Learning objectives. By the end of class, you should be able to

1. recall when the KKT conditions certify constrained first-order optimality, and use it to solve convex and nonconvex problems;
2. define the critical cone, and state the second-order optimality conditions;
3. check the second-order conditions with the projected Hessian under strict complementarity.

1 Recap: First-order optimality

Recall from basic calculus that

$$f(x+td) = f(x) + t\nabla f(x)^T d + o(t).$$

More generally, if $\gamma(t)$ is a feasible path on some domain Ω

$$\gamma : [0,1) \to \Omega, \quad \gamma(0) = x, \quad \dot{\gamma}(0) = d,$$

then

$$f(\gamma(t)) = f(\underbrace{\gamma(0)}_{x}) + t\nabla f(\underbrace{\gamma(0)}_{x})^T \underbrace{\dot{\gamma}(0)}_{d} + o(t).$$

Without constraints, so $\Omega = \mathbb{R}^n$, then x is **first-order optimal** if no direction decreases f to first order,

$$\nabla f(x)^T d \geq 0 \quad \text{for all } d \in \mathbb{R}^n.$$

This holds if and only if $\nabla f(x) = 0$, because $d = -\nabla f(x)$ gives $-\|\nabla f(x)\|^2 \geq 0$.

With constraints, the problem is

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0 \text{ for all } i \in \{1,\dots,m\}, \\ & h_j(x) = 0 \text{ for all } j \in \{1,\dots,p\}, \end{aligned}$$

now with twice continuously differentiable functions and feasible set Ω. Only directions that stay feasible count. The **tangent cone**

$$\mathcal{T}(x) = \{\dot\gamma(0) : \gamma(0) = 0, \quad \gamma : [0,1) \to \Omega \text{ is feasible path}\}$$

is the set of initial velocities of feasible paths from x, and first-order optimality reads

$$\nabla f_0(x)^T d \ge 0 \quad \text{for all } d \in \mathcal{T}(x). \tag{FOC}$$

Every local minimizer satisfies (FOC). But $\mathcal{T}(x)$ is defined through paths, and checking (FOC) is coNP-complete in general.

Instead, linearize the active constraints, $\mathcal{A}(x) = \{i : f_i(x) = 0\}$. The **linearized cone** is

$$\mathcal{F}(x) = \left\{ d \in \mathbb{R}^n : \begin{array}{l} \nabla f_i(x)^T d \leq 0 \text{ for all } i \in \mathcal{A}(x), \\ \nabla h_j(x)^T d = 0 \text{ for all } j \in \{1, \ldots, p\} \end{array} \right\},$$

and can be used as a surrogate for $\mathcal{T}(x) \subseteq \mathcal{F}(x)$.

Theorem 1. KKT conditions. The following holds

$$\nabla f_0(x)^T d \geq 0 \quad \text{for all } d \in \mathcal{F}(x) \qquad \text{(KKT)}$$

for a primal point x satisfying

$$\text{(primal feasibility) } f_i(x) \leq 0, \quad h_j(x) = 0 \text{ for all } i, j,$$

if and only if there exist dual multipliers $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^p$ with

$$\text{(stationarity) } \nabla f_0(x) + \sum_{i=1}^m y_i \nabla f_i(x) + \sum_{j=1}^p z_j \nabla h_j(x) = 0,$$

$$\text{(dual feasibility) } y_i \geq 0 \text{ for all } i \in \{1, \ldots, m\},$$

$$\text{(complementarity) } y_i f_i(x) = 0 \text{ for all } i \in \{1, \ldots, m\}.$$

The proof is Farkas' lemma. Some $d \in \mathcal{F}(x)$ descends exactly when the following primal problem has a solution

$$\text{find } d \in \mathbb{R}^n \text{ such that } \begin{bmatrix} \nabla f_0(x)^T \\ \nabla f_i(x)^T, \ i \in \mathcal{A}(x) \\ \nabla h_j(x)^T \\ -\nabla h_j(x)^T \end{bmatrix} d \leq \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Its dual problem is the KKT conditions; it is feasible *if and only if* the primal problem is infeasible.

