

Lecture 7: Constrained Optimization

1 Overview: constraints, tangent directions, and multipliers

Lecture 6 solved an optimization problem in three separate steps: establish that a solution exists, generate candidates from first-order conditions, and certify a candidate as global, by concavity or by comparing values. Its first-order condition at a boundary point was one-sided: at a local maximizer x^* ,

$$\nabla f(x^*) \cdot v \leq 0 \quad \text{for every feasible direction } v \text{ from } x^*, \quad (\text{FOC})$$

where v is a feasible direction if the segment $x^* + tv$ stays in the feasible set for all small $t > 0$.

In economic problems the feasible set is typically described by equations and inequalities. For example, in a consumer problem, the budget constraint $p \cdot x \leq m$ is a linear inequality. In a planner's problem, the requirement $x_1 + x_2 = m$ that a resource be split between two uses is a linear equality. In a producer's problem, the technology constraint $q = F(z)$, linking output q to inputs z , is a nonlinear equality whenever the production function F is nonlinear.

Depending on the shape of the feasible set, the condition (FOC) may be vacuous, i.e., it may rule out nothing. Suppose first that there is just one *equality* constraint $h(x) = 0$. If h is affine, as in the planner's problem, the feasible set is a line: segments along the line stay in the set, so nonzero feasible directions exist and (FOC) is not vacuous (Figure 1.1, left). If instead the feasible set is curved, such as the circle $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 2\}$, it contains no segments: every straight move away from a point of the circle leaves the circle. The only feasible direction is then $v = 0$, so (FOC) holds at every point of the circle and rules out nothing (Figure 1.1, right).



Figure 1.1: Feasible movement under an equality constraint. Left: an affine constraint. The feasible set is the line $h(x) = 0$ with $h(x) = m - x_1 - x_2$; the two blue arrows from x^* are feasible directions, since segments along the line stay in the set. Right: a curved constraint. The feasible set is the circle $h(x) = 0$ with $h(x) = 2 - x_1^2 - x_2^2$; every straight move from x^* leaves the circle. The feasible movement is the curve γ , which bends with the circle; the green vector v is its velocity at x^* , tangent to the circle (the dashed tangent line).

For a curved feasible set, directions of movement come from curves rather than segments. A feasible movement through x^* is a curve $\gamma(t)$ that stays in the feasible set, with $\gamma(0) = x^*$:

think of $\gamma(t)$ as the position of a moving point at time t . Its *velocity* $v = \gamma'(0)$ is the vector of the instantaneous rates of change of its coordinates — the direction in which the moving point passes through x^* . The path bends with the set, but the velocity is an ordinary vector: on the circle it is tangent to the circle, and on the line the curve can be the line itself, whose velocity is one of the old feasible directions.

At points satisfying a regularity condition (LICQ), these velocities — the *tangent directions* — are exactly the vectors orthogonal to every constraint gradient $\nabla h_j(x^*)$, and the first-order condition becomes

$$\nabla f(x^*) + \mu_1^* \nabla h_1(x^*) + \cdots + \mu_k^* \nabla h_k(x^*) = 0$$

for some coefficients μ_j^* , which are the *Lagrange multipliers*. The function $L(x, \mu) = f(x) + \mu \cdot h(x)$ is the *Lagrangian*; its critical points are precisely the pairs (x, μ) that satisfy this equation together with $h(x) = 0$.

Section 3 then adds inequality constraints, written $g_i(x) \geq 0$. At a candidate x^* , a constraint is *active* if $g_i(x^*) = 0$ and *slack* if $g_i(x^*) > 0$; a slack constraint does not restrict small movements, so it drops out of the first-order condition. Stationarity, feasibility, nonnegative multipliers on the inequalities, and the rule that each multiplier or its constraint is zero (*complementary slackness*) together form the *Karush–Kuhn–Tucker (KKT) conditions*.

Section 4 then provides the global certification, specialized to the cases of Sections 2 and 3. As in Lecture 6, existence must be shown separately, using Weierstrass, coercivity, or a direct argument.

2 Equality constraints and the Lagrangian

2.1 Tangent directions and constraint normals

Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \rightarrow \mathbb{R}$ and $h : U \rightarrow \mathbb{R}^k$ be C^1 , and consider

$$\max_{x \in U} f(x) \quad \text{subject to} \quad h(x) = 0.$$

Here, the feasible set $\{x \in U : h(x) = 0\}$ is described by a system of k equations. When one of these equations is nonlinear, the feasible set may be curved. We thus work with feasible curves and their velocities.

