ORF522 – Linear and Nonlinear Optimization

10. Optimality conditions for nonlinear optimization

Bartolomeo Stellato — Fall 2024

Today's lecture
[Chapter 2 and 12, NO][Chapter 4 and 5, CO]

Optimality conditions for nonlinear optimization

- Unconstrained optimization
- Constrained optimization (KKT conditions)
- Duality

Unconstrained optimization

First-order necessary conditions

Fermat's Theorem

Theorem

If $x^\star$ is a local optimizer for the continuously differentiable function f, then

$$\nabla f(x^\star) = 0$$

First-order necessary condition

Proof (contraposition)

Assume that $\nabla f(x^\star) \neq 0$. Define $d = -\nabla f(x^\star)$. Then,

$$\nabla f(x^\star)^T d = -\|\nabla f(x^\star)\|^2 < 0$$

Then, by Taylor approximation

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

With small enough t, we can find $y = x^\star + td$ in the neighborhood of $x^\star$ such that

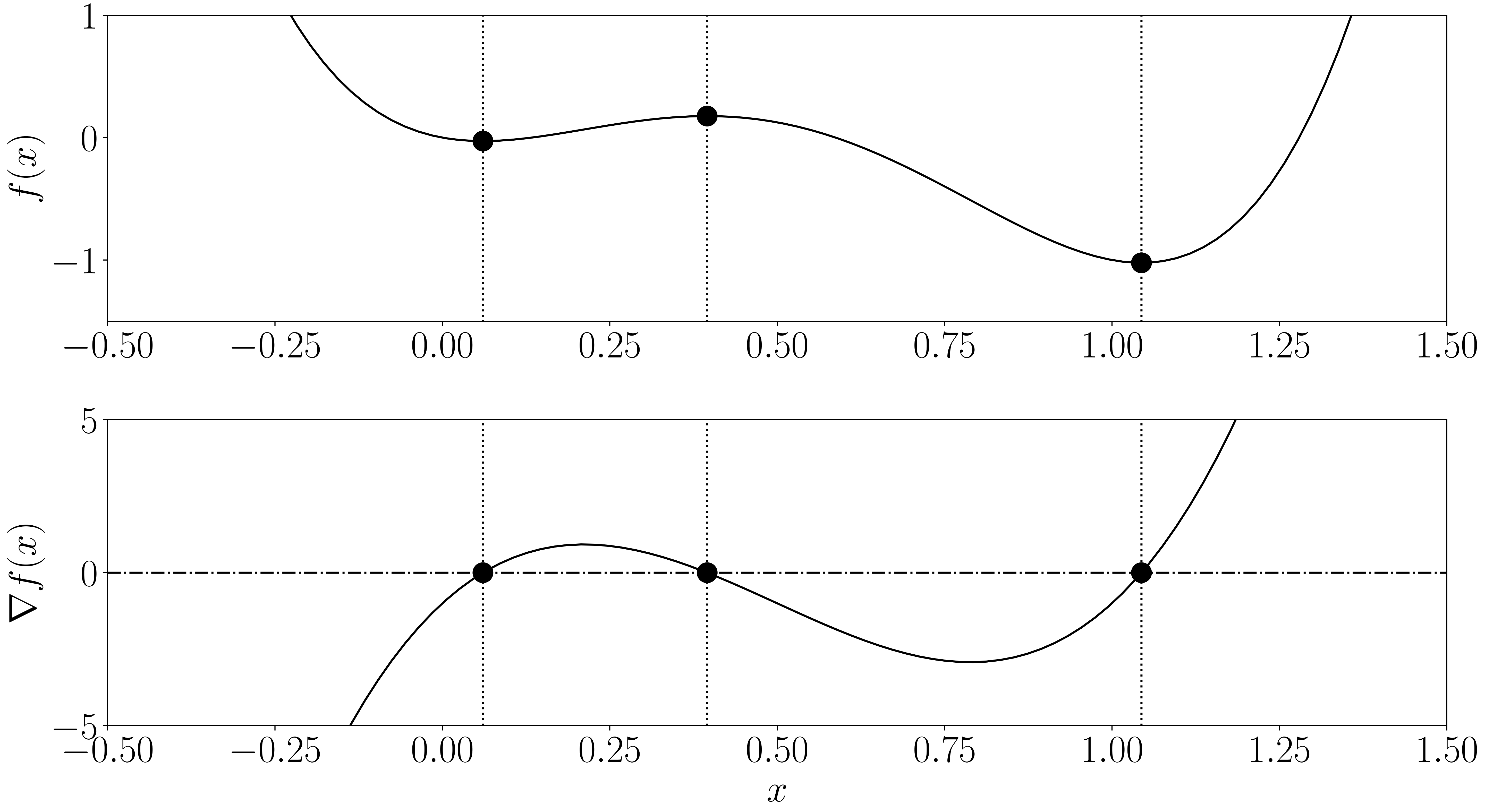
$$f(y) < f(x^\star)$$

$\blacksquare$

First-order necessary condition is not sufficient

$$f(x) = 10x^2(1-x)^2 - x$$

$$\nabla f(x) = 40x^3 - 60x^2 + 20x - 1$$



Each local minimum/maximum satisfies

$$\nabla f(x) = 0$$

Second-order necessary condition

Theorem

If $x^\star$ is a local optimizer for the continuously differentiable function f, then

$$\nabla f(x^\star) = 0 \quad \text{and} \quad \nabla^2 f(x^\star) \succeq 0 \text{ (positive semidefinite)}$$

