

ORF522 – Linear and Nonlinear Optimization

13. Optimality conditions for nonlinear optimization

Ed forum

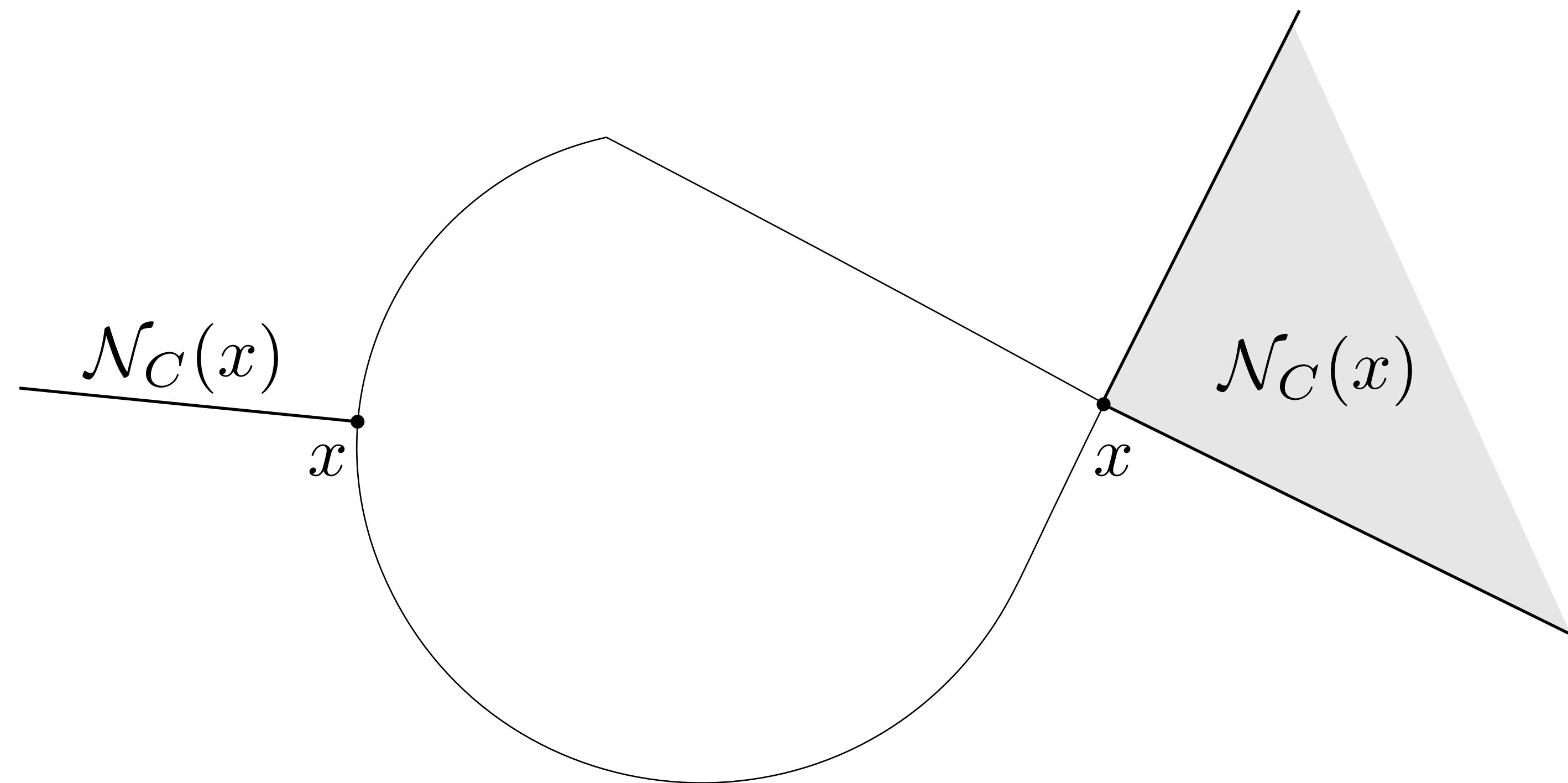
- Isn't the normal cone translated? (Answer in next slide)

Recap

Normal cone

For any set C and point $x \in C$, we define

$$\mathcal{N}_C(x) = \{g \mid g^T(y - x) \leq 0, \text{ for all } y \in C\}$$



Gradient

Derivative

If $f(x) : \mathbf{R}^n \rightarrow \mathbf{R}^m$ continuously differentiable, we define

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

Gradient

If $f : \mathbf{R}^n \rightarrow \mathbf{R}$, we define

$$\nabla f(x) = Df(x)^T$$

Example

$$f(x) = (1/2)x^T Px + q^T x$$

$$\nabla f(x) = Px + q$$

First-order approximation

$$f(y) \approx f(x) + \nabla f(x)^T (y - x)$$

(affine function of y)

Hessian

Hessian matrix (second derivative)

If $f(x) : \mathbf{R}^n \rightarrow \mathbf{R}$ second-order differentiable, we define

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

Example

$$f(x) = (1/2)x^T Px + q^T x$$

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

Second-order approximation

$$f(y) \approx f(x) + \nabla f(x)^T (y - x) + (1/2)(y - x)^T \nabla^2 f(x)(y - x)$$

(quadratic function of y)

Today's lecture

[Chapter 2 and 12, N and W][Chapter 4 and 5, B and V]

Optimality conditions for nonlinear optimization

- Unconstrained optimization
- Constrained optimization
- KKT optimality conditions
- Convex constrained nonconvex optimization