1.1 Constraint qualification

Every $d \in \mathcal{T}(x)$ lies in $\mathcal{F}(x)$, so

$$\text{(KKT)} \quad \Longrightarrow \quad \text{(FOC)}.$$

Simultaneously

$$\text{local optimal} \quad \Longrightarrow \quad \text{(FOC)}.$$

A **constraint qualification** guarantees $\mathcal{T}(x) = \mathcal{F}(x)$, such as LICQ, MFCQ, or Slater's condition. In this case,

$$\text{local optimal} \quad \Longrightarrow \quad \text{(FOC)} \quad \overset{\text{CQ}}{\Longleftrightarrow} \quad \text{(KKT)}.$$

1.2 Relationship to the dual

With the Lagrangian

$$\mathscr{L}(x, y, z) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) + \sum_{j=1}^{p} z_j h_j(x),$$

stationarity reads $\nabla_x \mathscr{L}(x, y, z) = 0$. For convex problems, the KKT conditions hold exactly at primal--dual pairs with zero gap,

$$f_0(\hat{x}) = g(y, z) = \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, y, z).$$

When no such pair exists, the KKT conditions have no solution.

1.3 Example problem

Example 2. Trust-region subproblem. For the following nonconvex problem,

$$\min_{x \in \mathbb{R}^2} \; x^T \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x \text{ s.t. } x^T x = 1,$$

check that constraint qualification holds, write the KKT conditions, find all KKT points, and identify the global min.

2 Second-order optimality

Without constraints, a point with $\nabla f(\hat{x}) = 0$ can be a local minimizer, a local maximizer, or a saddle point. The next term of the Taylor expansion decides,

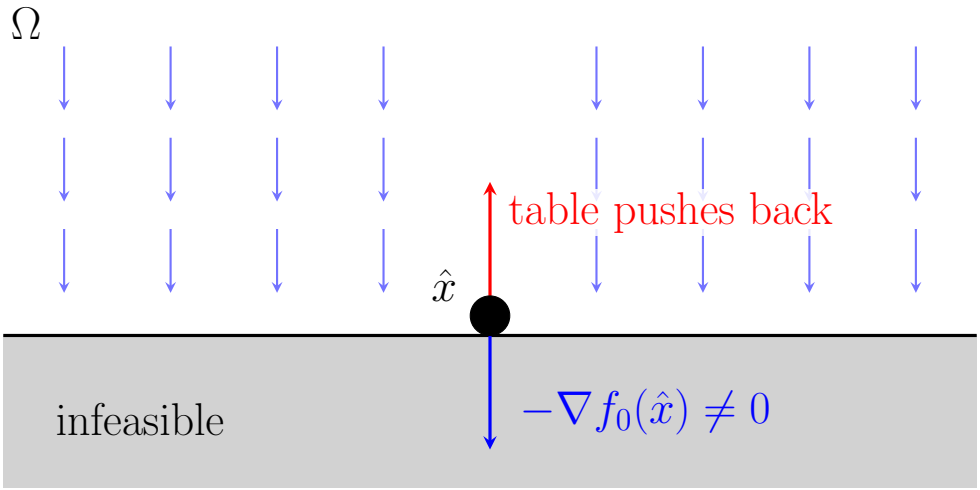
$$f(\hat{x} + d) \approx f(\hat{x}) + \frac{1}{2} d^T \nabla^2 f(\hat{x}) d.$$

Every local minimizer has $\nabla^2 f(\hat{x}) \succeq 0$, and $\nabla^2 f(\hat{x}) \succ 0$ guarantees a strict local minimizer.

With constraints, most solvers, though not all, compute KKT points. A KKT point is only first-order optimal, exactly like a solution of $\nabla f(x) = 0$. It can be a local minimizer, a local maximizer, or a saddle point. In Example 2, the points $\pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ are KKT points and maximizers.