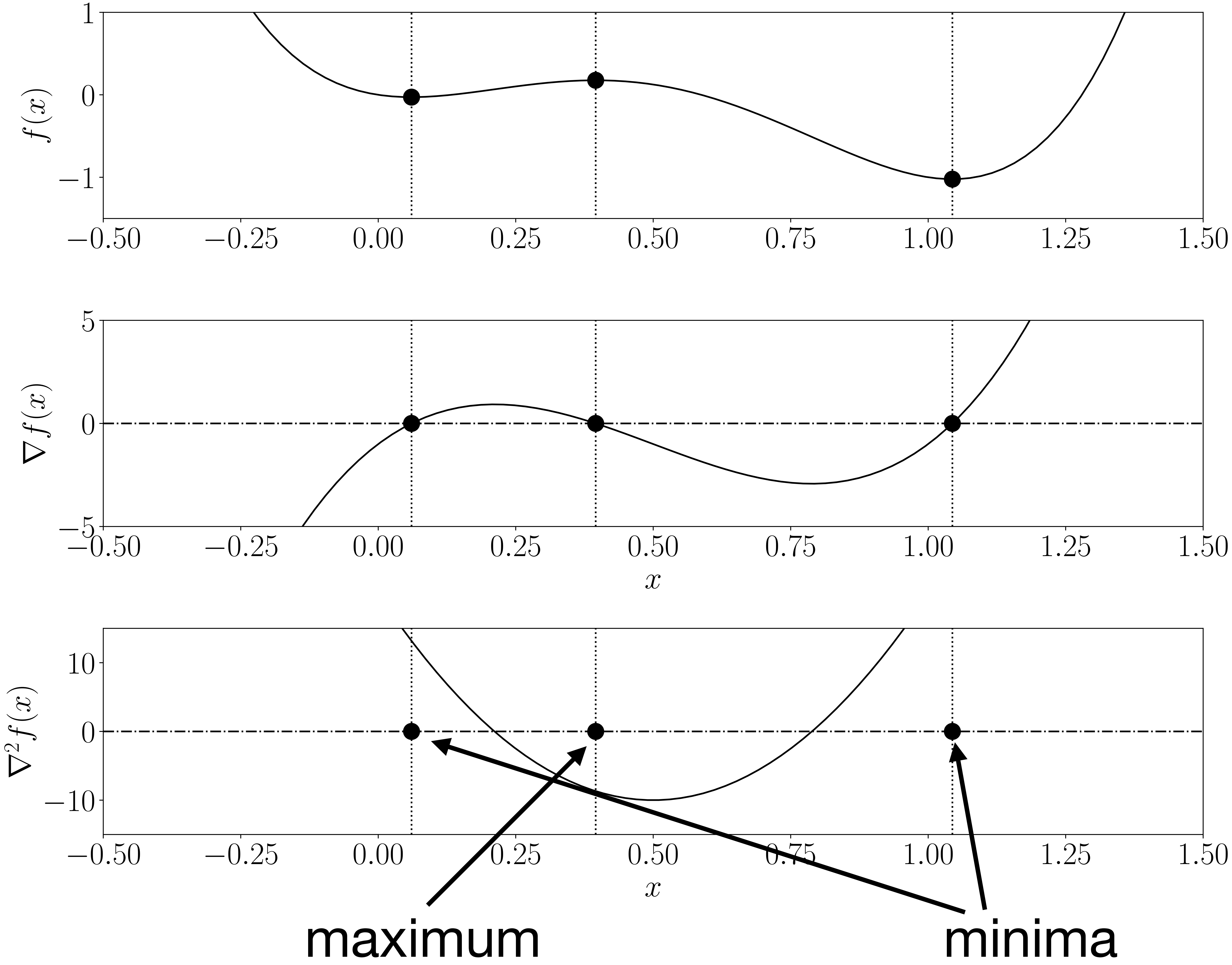
Proof

If $\nabla f(x^\star) = 0$, then the second-order approximation is

$$\begin{aligned} f(x^\star + td) &= f(x^\star) + \cancel{t\nabla f(x^\star)^T d} + t^2(1/2)d^T \nabla^2 f(x^\star)d + o(t^2) \\ &= f(x^\star) + t^2(1/2)d^T \nabla^2 f(x^\star)d + o(t^2) \end{aligned}$$