Unconstrained optimization

First-order necessary conditions

Fermat's Theorem



Theorem

If x^* is a local optimizer for the continuously differentiable function f , then

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

First-order necessary condition

Proof (contraposition)

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

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

Then, by Taylor approximation

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

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

$$f(y) < f(x^*)$$



Example: least-squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) = x^T A^T A x - 2x^T A^T b + b^T b$$

First-order optimality condition

$$\nabla f(x) = 2A^T(Ax - b) = 0$$



Normal-equations

$$A^T A x = A^T b$$

(they always
have
a solution)

Explicit solution

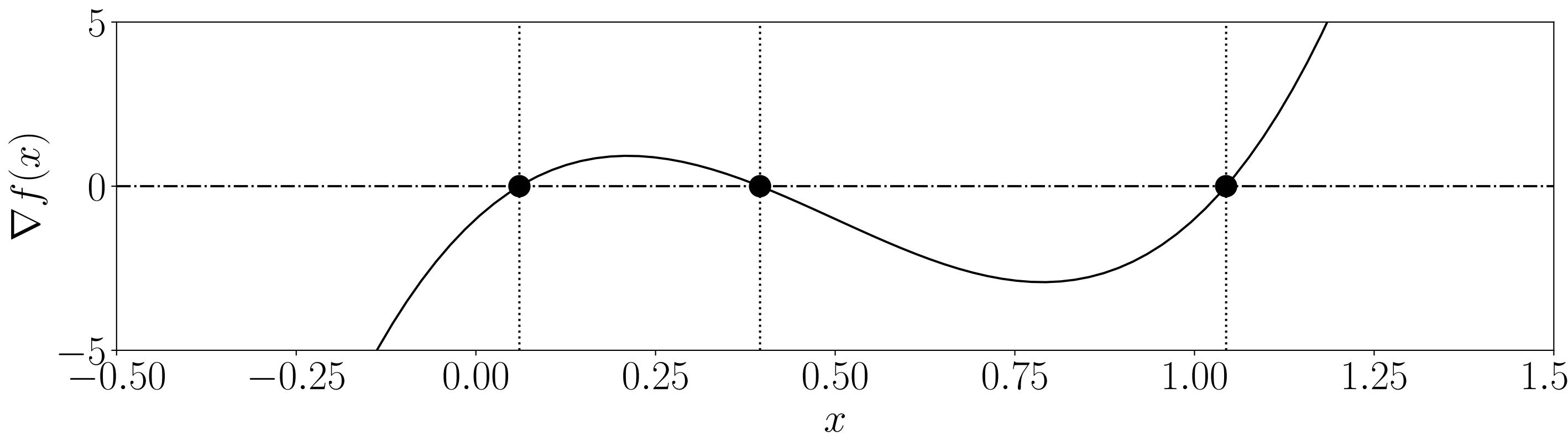
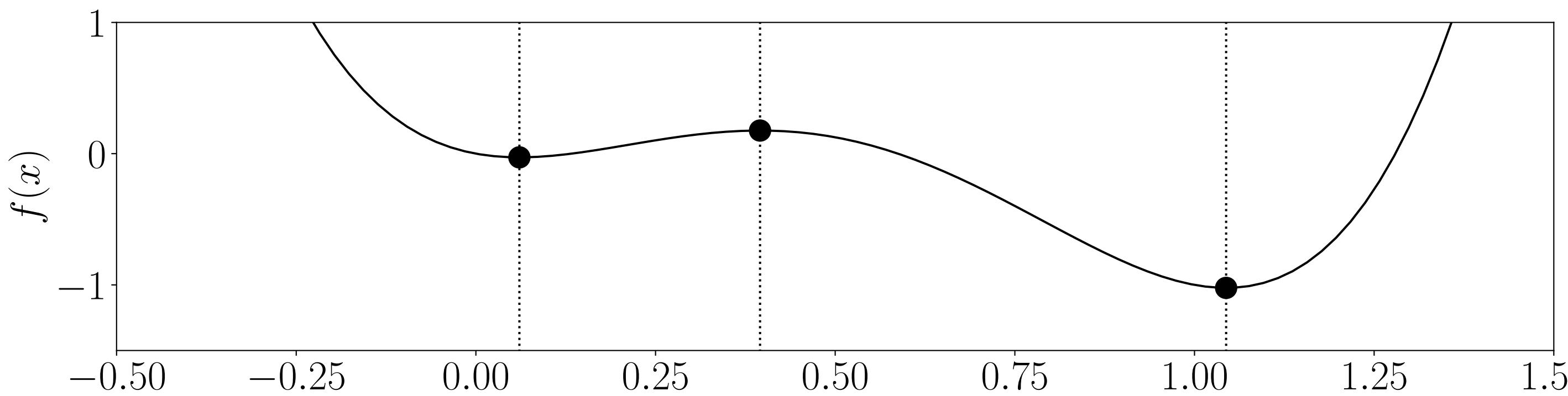
$$x^\star = (A^T A)^{-1} A^T b = A^\dagger b$$

(pseudoinverse)

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^* is a local optimizer for the continuously differentiable function f , then

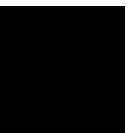
$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \succeq 0 \quad (\text{positive semidefinite})$$