We call γ a *feasible curve through* x^* if γ is differentiable, $\gamma(0) = x^*$, and $h(\gamma(t)) = 0$ for all t near 0. Its velocity at x^* is $v = \gamma'(0)$. Such a velocity is called a *tangent direction* to the constraint set at x^* .

Lemma 2.1 (Tangent directions lie in the null space). *Let γ be a feasible curve through x^* with velocity v . Then*

$$Dh(x^*)v = 0.$$

Proof. $h(\gamma(t)) = 0$ for all t near 0, and differentiating using the chain rule from Lecture 4 gives

$$Dh(\gamma(t))\gamma'(t) = 0 \quad \text{at } t = 0, \quad \text{so} \quad Dh(x^*)v = 0.$$

□

Every tangent direction thus lies in the null space $N(Dh(x^*))$ from Lecture 3. Row by row, the display says

$$\nabla h_j(x^*) \cdot v = 0 \quad \text{for each } j.$$

In other words, each constraint gradient is *orthogonal* (perpendicular) to every tangent direction. A vector perpendicular to the tangent directions is called a *normal* to the constraint set. Thus the constraint gradients are normals to the constraint set.

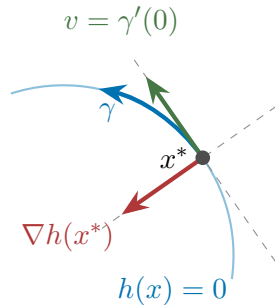


Figure 2.1: The local geometry at x^* for one constraint ($k = 1$). The feasible curve γ stays on the constraint set. Its velocity $v = \gamma'(0)$ is tangent to the constraint set, while $\nabla h(x^*)$ is normal to it; hence $\nabla h(x^*) \cdot v = 0$. The dashed lines are the tangent and normal lines at x^* .

The converse is generally not true. That is, if $Dh(x^*)v = 0$ for some vector v , there may be no feasible curve through x^* with velocity v . For example, if $h(x) = x^2$, then the feasible set is the single point $\{0\}$, so the only feasible velocity at 0 is 0. Yet, the null space of $h'(0) = 0$ is all of $\mathbb{R}$, so any vector $v \in \mathbb{R}$ satisfies $Dh(0)v = 0$.

The following condition ensures that all vectors in the null space of $Dh(x^*)$ are velocities of feasible curves through x^* .

Definition 2.2 (LICQ for equality constraints). The *linear independence constraint qualification* holds at a feasible point x^* if the constraint gradients $\nabla h_1(x^*), \dots, \nabla h_k(x^*)$ are linearly independent; equivalently, if $\text{rank } Dh(x^*) = k$.

Remark 2.3 (How many constraints?). The number k of equations need not equal the number of independent restrictions they place near x^* ; that number is $r = \text{rank } Dh(x^*)$. When the constraints have a locally equivalent description by r independent equations, the feasible set is locally a surface of dimension $n - r$. Thus $r < n$ leaves feasible movement, while $r = n$ makes the feasible point locally isolated. LICQ says $r = k$, so it requires $k \leq n$: more than n vectors in $\mathbb{R}^n$ cannot be linearly independent. When $k > n$, no description with that many equations satisfies LICQ; when possible, re-describe the same feasible set by independent equations, and otherwise analyze it directly.

Lemma 2.4 (Tangent directions under LICQ). *Let h be C^1 and let LICQ hold at the feasible point x^* . If $Dh(x^*)v = 0$, then v is the velocity of a differentiable feasible curve through x^* .*

Intuitively, if a vector v is orthogonal to the constraint normals, then it is tangent to the constraint set, and a curve can be drawn along the constraint set in the direction of v . LICQ thus ensures that the constraints form a smooth surface locally around x^* , so that every vector of the null space is realized as the velocity of a feasible curve.

Combining Lemmas 2.1 and 2.4, we conclude that under LICQ, the set of feasible velocities at x^* is exactly $N(Dh(x^*))$.