To have a local minimum $d^T \nabla^2 f(x^\star)d \geq 0$ for any d ■

Example fixed



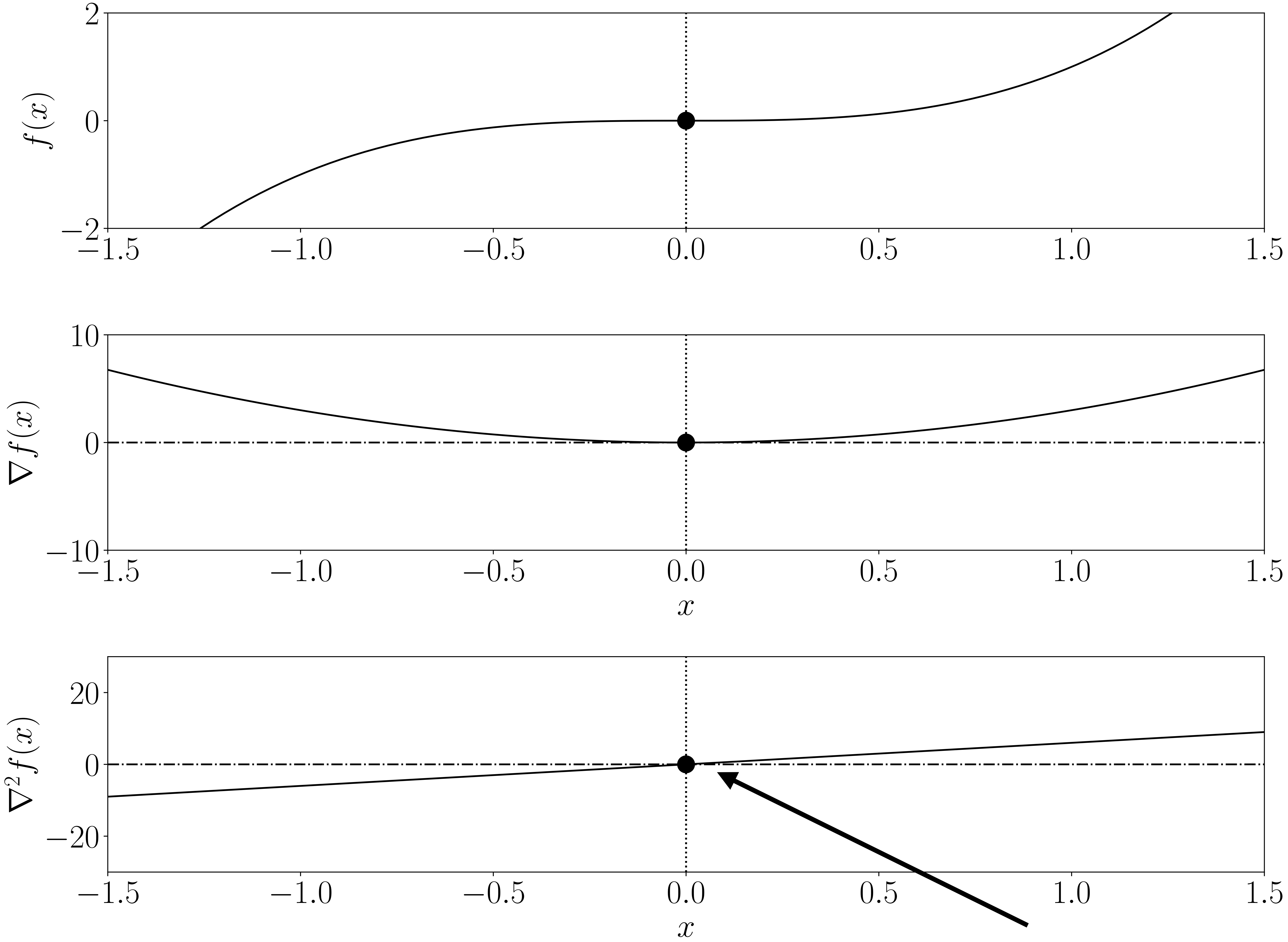
$$f(x) = 10x^2(1-x)^2 - x$$

$$\nabla f(x) = 40x^3 - 60x^2 + 20x - 1$$

$$\nabla^2 f(x) = 120x^2 - 120x + 20$$

Are they sufficient as well?

Second-order necessary condition is not sufficient



not local minimum

Cubic function

$$f(x) = x^3$$

$$\nabla f(x) = 3x^2$$

$$\nabla^2 f(x) = 6x$$

Conditions satisfied

$$\nabla f(0) = 0$$

$$\nabla^2 f(0) = 0 \succeq 0$$

Second-order sufficient condition

Theorem

Let f be a continuously differentiable function. If $x^\star$ satisfies

$$\nabla f(x^\star) = 0 \quad \text{and} \quad \nabla^2 f(x^\star) \succ 0$$

then $x^\star$ is a local minimum of f

Proof

If $\nabla^2 f(x^\star) \succ 0$, then $\exists \lambda > 0$ such that $d^T \nabla^2 f(x^\star) d > \lambda \|d\|_2^2$