Proof

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

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

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



Least-squares continued

$$\text{minimize} \quad \|Ax - b\|_2^2$$

$$f(x) = x^T A^T Ax - 2x^T A^T b + b^T b$$

First-order optimality condition

$$\nabla f(x) = 2A^T(Ax - b) = 0$$

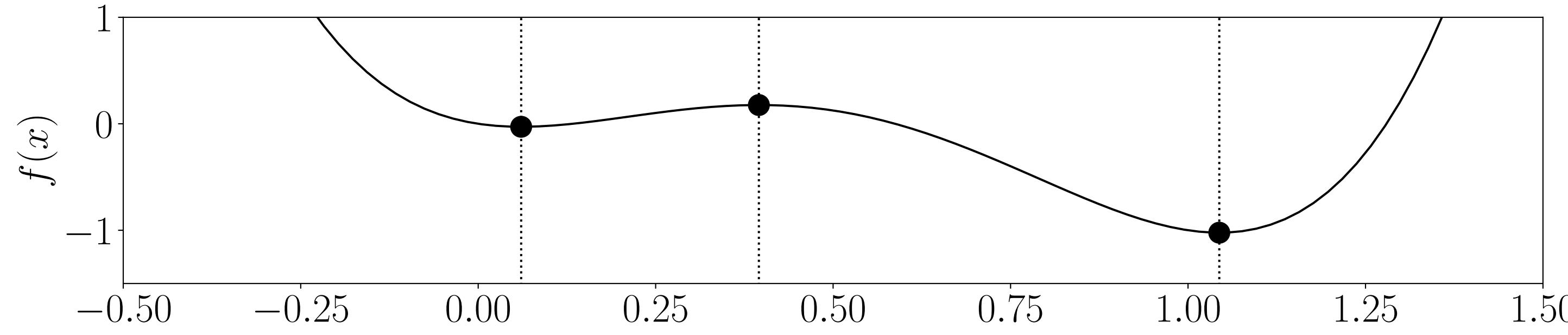
Explicit solution

$$x^* = (A^T A)^{-1} A^T b = A^\dagger b$$

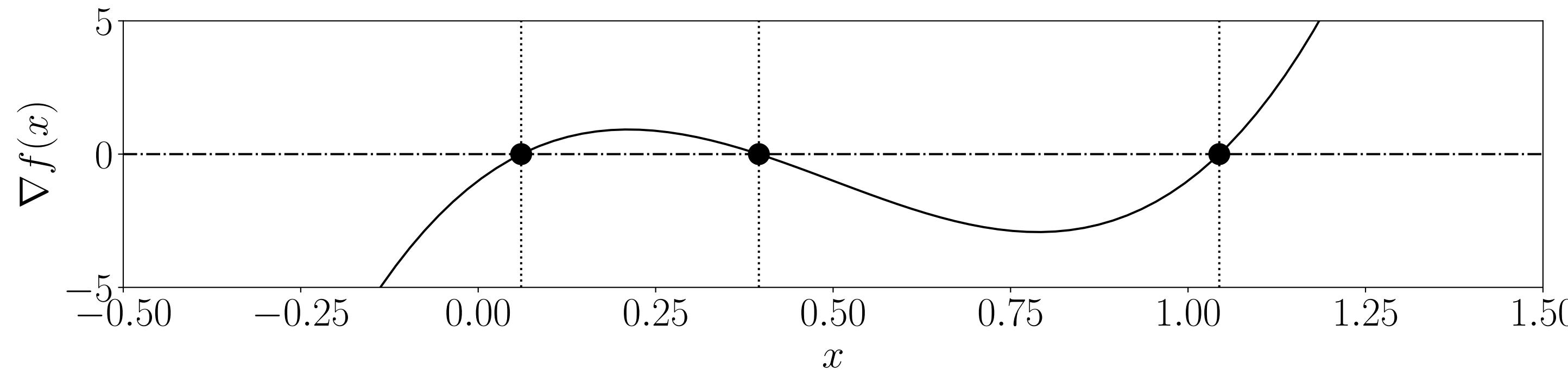
Second-order optimality condition

$$\nabla^2 f(x) = 2A^T A \succeq 0 \quad (\text{for any } A)$$

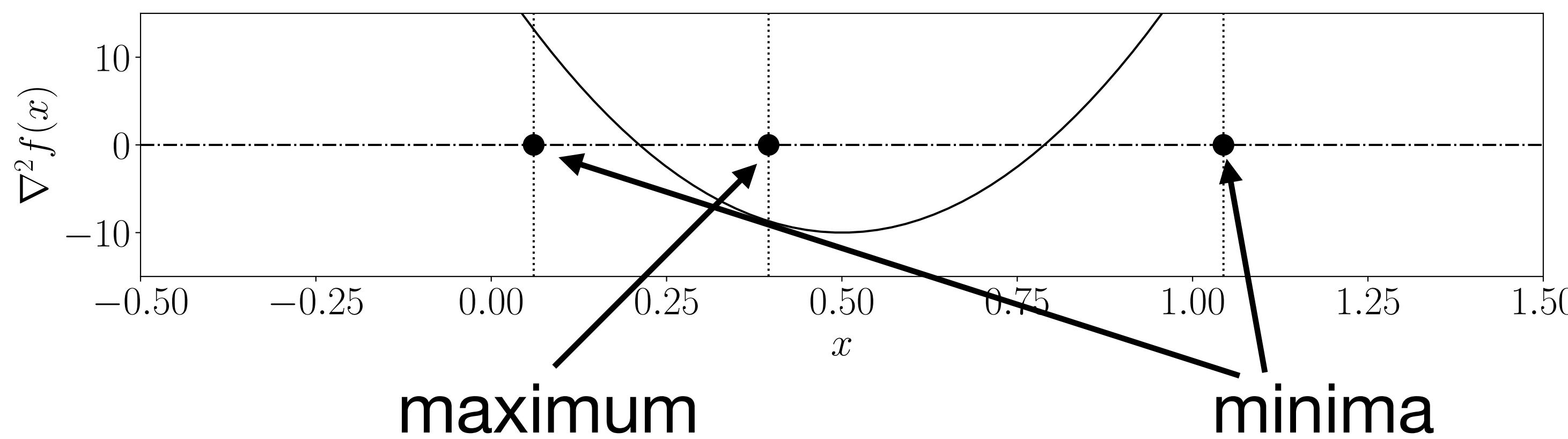
Example fixed