Some methods do reach second-order points. Most of them handle equality constraints, whose feasible set is a smooth surface called a manifold. Examples are Riemannian trust-region methods and Riemannian cubic regularization.

3 Curvature of the constraints



Think of f_0 as the height of a landscape, so that $-\nabla f_0$ is the force on a ball. Here $f_0 = x_2$, so the force $-\nabla f_0 = (0, -1)$ is gravity, and the table $x_2 < 0$ is infeasible.

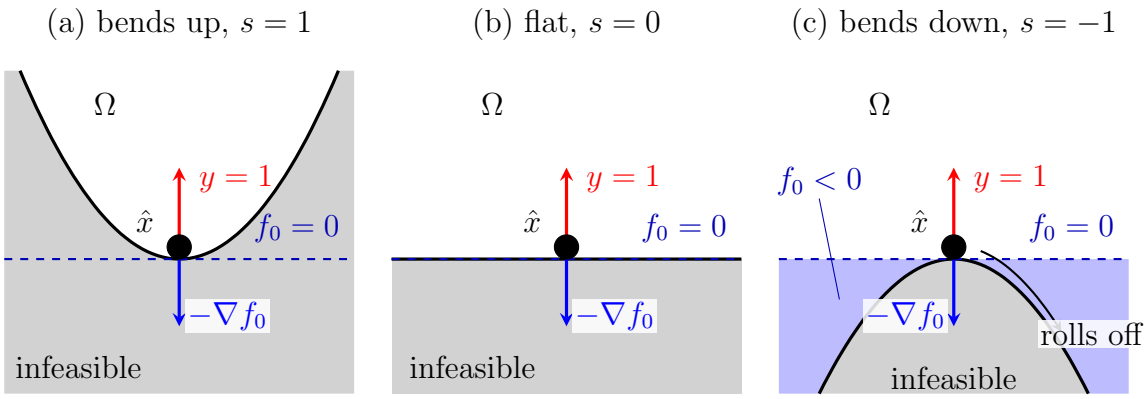
The ball rests on the table at the KKT point $\hat{x} = 0$. The force $-\nabla f_0(\hat{x})$ is not zero. It presses the ball into the table, and the table pushes back.

On a flat table, the ball can roll along the table without changing its height. What if the table is curved?

Example 3. A curved table. Bend the table into $x_2 = sx_1^2$,

$$\min_{x \in \mathbb{R}^2} \; x_2 \text{ s.t. } sx_1^2 - x_2 \leq 0,$$

for $s = 1$ (bends up), $s = 0$ (flat), and $s = -1$ (bends down). Show that $\hat{x} = 0$ is a KKT point in all three cases. Is it a local minimizer?



In all three panels, the force $-\nabla f_0 = (0, -1)$ and the push $y = 1$ of the table balance, and $\nabla^2 f_0 = 0$. The dashed line is the level set $f_0 = 0$ through $\hat{x}$.

- In (a), the table stays above the dashed line. Nearby feasible points have $f_0 > 0$.
- In (b), the table is the dashed line. The ball can roll along it at the same height.
- In (c), the table cuts below it. The shaded points are feasible with $f_0 < 0$, and the ball rolls off.

The objective is linear, so $\nabla^2 f_0 = 0$ cannot tell the three tables apart. Yet along the table, $f_0 = sx_1^2$ is curved. The curvature of the constraint must enter the test.

At a KKT point, $\nabla f_0(x) \neq 0$ in general. The function whose gradient vanishes is the Lagrangian,

$$\nabla_x \mathscr{L}(x, y, z) = 0.$$

We say γ is a **critical** feasible path if

$$\gamma : [0, 1) \to \Omega, \quad \gamma(0) = x, \quad \dot{\gamma}(0) = d,$$
$$f_i(\gamma(t)) = 0 \text{ for all } i \in \mathcal{A}(x).$$

Question 1. Starting from a KKT point x, let γ denote a **critical** feasible path. Then

$$f_0(\gamma(t)) = f_0(x) + \frac{t^2}{2} d^T \nabla^2_{xx} \mathscr{L}(x, y, z) d + o(t^2).$$

The **Hessian of the Lagrangian** is

$$\nabla^2_{xx}\mathscr{L}(x,y,z) = \nabla^2 f_0(x) + \sum_{i=1}^{m} y_i \nabla^2 f_i(x) + \sum_{j=1}^{p} z_j \nabla^2 h_j(x).$$

Observe that this term is not zero even when $\nabla^2 f_0(x) = 0$.