Then, if $\nabla f(x^\star) = 0$, in a neighborhood of $x^\star$ we have

$$f(x^\star + td) = f(x^\star) + t^2(1/2) d^T \nabla^2 f(x^\star) d + o(t^2) > f(x^\star)$$

for any d ■

Examples

Cubic function

$f(x) = x^3 \longrightarrow \nabla^2 f(x) = 6x \longrightarrow \nabla^2 f(0) = 0$ (does not satisfy sufficient condition)

Least-squares

$$f(x) = \|Ax - b\| = x^T A^T A x - 2x^T A^T b + b^T b \longrightarrow \nabla^2 f(x) = 2A^T A$$

$2A^T A \succ 0$ if A is full rank
(linear independent columns in A)

Constrained optimization

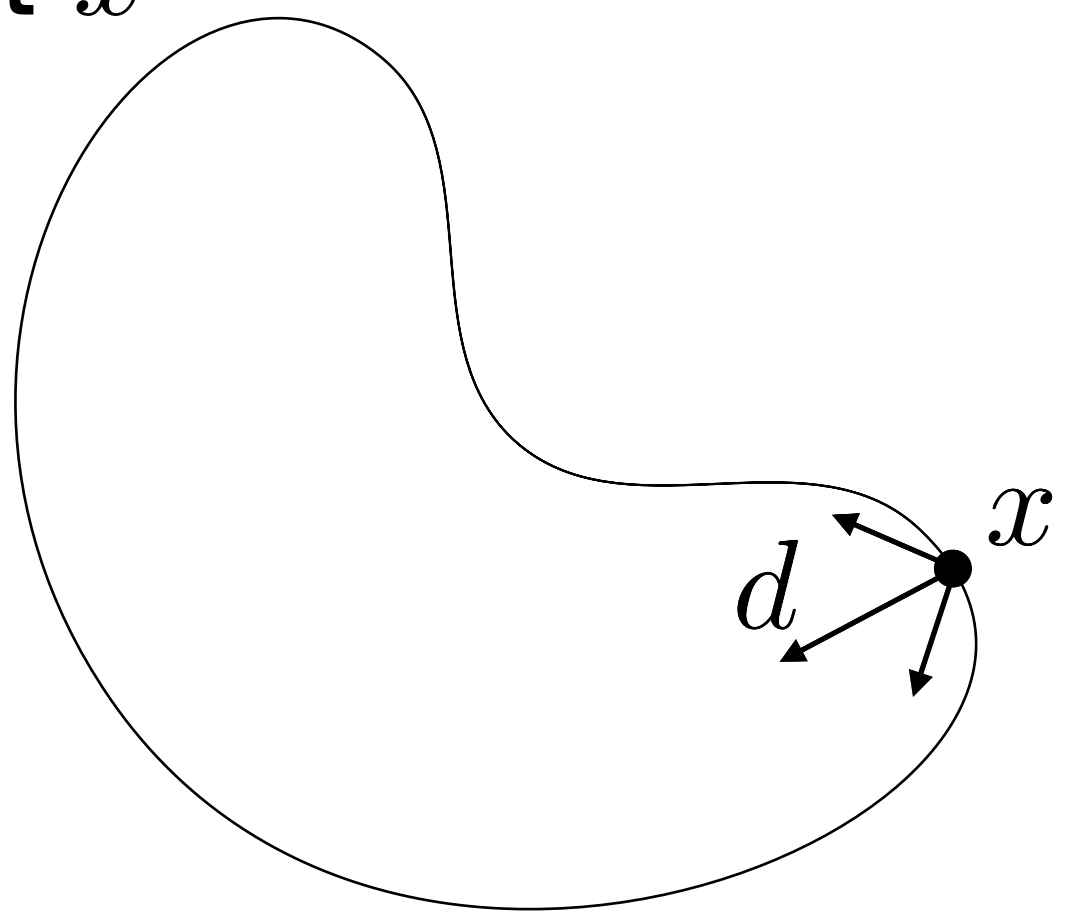
Feasible direction

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array}$$

Given $x \in C$, we call d a **feasible direction** at x if there exists $\bar{t} > 0$ such that

$$x + td \in C, \quad \forall t \in [0, \bar{t}]$$

$F(x)$ is the **set of all feasible directions** at x



Examples

$$C = \{Ax = b\} \quad \Longrightarrow F(x) = \{d \mid Ad = 0\}$$

$$C = \{Ax \leq b\} \quad \Longrightarrow F(x) = \{d \mid a_i^T d \leq 0 \quad \text{if } a_i^T x = b_i\}$$

$$C = \{g_i(x) \leq 0, \text{ (nonlinear)}\} \quad \Longrightarrow F(x) = \{d \mid \nabla g_i(x)^T d < 0 \quad \text{if } g_i(x) = 0\}$$