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

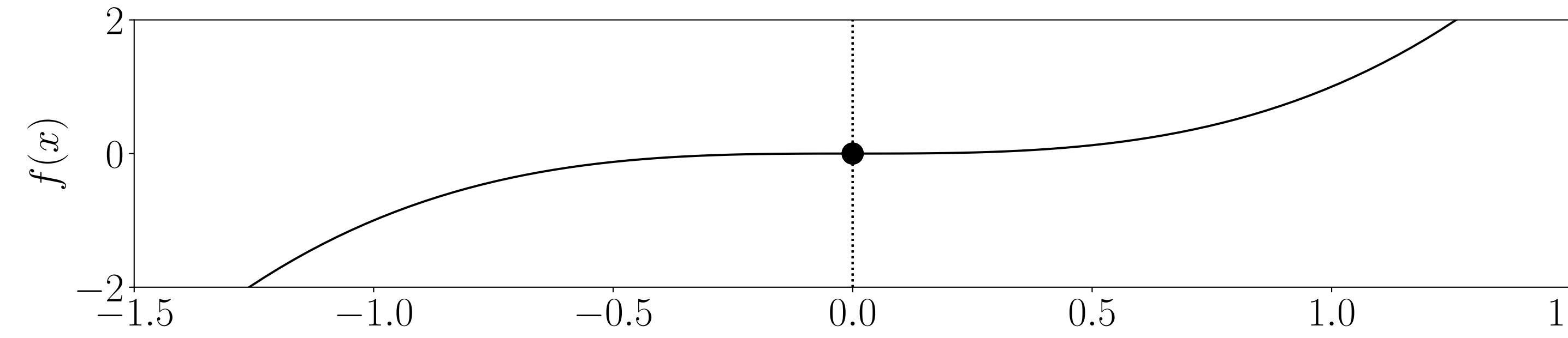


$$\nabla f(x) = 40x^3 - 60x^2 + 20x - 1$$



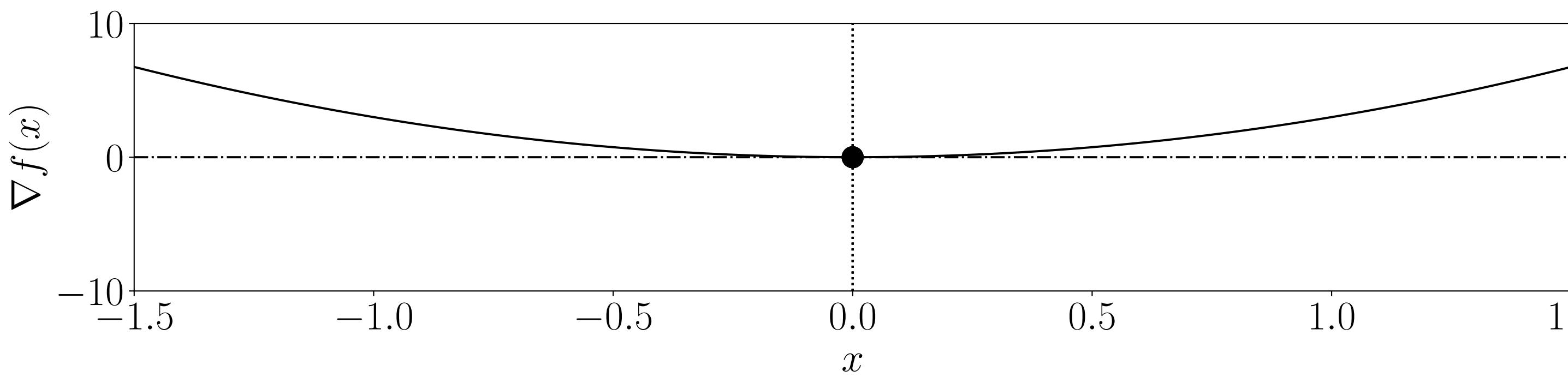
$$\nabla^2 f(x) = 120x^2 - 120x + 20$$

Second-order necessary condition is not sufficient



Cubic function

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

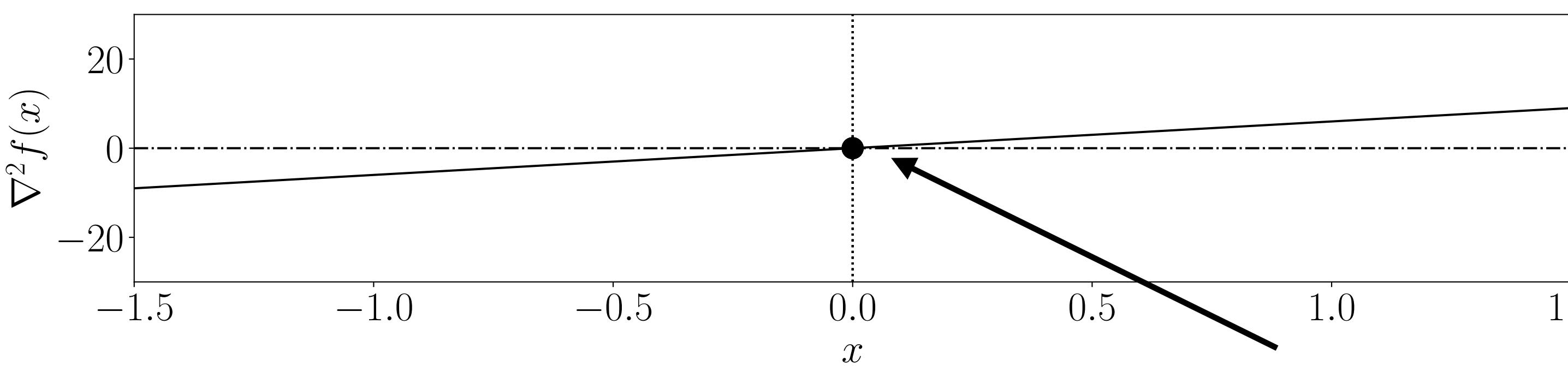


$$\nabla f(x) = 3x^2$$

Conditions satisfied

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

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



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

not local minimum

Second-order sufficient condition

Theorem

Let f be a continuously differentiable function. If x^* satisfies