2.2 First-order conditions and the Lagrangian

Now suppose x^* is a local maximizer at which LICQ holds. For any feasible curve through x^* , the restriction $t \mapsto f(\gamma(t))$ has a local maximum at the interior point $t = 0$, so its derivative at $t = 0$ is zero. Chain rule and Lemma 2.4 thus yield

$$\nabla f(x^*) \cdot v = 0 \quad \text{for every } v \in N(Dh(x^*)).$$

This is the equality-constraint analogue of (FOC). It holds with equality because both directions along the constraint set are available: if v is a tangent direction, then so is $-v$, the velocity of the reversed curve $\gamma(-t)$. Simply put, the gradient of the objective f must be orthogonal to every tangent direction v , or else moving along a feasible curve in one of the directions v and $-v$ would increase f . Under LICQ, the tangent directions are exactly the vectors orthogonal to the constraint gradients.¹ Since $\nabla f(x^*)$ is orthogonal to every tangent direction, it must be a linear combination of the constraint gradients; that is, there are coefficients $\mu_1^*, \dots, \mu_k^*$ such that

$$\nabla f(x^*) + \mu_1^* \nabla h_1(x^*) + \dots + \mu_k^* \nabla h_k(x^*) = 0.$$

These coefficients are called the *Lagrange multipliers*. The first-order condition is summarized by the following theorem.

Theorem 2.5 (Lagrange multiplier theorem). *Let f and h be C^1 , let x^* be a local extremum (maximizer or minimizer) of f subject to $h(x) = 0$, and let LICQ hold at x^* . Then there is a unique vector $\mu^* \in \mathbb{R}^k$ such that*

$$\nabla f(x^*) + Dh(x^*)^\top \mu^* = 0.$$

Proof (optional, using Lecture 8's implicit function theorem). Stack the objective gradient on top of the constraint gradients:

$$M = \begin{pmatrix} \nabla f(x^*)^\top \\ \nabla h_1(x^*)^\top \\ \vdots \\ \nabla h_k(x^*)^\top \end{pmatrix}.$$

Suppose, for contradiction, that $\text{rank } M = k + 1$. Then some $k + 1$ columns of M form an invertible matrix. Hold the remaining coordinates fixed at their x^* -values and apply the implicit function theorem to the selected coordinates in the system

$$f(x) = c_0, \quad h(x) = c.$$

It follows that every right-hand side (c_0, c) sufficiently close to $(f(x^*), 0)$ has a solution near x^* . In particular, take $c = 0$ and $c_0 = f(x^*) + \varepsilon$. If x^* is a local maximizer, choose $\varepsilon > 0$; if it is a local minimizer, choose $\varepsilon < 0$. Either choice produces a nearby feasible point that contradicts the optimality of x^* . Therefore, $\text{rank } M \leq k$.

¹The row interpretation of the null space from Lecture 3 says that $Av = 0$ if and only if v is orthogonal to every row of A . We also use the following row-space/null-space fact without proof: every vector orthogonal to all of $N(A)$ is a linear combination of the rows of A . For $A = Dh(x^*)$, those rows are the transposed constraint gradients.

By LICQ, the k constraint gradients are linearly independent. Since adding $\nabla f(x^*)$ does not increase their rank, $\nabla f(x^*)$ must be a linear combination of them. Thus there is a vector $\mu^* \in \mathbb{R}^k$ such that

$$\nabla f(x^*) + Dh(x^*)^\top \mu^* = 0.$$

Finally, if both μ^* and $\tilde{\mu}$ satisfy this equation, then $Dh(x^*)^\top (\mu^* - \tilde{\mu}) = 0$. LICQ makes the constraint gradients linearly independent, so $\mu^* = \tilde{\mu}$. $\square$

As in Lecture 6, the theorem gives only a necessary condition for a local extremum. Note that the Lagrange multipliers are not restricted in sign, i.e., they can be positive or negative. For a fixed objective, constraint description, and candidate x^* , LICQ makes the multiplier vector unique. Rescaling or rewriting a constraint (e.g., multiplying both sides by 2) changes its multiplier. Finally, the theorem does not guarantee that an optimizer exists; existence must be established separately.

LICQ cannot be dropped from Theorem 2.5, and all points in the feasible set that do not satisfy LICQ must be treated separately.