Descent direction

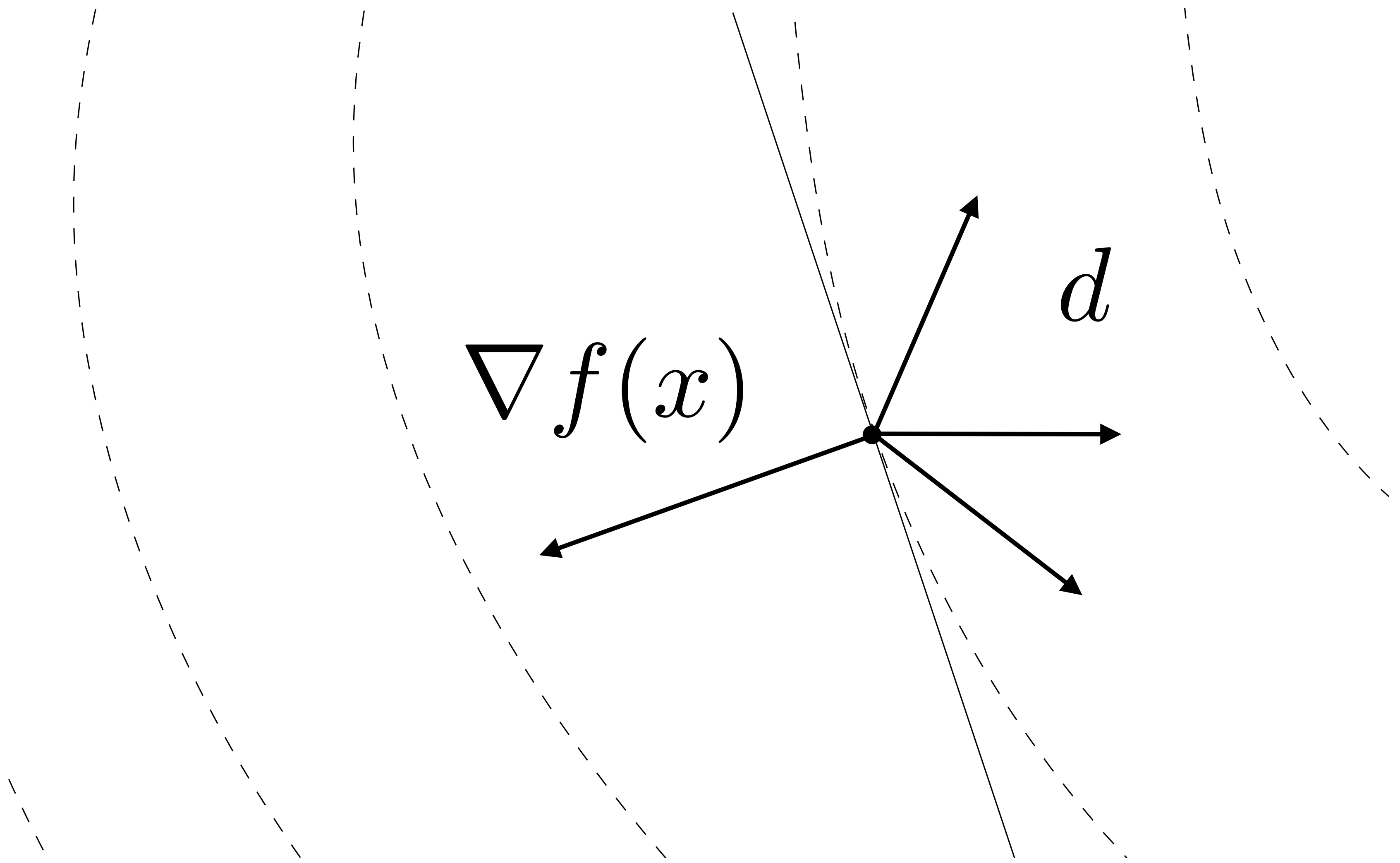
Given continuously differentiable f, we call d a **descent direction** at x if there exists $\bar{t}$ such that

$$f(x + td) < f(x), \quad \forall t \in [0, \bar{t}]$$

$D(x)$ is the **set of all descent directions**

Remark

For all descent directions d at x we have $\nabla f(x)^T d < 0$



Necessary optimality condition idea

All feasible directions are not descent directions

↓

There is no feasible descent direction

If $x^\star$ is a local optimum, then

$$F(x^\star) \cap D(x^\star) = \emptyset$$

Nonlinear optimization with equality constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

Theorem

If $x^\star$ is a local optimum, then $\exists y$ such that $\nabla f(x^\star) + A^T y = 0$

Interpretation

$$\nabla f(x^\star) \in \mathbf{range}(A^T) = \mathbf{null}(A)^\perp \longrightarrow \nabla f(x^\star) \perp \mathbf{null}(A)$$

(perpendicular to hyperplane)

Example: constrained least squares

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & Cx = d \end{array} \longrightarrow \begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ y \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$

optimality conditions

Proof of the theorem

Feasible directions
$F(x) = \{d \mid Ad = 0\}$
Descent directions
$D(x) = \{d \mid \nabla f(x)^T d < 0\}$