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

then x^* is a local minimum of f

Proof

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

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

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

for any d



Examples

Cubic function

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

Least-squares

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

$2A^T A \succ 0$ 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

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

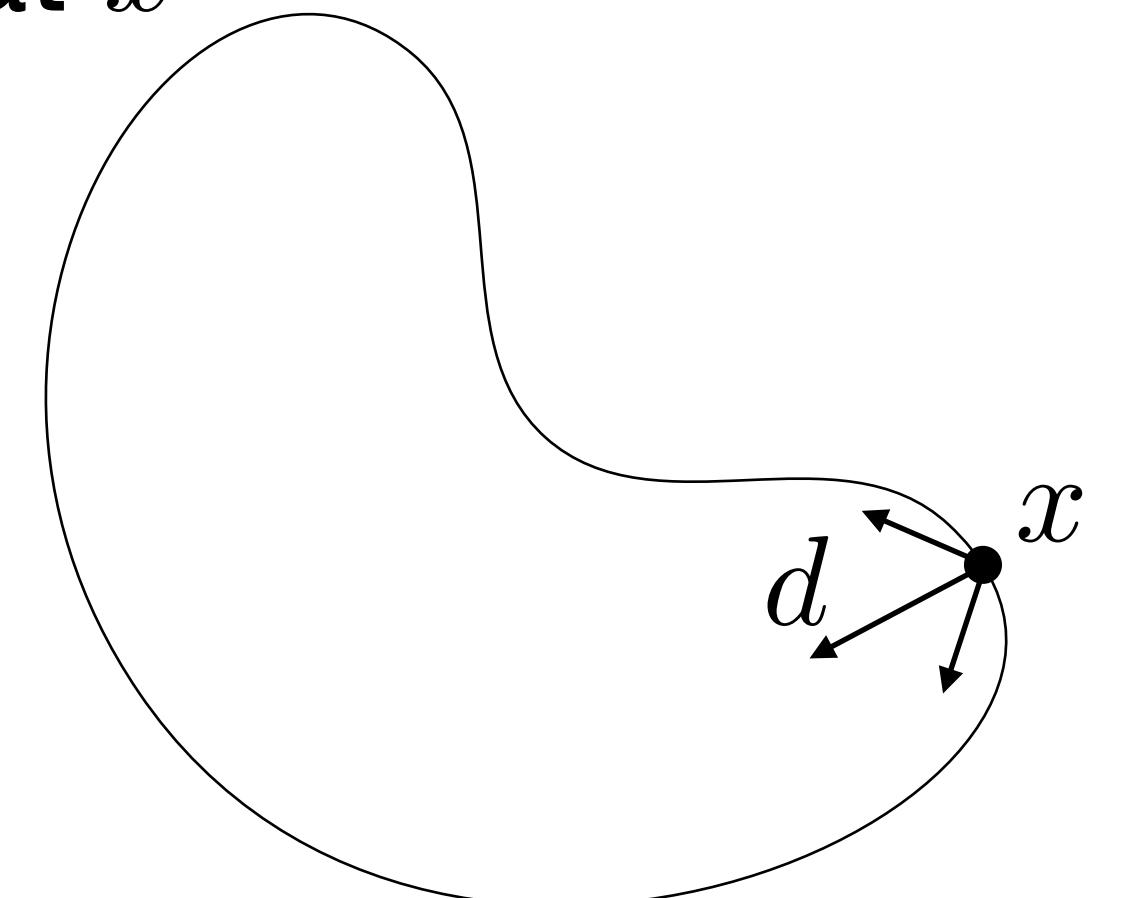
$F(x)$ is the **set of all feasible directions** at x