Example 2.6 (A singular description loses the multiplier). Consider $\max_{x \in \mathbb{R}} x$ subject to $h(x) = x^2 = 0$. We have $f'(x) = 1$ and $h'(x) = 2x$. If we were to apply the Lagrange multiplier theorem, we would get $1 + 2x\mu = 0$. However, the feasible set is the single point $\{0\}$, meaning that $x^* = 0$ is the global maximizer. Clearly, at $x^* = 0$ there does not exist a μ that satisfies the multiplier condition $1 + 2x^*\mu = 0$.

The Lagrange multiplier condition and the feasibility requirement $h(x) = 0$ form a system of $n+k$ equations in the $n+k$ unknowns (x, μ) . These equations can be written as the critical-point conditions of a single function of $n+k$ variables, the *Lagrangian*.

Definition 2.7 (Lagrangian). The *Lagrangian* of the problem $\max f(x)$ subject to $h(x) = 0$ is the function of x and μ given by

$$L(x, \mu) = f(x) + \mu \cdot h(x).$$

Its partial gradients are:

$$\nabla_x L(x, \mu) = \nabla f(x) + Dh(x)^\top \mu, \quad \nabla_\mu L(x, \mu) = h(x).$$

Consequently, the critical points of L are exactly the pairs (x, μ) that satisfy the first-order condition and the feasibility requirement. Note also that the Lagrangian is not being maximized: a constrained maximizer is a critical point of $L(\cdot, \mu^*)$, and can thus be a maximum, a minimum, or a saddle point of the Lagrangian.

Example 2.8 (Solving directly and with a Lagrangian). Consider $\max x_1 x_2$ subject to $x_1 + x_2 = 2$.

Direct solution. The constraint gives $x_2 = 2 - x_1$, so the problem reduces to

$$\max_{x_1 \in \mathbb{R}} x_1(2 - x_1) = \max_{x_1 \in \mathbb{R}} [1 - (x_1 - 1)^2].$$

The unique global maximizer is $x_1^* = 1$, and the constraint then gives $x_2^* = 1$. Thus $x^* = (1, 1)^\top$.

Lagrangian solution. Write the constraint as $h(x) = 2 - x_1 - x_2 = 0$. Since $\nabla h(x) = (-1, -1)^\top \neq 0$, LICQ holds. The Lagrangian is

$$L(x, \mu) = x_1 x_2 + \mu(2 - x_1 - x_2),$$

so every constrained optimum must be among the solutions to

$$\frac{\partial L}{\partial x_1} = x_2 - \mu = 0, \quad \frac{\partial L}{\partial x_2} = x_1 - \mu = 0, \quad \frac{\partial L}{\partial \mu} = 2 - x_1 - x_2 = 0.$$

The first two equations give $x_1 = x_2 = \mu$. Substituting into the third gives $\mu^* = 1$ and hence $x^* = (1, 1)^\top$, the same point found above. The direct solution established that this point is the global maximizer; the Lagrangian equations by themselves identified it only as a candidate.

In particular, we did not maximize the Lagrangian. As a function of x with $\mu = \mu^*$ fixed,

$$L(x, \mu^*) = x_1 x_2 + 2 - x_1 - x_2 \quad \text{has Hessian} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

whose characteristic equation $\lambda^2 - 1 = 0$ gives eigenvalues 1 and -1 . The Hessian is indefinite, so by the second-order classification of Lecture 6 the point x^* is a saddle point of $L(\cdot, \mu^*)$. On the feasible set, $h(x) = 0$ makes $L(x, \mu^*) = f(x)$, so x^* does maximize the Lagrangian among feasible points. The Lagrangian collects the first-order conditions of the constrained problem; it does not turn the problem into an unconstrained maximization of L .

3 Inequalities and the KKT conditions

3.1 Standard form and active sets

We fix one convention for inequality constraints. Every inequality is written as

$$g_i(x) \geq 0,$$

so $a(x) \leq b$ becomes $b - a(x) \geq 0$.

The optimization problem is

$$\max_{x \in U} f(x) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \geq 0,$$

where $h : U \rightarrow \mathbb{R}^k$ and $g : U \rightarrow \mathbb{R}^\ell$. Its Lagrangian is

$$L(x, \mu, \lambda) = f(x) + \mu \cdot h(x) + \lambda \cdot g(x), \quad \mu \in \mathbb{R}^k, \quad \lambda \geq 0.$$

The requirement $\lambda \geq 0$ is forced by the convention $g \geq 0$, and the one-dimensional calculation after Theorem 3.2 shows why.