→

alternative 1

$$Ad = 0$$
$$\nabla f(x^\star)^T d < 0$$

alternative 2

$\exists y$ such that
$$\nabla f(x^\star) + A^T y = 0$$

can't be both true if $\nabla f(x^\star) + A^T y = 0 \Longrightarrow \nabla f(x^\star)^T d + y^T A d = 0$ *(contradiction)*

can't be both false

minimize	$\nabla f(x^\star)^T d$	maximize	0
subject to	$Ad = 0$	subject to	$\nabla f(x^\star) + A^T y = 0$

if alternative 1, then $p^\star = -\infty \quad \Rightarrow d^\star = -\infty$ (dual infeasible)
if alternative 2, then $p^\star = 0 \quad \Rightarrow d^\star = 0$ (dual feasible) ■

Necessary conditions for smooth nonlinear optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, m \end{array} \qquad (g_i(x) \text{ nonlinear})$$

Linearly independence constraint qualification (LICQ)

Given x and the set of active constraints $\mathcal{A}(x) = \{i \mid g_i(x) = 0\}$, we say that LICQ holds if and only if

$$\{\nabla g_i(x), \quad i \in \mathcal{A}(x)\} \text{ is } \textbf{linearly independent}$$

Theorem

If $x^\star$ is a local minimum and LICQ holds, then there exists $y \geq 0$ such that

$$\nabla f(x^\star) + \sum_{i=1}^m y_i \nabla g_i(x^\star) = 0$$

$$y_i g_i(x^\star) = 0, \quad i = 1, \dots, m$$

Useful Lemma

Farkas lemma variation

Given A, exactly one of the following statements is true

1. There exists an d with $Ad < 0$
2. There exists a u with $A^T u = 0$, $\mathbf{1}^T u = 1$, and $u \geq 0$

Proof

They cannot be both true. $Ad < 0 \;\Rightarrow\; u^T A d < 0$ (contradiction)

They cannot be both false

1 is equivalent to $\tilde{A}\tilde{d} \geq 0,\; c^T \tilde{d} < 0$ with $\tilde{A} = \begin{bmatrix} A & \mathbf{1} \end{bmatrix}$, $c = (0, \dots, 0, 1)$ and $\tilde{d} = (-d, -\epsilon)$

By Farkas lemma (Lec 9) , we have the alternative

$\tilde{A}^T u = c,\; u \geq 0$, equivalent to 2. ■

Necessary conditions for smooth nonlinear optimization

Proof

Feasible directions

$$F(x) = \{d \mid \nabla g_i(x)^T d < 0, \quad i \in \mathcal{A}(x)\}$$

Descent directions

$$D(x) = \{d \mid \nabla f(x)^T d < 0\}$$

Optimality condition

$$F(x) \cap D(x) = \emptyset$$

⟶ Infeasible system

$$Ad < 0, \quad A = \begin{bmatrix} \nabla f(x) & \nabla g_{\mathcal{A}(x)_1}(x) & \dots & \nabla g_{\mathcal{A}(x)_n}(x) \end{bmatrix}^T$$

Farkas lemma variation ⟶ $\exists u \geq 0$ such that $A^T u = 0$ and $\mathbf{1}^T u = 1$

Therefore,

$$u_0 \nabla f(x^\star) + \sum_{i \in \mathcal{A}(x^\star)} u_i \nabla g_i(x^\star) = 0$$

$$u \geq 0, \quad \mathbf{1}^T u = 1$$

Necessary conditions for smooth nonlinear optimization

Proof (continued)

$$u_0 \nabla f(x^\star) + \sum_{i \in \mathcal{A}(x^\star)} u_i \nabla g_i(x^\star) = 0$$

$$u \geq 0, \quad \mathbf{1}^T u = 1$$

If $u_0 = 0$, then $\sum_{i \in \mathcal{A}(x^\star)} u_i \nabla g_i(x^\star) = 0$ (LICQ violated).

Hence, $u_0 > 0$. Let's define $y = u/u_0$, obtaining $\nabla f(x^\star) + \sum_{i \in \mathcal{A}(x)} y_i \nabla g_i(x^\star) = 0$

Which can be rewritten as $$\nabla f(x^\star) + \sum_{i=1}^{m} y_i \nabla g_i(x^\star) = 0$$

$$y_i g_i(x^\star) = 0, \quad i = 1, \dots, m \qquad \blacksquare$$

What happens if LICQ fails?