Examples

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

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

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



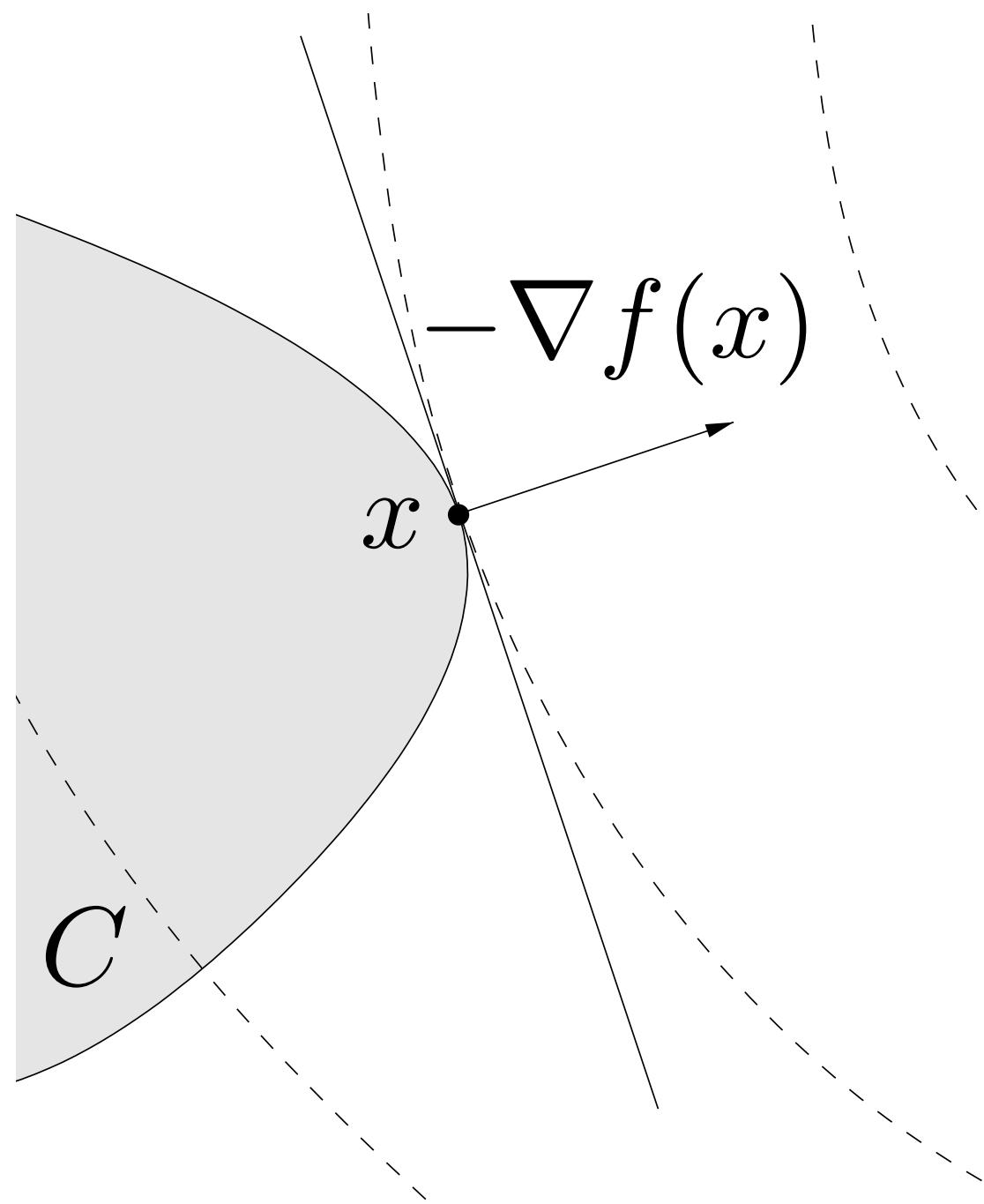
First-order necessary optimality condition