For inequality constraints, we will use the following classification.

Definition 3.1 (Active and slack constraints, mixed LICQ). At a feasible point x , the *active set* is

$$I(x) = \{i : g_i(x) = 0\},$$

and a constraint with $g_i(x) > 0$ is *slack*. LICQ holds at x for the mixed problem if the vectors

$$\{\nabla h_j(x) : j = 1, \dots, k\} \cup \{\nabla g_i(x) : i \in I(x)\}$$

are linearly independent.

A slack constraint stays satisfied throughout a neighborhood of x (by continuity of g_i), so it places no restriction on local movement and its gradient does not enter the constraint qualification.

3.2 The KKT theorem

Theorem 3.2 (Karush–Kuhn–Tucker). *Let f , h , and g be C^1 , let x^* be a local maximizer subject to $h(x) = 0$ and $g(x) \geq 0$, and let LICQ hold at x^* . Then there are $\mu^* \in \mathbb{R}^k$ and $\lambda^* \in \mathbb{R}^\ell$ such that*

$$\begin{aligned} \nabla f(x^*) + Dh(x^*)^\top \mu^* + Dg(x^*)^\top \lambda^* &= 0 && (\text{stationarity}) \\ h(x^*) &= 0, \quad g(x^*) \geq 0 && (\text{feasibility}) \\ \lambda^* &\geq 0 && (\text{multiplier signs}) \\ \lambda_i^* g_i(x^*) &= 0 \quad (i = 1, \dots, \ell) && (\text{complementary slackness}) \end{aligned}$$

We use this result without proof.

Complementary slackness forces $\lambda_i^* = 0$ whenever $g_i(x^*) > 0$, and permits $\lambda_i^* > 0$ only where $g_i(x^*) = 0$: slack constraints have zero multipliers. The converse implication fails: an *active* constraint can also have a zero multiplier (as in Example 3.3 below).

The sign restriction $\lambda \geq 0$ can be read off a one-dimensional picture. Suppose $n = \ell = 1$, the constraint $g(x^*) = 0$ is active, and $g'(x^*) > 0$, so $g(x^* + t) > 0$ for every small $t > 0$ and $v = 1$ is a feasible direction. If x^* is a local maximizer, the objective cannot increase into the feasible side, which is (FOC) with $v = 1$: $f'(x^*) \leq 0$. Stationarity reads $f'(x^*) + \lambda^* g'(x^*) = 0$, so

$$\lambda^* = -\frac{f'(x^*)}{g'(x^*)} \geq 0.$$

With the constraint written as $g \leq 0$ instead, the same calculation would produce the opposite sign, so the convention must be fixed once and recorded.

Like every first-order condition in this course, the KKT conditions generate candidates but do not certify them. On $\max_{x \geq 0} x^3$, the point $x^* = 0$ with $\lambda^* = 0$ satisfies every condition of Theorem 3.2, yet 0 is not even a local maximizer, since $x^3 > 0$ for every $x > 0$. Existence and certification remain separate steps, and here existence in fact fails.

3.3 Active-set reasoning in a capacity problem

With ℓ inequalities, the practical method is to reason/guess about which constraints bind. Each guess of an active set of constraints turns the KKT conditions into equalities that can be solved. The remaining KKT conditions — feasibility and multiplier signs — then confirm that the guess is a solution, or reject it.

Example 3.3 (KKT by guess-and-verify). Let $a > 0$ and $K > 0$, and consider

$$\max_q \quad aq - \frac{1}{2}q^2 \quad \text{subject to} \quad q \geq 0, \quad K - q \geq 0.$$

The feasible set $[0, K]$ is nonempty and compact and the objective is continuous, so a maximizer exists by Weierstrass. With $g_1(q) = q$ and $g_2(q) = K - q$, the Lagrangian is

$$L(q, \lambda) = aq - \frac{1}{2}q^2 + \lambda_1 q + \lambda_2 (K - q),$$

and the KKT conditions are

$$a - q + \lambda_1 - \lambda_2 = 0, \quad 0 \leq q \leq K, \quad \lambda_1, \lambda_2 \geq 0, \quad \lambda_1 q = 0, \quad \lambda_2 (K - q) = 0.$$

There are four cases to check.