4 The critical cone

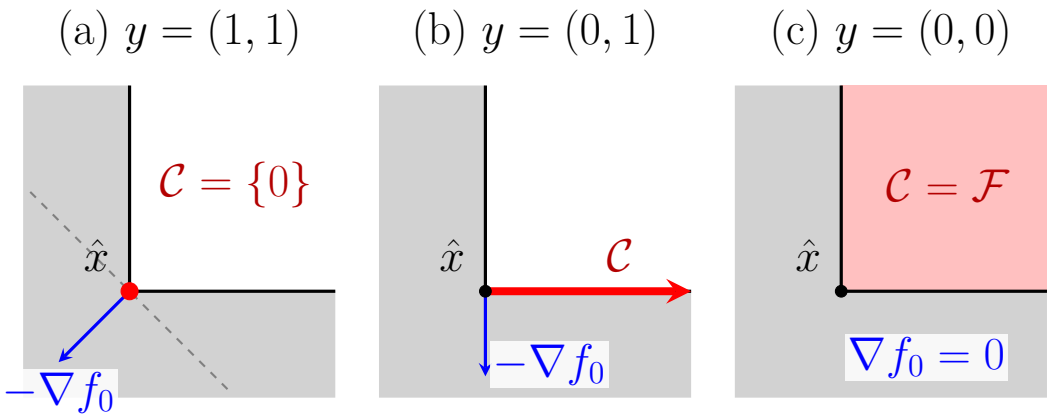
At a KKT point, an active constraint $i \in \mathcal{A}(\hat{x})$ is **strongly active** if $y_i > 0$, and **weakly active** if $y_i = 0$. The **critical cone** keeps the directions that first order leaves undecided,

$$\mathcal{C}(\hat{x}, y) = \left\{ d \in \mathcal{F}(\hat{x}) : \ \nabla f_0(\hat{x})^T d = 0 \right\}.$$

Lemma 4. The critical cone. At a KKT point $(\hat{x}, y, z)$, a direction d lies in $\mathcal{C}(\hat{x}, y)$ if and only if

$$\begin{aligned} &\nabla h_j(\hat{x})^T d = 0 \text{ for all } j \in \{1, \dots, p\}, \\ &\nabla f_i(\hat{x})^T d = 0 \text{ for every strongly active } i, \\ &\nabla f_i(\hat{x})^T d \leq 0 \text{ for every weakly active } i. \end{aligned}$$

Three objectives at the corner $\hat{x} = 0$ of $\Omega = \{x \geq 0\}$, where stationarity gives $y = \nabla f_0(0)$.



In all three panels, $\mathcal{T}(0) = \mathcal{F}(0)$ is the quadrant. The critical cone depends on the objective.

- In (a), both walls push. Every direction that leaves the corner goes uphill at first order, and nothing is left to test.
- In (b), only the floor pushes. Sliding along the floor, away from the wall, is level at first order.
- In (c), no wall pushes, and the whole quadrant is undecided.

5 Second-order conditions

Theorem 5. Second-order necessary condition. Let $x^\star$ be a local minimizer at which LICQ holds, with KKT multipliers (y, z). Then

$$d^T \nabla^2_{xx} \mathscr{L}(x^\star, y, z) d \geq 0 \quad \text{for all } d \in \mathcal{C}(x^\star, y).$$

Theorem 6. Second-order sufficient condition. Let $(\hat{x}, y, z)$ be a KKT point. If

$$d^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) d > 0 \quad \text{for all nonzero } d \in \mathcal{C}(\hat{x}, y),$$

then $\hat{x}$ is a strict local minimizer.