All feasible directions do not decrease the cost

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

Theorem

If x^* is a local minimum, then

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in F(x^*)$$


Unconstrained case

$F(x^*) = \mathbf{R}^n$, therefore $\nabla f(x^*) = 0$

Descent direction

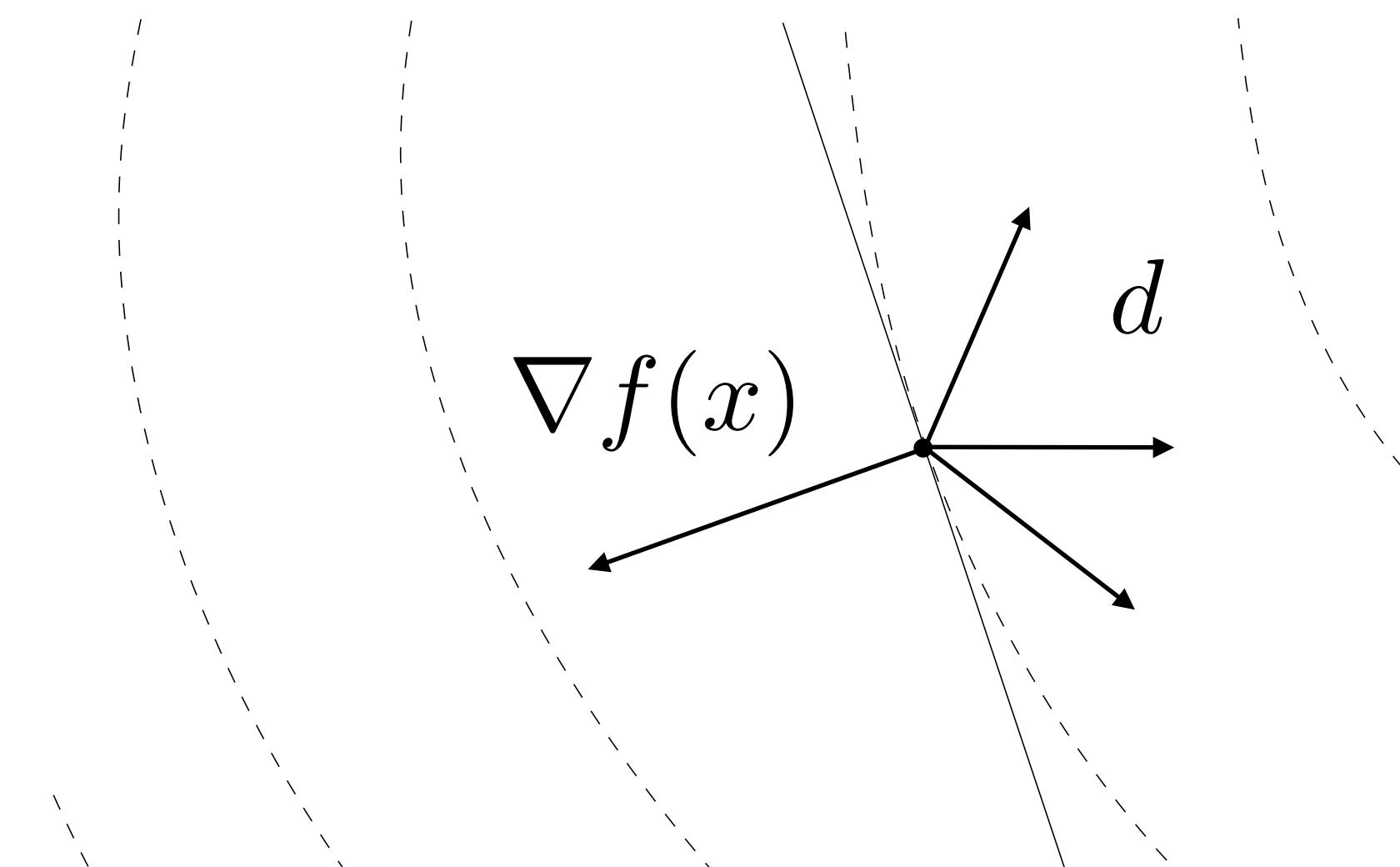
Given continuously differentiable f , we call d a **descent direction** at x if there exists \bar{t} such that

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

$D(x)$ 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^* is a local optimum, then

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

Nonlinear optimization with equality constraints

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

Theorem

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

Proof

Feasible directions

$$F(x) = \{d \mid Ad = 0\}$$

Descent directions

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

$F(x^*) \cap D(x^*) = \emptyset$ if and only if $\exists \nu$ such that $A^T \nu = \nabla f(x^*)$ (thm. of alternatives)

Let $y = -\nu$



Interpretation

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

(perpendicular
to
hyperplane)

Example: constrained least squares

$$\begin{array}{ll}\text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & Cx = d\end{array}$$

$$\begin{aligned}f(x) &= x^T A^T A x - 2x^T A^T b + b^T b \\ \nabla f(x) &= 2A^T(Ax - b)\end{aligned}$$

Optimality conditions

$$\text{Feasibility } Cx = d$$

$$\text{Optimality } 2A^T(Ax - b) + C^T y = 0$$



Linear system solution

$$\begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$

Necessary conditions for smooth nonlinear optimization

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

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

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

Theorem

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

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m y_i \nabla g_i(x^*) &= 0 \\ y_i g_i(x^*) &= 0, \quad i = 1, \dots, m \end{aligned}$$

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 (easy to show)

Let's show they are alternatives:

We can write 1. as $B\tilde{d} \leq 0$, $c^T \tilde{d} > 0$

where $B = [A \quad \mathbf{1}]$, $c = (0, \dots, 0, 1)$ and $\tilde{d} = (x, \epsilon)$

By Farkas lemma, we have the alternative $B^T u = c$, $u \geq 0$, equivalent to 2. ■ 27

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 $1^T u = 1$

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

Therefore,

$$u \geq 0, \quad 1^T u = 1$$

Necessary conditions for smooth nonlinear optimization

Proof (continued)