1. $\lambda_1 = \lambda_2 = 0$. Then stationarity gives $q = a$. To satisfy the second constraint, we must have $K - q \geq 0 \iff K \geq a$.
2. $\lambda_1 = 0, K - q = 0$. Then stationarity gives $\lambda_2 = a - K$, and $\lambda_2 \geq 0$ is satisfied if and only if $K \leq a$.
3. $\lambda_2 = 0, q = 0$. Then stationarity gives $\lambda_1 = -a < 0$, which violates the sign restriction; this guess is rejected for every $a > 0$.
4. $\lambda_1, \lambda_2 > 0$. This case is impossible because we cannot have $q = 0$ and $q = K$ for $K > 0$.

The solution is thus

$$q^*(a, K) = \min\{a, K\}, \quad \lambda_2^*(a, K) = \max\{a - K, 0\}.$$

Note that when $a = K$, the solution is $q^* = K$ and the multiplier is $\lambda_2^* = 0$, meaning that the capacity constraint is active and the multiplier is zero.

To certify that q^* is a global maximizer, observe that the objective can be rewritten as follows:

$$aq - \frac{1}{2}q^2 = -\frac{1}{2}(q - a)^2 + \frac{1}{2}a^2,$$

which shows that the objective is maximized by the point q^* that minimizes $(q - a)^2$, i.e., the point closest to a . Since the point of $[0, K]$ closest to a is $\min\{a, K\} = q^*$, it is indeed a global maximizer.

4 Global certification

4.1 Concave programs

In Lecture 6, if the objective function f is concave on a convex set, then the first-order condition is sufficient for a global maximum. The same logic extends to the following class of constrained problems.

Definition 4.1 (Concave program). Let $U \subseteq \mathbb{R}^n$ be convex. The problem

$$\max_{x \in U} f(x) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \geq 0,$$

is a *concave program* if f and every g_i are concave on U and h is affine.

The sufficiency proof again relies on the tangent inequality from Lecture 5, $f(y) \leq f(x) + \nabla f(x) \cdot (y - x)$.

Theorem 4.2 (Sufficiency in a concave program). *Consider a concave program on an open convex set U , and suppose that f and every g_i are differentiable. If a feasible x^* and multipliers (μ^*, λ^*) satisfy the KKT conditions of Theorem 3.2, then x^* is a global maximizer.*

Proof (optional). Let y be feasible and put $v = y - x^*$. Concavity of f gives

$$f(y) - f(x^*) \leq \nabla f(x^*) \cdot v,$$

and stationarity rewrites the right-hand side as

$$\nabla f(x^*) \cdot v = -\mu^* \cdot Dh(x^*)v - \lambda^* \cdot Dg(x^*)v.$$

Because h is affine and both points are feasible, $h(y) - h(x^*) = Dh(x^*)v = 0$, so the μ^* term vanishes. For each i , concavity of g_i gives $\nabla g_i(x^*) \cdot v \geq g_i(y) - g_i(x^*)$, and since $\lambda^* \geq 0$,

$$-\lambda^* \cdot Dg(x^*)v \leq -\lambda^* \cdot (g(y) - g(x^*)) = \lambda^* \cdot g(x^*) - \lambda^* \cdot g(y) \leq 0,$$

where the final step uses complementary slackness, $\lambda^* \cdot g(x^*) = 0$, and then $\lambda^* \geq 0$ with $g(y) \geq 0$. Chaining the displays gives $f(y) \leq f(x^*)$. $\square$

Recall from Lecture 6 that the set of maximizers of a strictly concave function on a convex set contains at most one point (see Theorem 3.1, part 3 in Lecture 6). The constraints of a concave program define a convex feasible set (possibly empty): it is the intersection of superlevel sets of concave functions and an affine equality set, each convex (see Lecture 5). Thus, if f is strictly concave, the program has at most one maximizer.

The following condition requires the set of strictly feasible points (points that satisfy the equality constraints and satisfy every inequality strictly) to be nonempty. In a concave program, it makes the KKT conditions necessary; they are already sufficient by Theorem 4.2.