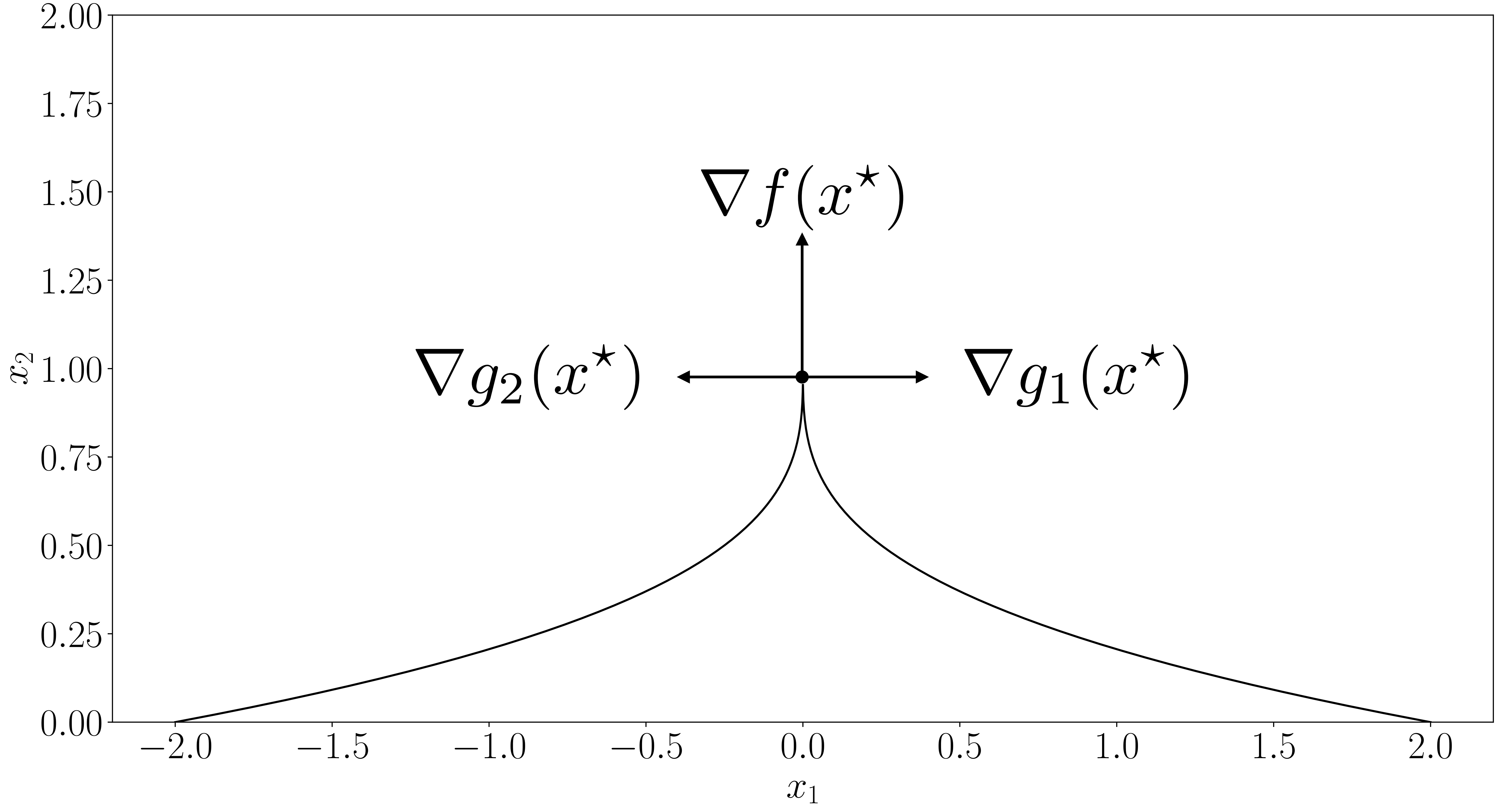
minimize $\quad -x_2$

subject to $\quad x_1 - 2(1 - x_2)^3 \leq 0$

$-x_1 - 2(1 - x_2)^3 \leq 0$

$x \geq 0$

$x^\star = (0, 1)$



KKT necessary conditions for nonlinear optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

Theorem

If $x^\star$ is a local minimizer and LICQ holds, then there exists $y^\star, v^\star$ such that

$$\nabla f(x^\star) + \sum_{i=1}^{m} y_i^\star \nabla g_i(x^\star) + \sum_{i=1}^{p} v_i^\star \nabla h_i(x^\star) = 0$$ **stationarity**

$$y^\star \geq 0$$ **dual feasibility**

$$g_i(x^\star) \leq 0, \quad i = 1, \ldots, m$$
$$h_i(x^\star) = 0, \quad i = 1, \ldots, p$$ **primal feasibility**

$$y_i^\star g_i(x^\star) = 0, \quad i = 1, \ldots, m$$ **complementary slackness**

Duality

Lagrangian dual function

$$p^\star = \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

Lagrangian

$$L(x, y, v) = f(x) + \sum_{i=1}^{m} y_i g_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

