring the dual problem to a standard form. Setting $s = c - A^T\lambda$, the dual inequalities are replaced by equalities $A^T\lambda + s = c$, with $s \geq 0$. The pair of variables (λ, s) together represent the complete set of dual unknowns. This form is particularly useful because it mirrors the structure of the primal, where constraints are also written in equality form with nonnegative slack variables.

Another useful reformulation rewrites the dual as a minimization problem: minimize $-b^T\lambda$, subject to $c - A^T\lambda \geq 0$. This way, both primal and dual are cast into consistent minimization frameworks, facilitating comparisons and theoretical developments.

Finally, the Lagrangian perspective offers another layer of interpretation. By treating the primal variables x as Lagrange multipliers for the dual constraints $A^T\lambda \leq c$, we can define the dual Lagrangian as $-b^T\lambda - x^T(c - A^T\lambda)$. This reveals a striking symmetry: primal variables act as multipliers for dual constraints, just as dual variables act as multipliers for primal constraints. The two problems are, in essence, different but equivalent views of the same optimization structure.

First-Order Necessary Conditions for the dual problem

For λ to be optimal for the dual problem $\max b^T\lambda,\ \text{s.t.} A^T\lambda \le c$, there exists x such that:

$$\begin{aligned} Ax &= b, \\ A^T\lambda &\le c, \\ x &\ge 0, \\ x_i(c - A^T\lambda)_i &= 0, \quad i = 1, 2, \ldots, n. \end{aligned}$$

- Defining $s = c - A^T\lambda$ we obtain that these conditions are identical to (11):

$$A^T\lambda + s = c, \quad Ax = b, \quad x \ge 0, \quad s \ge 0, \quad x^Ts = 0.$$

- The optimal λ in the primal problem are the optimal variables in the dual, and the optimal x in the dual are the optimal variables in the primal.
- Sufficiency: For x^*, λ^* satisfying these conditions (with $s = c - A^T\lambda^*$), and any dual feasible $\tilde{\lambda}$ ($A^T\tilde{\lambda} \le c$):

$$b^T\tilde{\lambda} = (x^*)^TA^T\tilde{\lambda} = (x^*)^T(A^T\tilde{\lambda} - c) + c^Tx^* \le c^Tx^* = b^T\lambda^*,$$

since $A^T\tilde{\lambda} - c \le 0$, $x^* \ge 0$. Thus, λ^* is optimal for the dual problem.

Comments

The first-order conditions for the dual problem play a central role in linking primal and dual optimality. Recall that the dual is to maximize $b^T\lambda$ subject to $A^T\lambda \le c$. For a vector λ to be optimal, there must exist a vector x satisfying several conditions simultaneously. These conditions mirror the Karush–Kuhn–Tucker system we have already encountered.

The requirements are: first, the primal feasibility condition $Ax = b$; second, the dual feasibility condition $A^T\lambda \le c$; third, nonnegativity of x; and fourth, complementarity, meaning that for each component i, the product $x_i(c - A^T\lambda)_i$ equals zero. In other words, either the primal variable is strictly positive and the corresponding constraint in the dual is tight, or the dual inequality is slack and the primal variable is zero.

Defining $s = c - A^T\lambda$, the system can be written more symmetrically as $A^T\lambda + s = c$, $Ax = b$, $x \ge 0$, $s \ge 0$, and $x^Ts = 0$. This is exactly the KKT system for linear programming.

Why do these conditions guarantee optimality? Suppose x^* and λ^* satisfy them. For any dual feasible vector $\tilde{\lambda}$, we can compare $b^T\tilde{\lambda}$ with c^Tx^*. Using feasibility and complementarity, it follows that $b^T\tilde{\lambda}$ is less than or equal to $b^T\lambda^*$, which equals c^Tx^*. Hence λ^* is indeed an optimal solution of the dual.

The primal-dual relationship is symmetric: the dual of the dual problem $\max b^T\lambda$, subject to $A^T\lambda \leq c$, is the primal problem.

- Weak Duality: For a primal feasible x ($Ax = b$, $x \geq 0$) and dual feasible (λ, s) ($A^T\lambda + s = c$, $s \geq 0$):

$$c^Tx - b^T\lambda = (c - A^T\lambda)^T x = s^T x \geq 0.$$

Thus, $c^Tx \geq b^T\lambda$, i.e., the dual objective is a lower bound on the primal objective.

Theorem 31: Strong Duality

(i) If either the primal problem ($\min c^Tx$, s. t. $Ax = b, x \geq 0$) or the dual problem ($\max b^T\lambda$, s. t. $A^T\lambda \leq c$), has a finite solution, then so does the other, and their objective values are equal.