Definition 4.3 (Slater's condition). A concave program with affine equalities satisfies *Slater's condition* if there is $\bar{x} \in U$ with

$$h(\bar{x}) = 0 \quad \text{and} \quad g_i(\bar{x}) > 0 \quad \text{for every } i.$$

Theorem 4.4 (Necessity under Slater's condition). *Consider a concave program on an open convex set U , and suppose that f and every g_i are differentiable. If Slater's condition holds, then every global maximizer satisfies the KKT conditions with some multipliers (μ^*, λ^*) .*

Slater and LICQ are both constraint qualifications, but they are different: LICQ is local and point-specific, while Slater is global and specific to concave programs. However, both are sufficient to guarantee the existence of Lagrange multipliers.

Here is another example of what goes wrong when a constraint qualification fails.

Example 4.5 (A concave program with no multipliers). Consider

$$\max_{(x_1, x_2) \in \mathbb{R}^2} x_2 \quad \text{subject to} \quad g_1(x) = x_1 \geq 0, \quad g_2(x) = -x_1 - x_2^2 \geq 0.$$

This is a concave program: the objective and g_1 are affine, and g_2 is concave. Feasibility requires $0 \leq x_1 \leq -x_2^2 \leq 0$, so the feasible set is $\{(0, 0)\}$ and $x^* = (0, 0)$ is the global maximizer. Slater's condition fails because $g_1(x) > 0$ requires $x_1 > 0$, whereas $g_2(x) > 0$ requires $x_1 < -x_2^2 \leq 0$.

At x^* , stationarity would require

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = 0,$$

which is impossible for any multipliers because the second component is always 1. Thus the global maximizer has no KKT multipliers. Yet, the global maximizer exists and is unique — it is the only feasible point $(0, 0)$.

4.2 Second-order conditions on the tangent space

For equality-constrained problems there is also a second-order local test. The relevant matrix is the Hessian of the *Lagrangian*, not of the objective: curvature only matters along the tangent directions, and along those directions both the objective and the constraints curve. The multiplier terms in L account for the curvature of the constraints.

Theorem 4.6 (Second-order conditions on the tangent space). *Let f and h be C^2 , let x^* be feasible with LICQ, and let μ^* satisfy the stationarity condition from the Lagrange multiplier Theorem 2.5:*

$$\nabla f(x^*) + \text{D}h(x^*)^\top \mu^* = 0.$$

Write

$$T = N(\text{D}h(x^*)), \quad H = H_f(x^*) + \sum_{j=1}^k \mu_j^* H_{h_j}(x^*),$$

so H is the Hessian of $L(\cdot, \mu^*)$ at x^* .

1. If x^* is a local constrained maximizer, then $v^\top H v \leq 0$ for every $v \in T$.
2. If $v^\top H v < 0$ for every nonzero $v \in T$, then x^* is a strict local constrained maximizer.

Example 4.7 (Classifying two candidates on a circle). Consider $\max x_1 + x_2$ subject to $h(x) = 2 - x_1^2 - x_2^2 = 0$.

Candidates. Since $\nabla h(x) = (-2x_1, -2x_2)^\top \neq 0$ at every point of the circle, LICQ holds at every feasible point. The stationarity system of $L = x_1 + x_2 + \mu(2 - x_1^2 - x_2^2)$ is

$$1 - 2\mu x_1 = 0, \quad 1 - 2\mu x_2 = 0, \quad x_1^2 + x_2^2 = 2,$$

with the two solutions $(x^*, \mu^*) = ((1, 1)^\top, \frac{1}{2})$ and $((-1, -1)^\top, -\frac{1}{2})$.

Classification. The objective is linear, so $H_f = 0$ and the curvature comes entirely from the constraint:

$$H = \mu^* H_h = -2\mu^* I,$$

which is $-I$ at the first candidate and I at the second. Both tangent spaces are $T = \{v : v_1 + v_2 = 0\}$. On T , the first candidate has $v^\top (-I)v = -\|v\|^2 < 0$ for $v \neq 0$, a strict local maximum, and the second has $v^\top I v > 0$, a strict local minimum. The Hessian of the objective alone is zero at both points, so it cannot distinguish them; the classification comes from the constraint curvature term $\mu^* H_h$.

Certification. The circle is compact, so Weierstrass gives a global maximizer and a global minimizer, and by Theorem 2.5 both are among the two candidates. Comparing values, $f(1, 1) = 2$ and $f(-1, -1) = -2$, so the local classification is also the global one.

There is also a second-order local test for inequality constraints, but it goes beyond the scope of this course due to the fact that the relevant set of directions is a cone rather than a subspace.