Lagrange dual function

$$q(y, z) = \inf_x L(x, y, v)$$

Lower bound property

For any $y \geq 0$ and v, $\quad q(y, v) \leq p^\star$

Proof. Let $\tilde{x}$ be a feasible point. Then,

$$q(y, v) = \inf_x L(x, y, v) \leq f(\tilde{x}) + \sum_{i=1}^{m} y_i \underbrace{g_i(\tilde{x})}_{\leq 0} + \sum_{i=1}^{p} v_i \underbrace{h_i(\tilde{x})}_{= 0} \leq f(\tilde{x})$$

$$\implies \quad g(y, v) \leq p^\star$$

■

Dual problem and weak duality

primal problem

$$p^\star = \text{minimize} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$h_i(x) = 0, \quad i = 1, \ldots, p$$

dual function

$$q(y, z) = \inf_x f(x) + \sum_{i=1}^m y_i g_i(x) + \sum_{i=1}^p v_i h_i(x)$$

dual problem

(find best lower bound)

$$d^\star = \text{maximize} \quad q(y, v)$$
$$\text{subject to} \quad y \geq 0$$

always convex optimization problem
(even when primal is not)

weak duality

(from lower bound property)

$$d^\star \leq p^\star$$

$$p^\star = \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

Strong duality

When is $p^\star = d^\star$?

- Does not hold in general
- (usually) holds for convex problems
- needs conditions (constraint qualifications)

theorem

If the problem is convex and there exists at least a strictly feasible x, *i.e.*,

$$g_i(x) \leq 0, \quad \text{(for all affine } g_i\text{)}$$

$$g_i(x) < 0, \quad \text{(for all non-affine } g_i\text{)}$$

$$h_i(x) = 0, \quad i = 1, \dots, p$$

Slater's condition

then $p^\star = d^\star$ (**strong duality holds**)

remarks

- Slater's condition implies that dual is not unbounded
- Generalizes LP duality

KKT necessary conditions revisited

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, \quad i = 1, \dots, p
\end{array}
$$

Theorem
If $x^\star$ is a local minimizer and strong duality holds, then $\exists y^\star, v^\star$ such that

$$\nabla f(x^\star) + \sum_{i=1}^m y_i^\star \nabla g_i(x^\star) + \sum_{i=1}^p v_i^\star \nabla h_i(x^\star) = 0$$

stationarity
$(\nabla_x L(x, y, v) = 0)$

$$y^\star \geq 0$$

dual feasibility

$$g_i(x^\star) \leq 0, \quad i = 1, \dots, m$$
$$h_i(x^\star) = 0, \quad i = 1, \dots, p$$

primal feasibility

$$y_i^\star g_i(x^\star) = 0, \quad i = 1, \dots, m$$

complementary slackness

KKT conditions for convex problems

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

f, g_i convex
h_i affine

conditions are also sufficient

If $x^\star, y^\star, v^\star$ satisfy KKT conditions for convex problem, then they are optimal.

Proof

$$f(x^\star) = f(x^\star) + \sum_{i=1}^{m} y_i^\star g_i(x^\star) + \sum_{i=1}^{p} v_i^\star h_i(x^\star) = L(x^\star, y^\star v^\star)$$

from complementary slackness

Since $L(x, y, v)$ is convex in x and $\nabla_x L(x^\star, y^\star, v^\star) = 0 \quad \Rightarrow q(y^\star, v^\star) = \inf_x L(x, y^\star, v^\star) = L(x^\star, y^\star, v^\star)$

$$\Rightarrow p^\star = f(x^\star) = q(y^\star, v^\star) = d^\star \quad \blacksquare$$

KKT remarks

History

- First appeared in publication by Kuhn and Tucker (1951)
- It already existed in Karush's unpublished master thesis (1939)

Unconstrained problems

They reduce to necessary first-order condition $\nabla f(x) = 0$

Strong duality

In general, we can replace LICQ assumption with strong duality

Convex problems

KKT conditions are always **sufficient**

If Slater condition holds, KKT conditions are **necessary and sufficient**

Example: KKT conditions for convex QP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x \\ \text{subject to} & Ax = b \\ & Cx \leq d \end{array}$$

Lagrangian

$$L(x, y, v) = (1/2)x^T P x + q^T x + y^T(Cx - d) + v^T(Ax - b)$$

where $y \geq 0$

Stationarity condition

$$\nabla_x L(x, y, u) = Px + q + C^T y + A^T v = 0$$

Example: KKT conditions for convex QP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x \\ \text{subject to} & Ax = b \\ & Cx \leq d \end{array}$$

KKT Optimality conditions

$$Px^\star + q + C^T y^\star + A^T v^\star = 0$$ **stationarity condition**

$$y^\star \geq 0$$ **dual feasibility**

$$Ax - b = 0$$

$$Cx - d \leq 0$$ **primal feasibility**

$$y_i(c_i^T x^\star - d_i) = 0, \quad i = 1, \dots, m$$ **complementary slackness**

Optimality conditions in nonlinear optimization

Today, we learned to:

- **Prove** optimality conditions for unconstrained optimization
- **Compute** feasible and descent directions
- **Derive** optimality conditions for constrained optimization
- **Connect** optimality conditions to duality theory

Next lecture

- Optimization algorithms