(ii) If either problem is unbounded, then the other is infeasible.

Comments

The relationship between primal and dual problems is not merely one-sided but deeply symmetric. In fact, the dual of the dual problem is the primal problem itself. This fundamental symmetry underlies the power of linear programming duality.

The first observation is the principle of weak duality. For any feasible primal solution x satisfying $Ax = b$ and $x \geq 0$, and any feasible dual pair (λ, s) satisfying $A^T\lambda + s = c$ with $s \geq 0$, we have the inequality $c^Tx - b^T\lambda = s^Tx \geq 0$. This shows that the primal objective value is always greater than or equal to the dual objective value. In other words, the dual provides a guaranteed lower bound on the primal.

The deeper result is strong duality. It asserts two things. First, if either the primal minimization problem or the dual maximization problem has a finite optimal solution, then so does the other, and their optimal objective values coincide exactly. Second, if one problem is unbounded, the other must be infeasible. Thus, in the context of linear programming, the primal and dual problems are inseparably linked through the strong duality theorem: in fact, they represent equivalent formulations of the same optimization problem, with their optimal solutions interconnected via the KKT conditions. That's why KKT conditions are necessary and sufficient for optimality of linear program's problems.

This interrelation provides the theoretical basis for efficient algorithms, such as the simplex method and interior-point methods, because checking dual feasibility and complementary slackness is often computationally advantageous. It also gives interpretive power: dual solutions identify shadow prices that quantify the value of constraints, offering direct economic and engineering insights. It also demonstrates the power of formalization, which we discussed in the first lesson.

Proof of Strong Duality

Proof of (i): Assume the primal $\min c^T x$, subject to $Ax = b$, $x \geq 0$, has a finite optimal solution x^*.

- By Theorem 21 (with a caveat, that for LP $T_\Omega(x) = \mathcal{F}(x)$) there exist λ^*, s^* such that $\langle x^*, \lambda^*, s^* \rangle$ satisfies:
$$A^T\lambda^* + s^* = c, \quad Ax^* = b, \quad x^* \geq 0, \quad s^* \geq 0, \quad x^{*T}s^* = 0.$$
- These conditions are equivalent to:
$$Ax^* = b, \quad A^T\lambda^* \leq c, \quad x^* \geq 0, \quad x_i^*(c - A^T\lambda^*)_i = 0, \quad i = 1, \dots, n,$$
which are sufficient for λ^* to solve the dual $\max b^T\lambda$, subject to $A^T\lambda \leq c$.
- Then, $c^T x^* = b^T\lambda^*$. A symmetric argument holds if the dual has a solution.

Proof of (ii): Assume the primal is unbounded, i.e., there exists a sequence x^k such that:
$$c^T x^k \to -\infty, \quad Ax^k = b, \quad x^k \geq 0.$$

- If the dual is feasible, i.e., $\exists \bar{\lambda}$ such that $A^T\bar{\lambda} \leq c$, then:
$$\bar{\lambda}^T b = \bar{\lambda}^T A x^k \leq c^T x^k \to -\infty,$$
leading to a contradiction. Thus, the dual is infeasible.
- Similarly, unboundedness of the dual implies infeasibility of the primal. $\square$

Comments

The proof of strong duality demonstrates beautifully the unity of primal and dual problems. Suppose the primal problem of minimizing $c^T x$, subject to $Ax = b$ and $x \geq 0$, has a finite optimal solution x^*. Then, by the KKT theorem, there exist multipliers λ^* and s^* satisfying the conditions $A^T\lambda^* + s^* = c$, $Ax^* = b$, $x^* \geq 0$, $s^* \geq 0$, and $x^{*T}s^* = 0$.

These conditions can be restated as primal feasibility, dual feasibility, and complementarity, showing that the pair (x^*, λ^*) satisfies both systems simultaneously. This suffices to prove that λ^* is an optimal solution of the dual problem maximizing $b^T\lambda$. Moreover, equality of objectives follows: $c^T x^* = b^T\lambda^*$. By symmetry, if we start with an optimal dual solution, we can construct a corresponding primal solution with the same value.

The second part of the theorem deals with unboundedness. If the primal is unbounded below, there exists a sequence x^k with $c^T x^k$ tending to minus infinity, while satisfying feasibility. If the dual were feasible, then for some feasible multiplier $\bar{\lambda}$, we would have $\bar{\lambda}^T b = \bar{\lambda}^T A x^k$, which is less than or equal to $c^T x^k$. But as the latter tends to minus infinity, this produces a contradiction. Hence, the dual must be infeasible. By symmetry, the same reasoning applies in reverse.

Thus, the proof of strong duality is not only rigorous but also elegantly demonstrates the inseparable balance between primal and dual optimization problems.