We state this without proof. It needs no constraint qualification.

6 The projected Hessian

At a KKT point, **strict complementarity** holds if

$$y_i > 0 \quad \text{for every } i \in \mathcal{A}(\hat{x}).$$

Then every active constraint enters $\mathcal{C}$ as an equality, and the critical cone is a subspace,

$$\mathcal{C}(\hat{x}, y) = \left\{ d \in \mathbb{R}^n : \begin{array}{l} \nabla h_j(\hat{x})^T d = 0 \text{ for all } j \in \{1, \dots, p\}, \\ \nabla f_i(\hat{x})^T d = 0 \text{ for all } i \in \mathcal{A}(\hat{x}) \end{array} \right\}.$$

This includes every problem with only equality constraints. Let the columns of Z form a basis of $\mathcal{C}(\hat{x}, y)$. Every critical direction is $d = Zu$, and

$$d^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) d = u^T \left(Z^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) Z \right) u.$$

The matrix $Z^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) Z$ is the **projected Hessian**.

Proposition 7. Projected Hessian test. Under strict complementarity,

$$\begin{array}{rcl} \hat{x} \text{ local minimizer, LICQ} & \Longrightarrow & Z^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) Z \succeq 0, \\ Z^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) Z \succ 0 & \Longrightarrow & \hat{x} \text{ strict local minimizer.} \end{array}$$

Both are eigenvalue tests, which cost $O(n^3)$ operations. For one constraint in $\mathbb{R}^2$, take Z to be its gradient rotated by 90 degrees.

Without strict complementarity, each weakly active constraint keeps its inequality, and the critical cone is no longer a subspace. The test changes from a subspace to a cone,

$$Z^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) Z \succeq 0$$

$$\downarrow$$

$$d^T \nabla^2_{xx} \mathscr{L}(\hat{x}, y, z) d \geq 0 \text{ for all } d \in \mathcal{C}(\hat{x}, y).$$

This becomes NP-hard in general.

7 Summary

First order. A local minimizer satisfies (FOC). Every KKT point satisfies (FOC), and a constraint qualification makes every local minimizer a KKT point. For convex problems, a KKT point is a primal-dual pair with zero gap, and it can fail to exist.

Solving. Check a constraint qualification, write the KKT conditions, and solve them. Complementary slackness sets $y_i = 0$ for every inactive constraint.

Curvature of the constraints. At a KKT point, $\nabla f_0 \neq 0$, but $\nabla_x \mathscr{L} = 0$. Along a feasible path that keeps the strongly active constraints at zero, $f_0 = \mathscr{L}$. So the second-order term of f_0 is governed by $\nabla^2_{xx}\mathscr{L}(\hat{x}, y, z)$, which adds the curvature of the constraints, weighted by the multipliers.

Critical cone. Directions with $\nabla f_0^T d > 0$ are decided at first order. The rest form $\mathcal{C}$, where strongly active constraints act as equalities and weakly active ones keep their inequality.

Conditions. Under LICQ, every local minimizer $x^\star$ has

$$d^T \nabla^2_{xx}\mathscr{L}(x^\star, y, z)d \geq 0 \quad \text{for all } d \in \mathcal{C}(x^\star, y).$$

Every KKT point $\hat{x}$ with

$$d^T \nabla^2_{xx}\mathscr{L}(\hat{x}, y, z)d > 0 \quad \text{for all nonzero } d \in \mathcal{C}(\hat{x}, y)$$

is a strict local minimizer.

Checking. Under strict complementarity, $\mathcal{C}$ is a subspace, and the test is the projected Hessian $Z^T \nabla^2_{xx}\mathscr{L}(\hat{x}, y, z)Z$. Otherwise $\mathcal{C}$ is a cone, and the test includes copositivity, which is NP-hard.