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

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

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

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

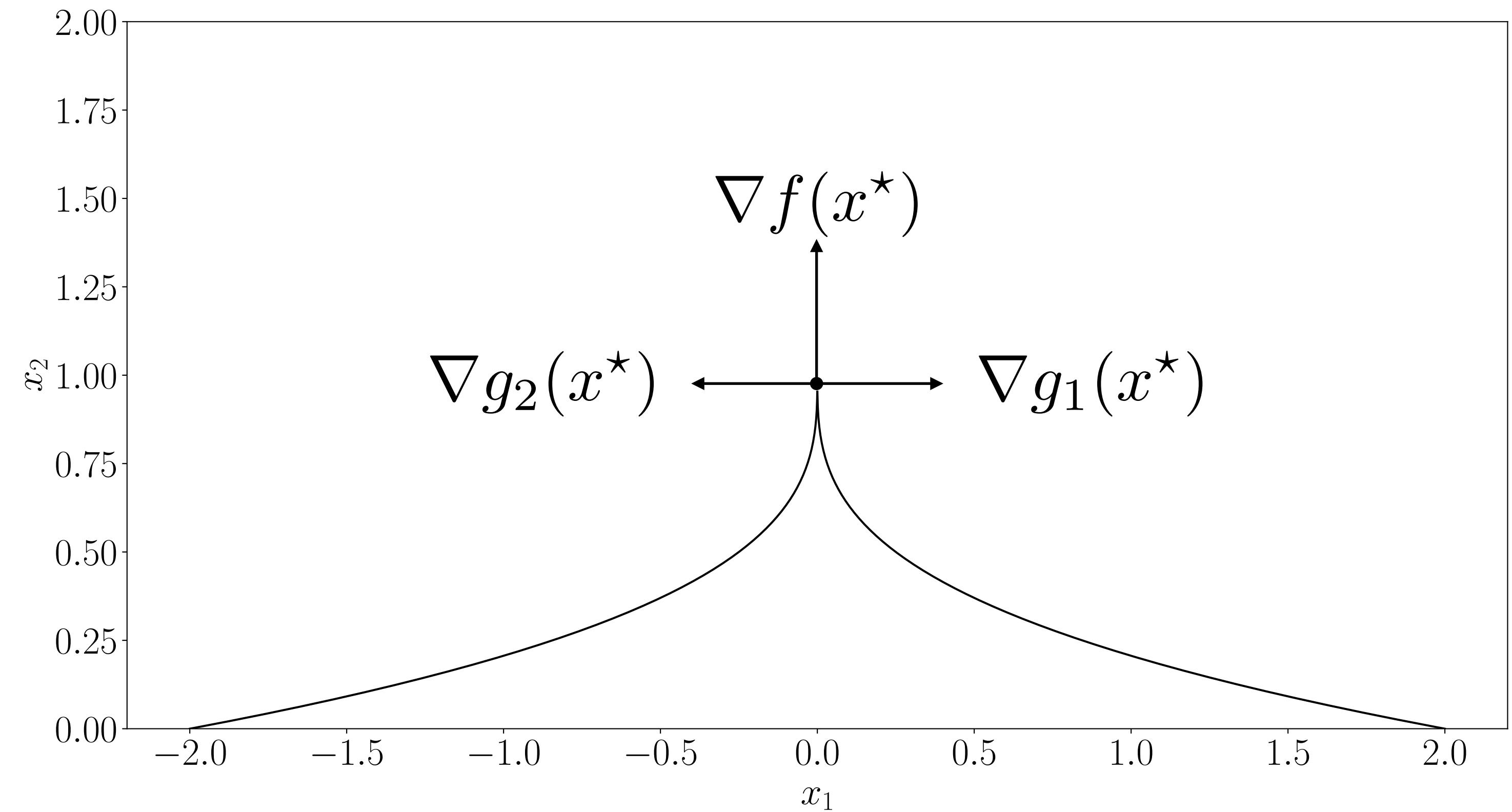
Which can be rewritten as $\nabla f(x^*) + \sum_{i=1}^m y_i \nabla g_i(x^*) = 0$

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



What happens if LICQ fails?

minimize $-x_2$
subject to $x_1 - 2(1 - x_2)^3 \leq 0$
 $-x_1 - 2(1 - x_2)^3 \leq 0$
 $x \geq 0$
 $x^* = (0, 1)$



Lagrangian function and duality

Lagrangian

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

Optimal cost
 $f(x^*) = p^*$

Lagrange multipliers

$$\begin{aligned}g_i(x) \leq 0 &\implies y_i \geq 0 \\h_i(x) = 0 &\implies v_i\end{aligned}$$

Lagrangian

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

Lagrangian Interpretation

Lower bound

$f(x) \geq L(x, y, v)$ for each feasible x

Proof

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

≤ 0 $= 0$

■

Dual function

$g(y, v) = \underset{x}{\text{minimize}} L(x, y, v)$

$\text{dom } g = \{(y, v) \mid g(y, v) > -\infty\}$

Lagrange dual problem

Finding the best lower bound

Always concave (-convex) problem

$$\begin{array}{ll} \text{maximize} & g(y, v) \\ \text{subject to} & y \geq 0 \end{array}$$



Dual problem

$$d^* = \max_{y \geq 0, v} \min_x L(x, y, v)$$

Lower bound condition always holds

Weak duality

$$d^* \leq p^*$$

Stationarity condition

minimize $f(x)$
subject to $g_i(x) \leq 0, \quad i = 1, \dots, m$
 $h_i(x) = 0, \quad i = 1, \dots, p$

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

Min-max formulation

$$p^* = \min_x \max_{y \geq 0, v} L(x, y, v) \quad (\text{minimize unconstrained version})$$

Stationarity condition on the Lagrangian

$$\nabla_x L(x, y, v) = \nabla f(x) + \sum_{i=1}^m y_i \nabla g_i(x) + \sum_{i=1}^p v_i \nabla h_i(x) = 0$$

KKT necessary conditions for optimality

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

Theorem

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

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

$$y^* \geq 0 \quad \textbf{dual feasibility}$$

$$g_i(x^*) \leq 0, \quad i = 1, \dots, m \quad \textbf{primal feasibility}$$

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

$$y_i^* g_i(x^*) = 0, \quad i = 1, \dots, m \quad \textbf{complementary slackness}$$

Strong duality theorem

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

Theorem

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

$$g_i(x) < 0, \quad i = 1, \dots, m, \quad (\text{for non-affine } g_i)$$

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

Slater's condition

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

Remarks

- For nonconvex optimization, we need harder conditions
- Generalizes LP conditions [Lecture 7]

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 **necessary and sufficient**

Example: KKT conditions for convex QP

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

Lagrangian

$$L(x, y, v) = (1/2)x^T Px + q^T x + y^T(Cx - d) + v^T(Ax - b) \quad \text{where } y \geq 0$$

Stationarity condition

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

Example: KKT conditions for convex QP

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

KKT Optimality conditions

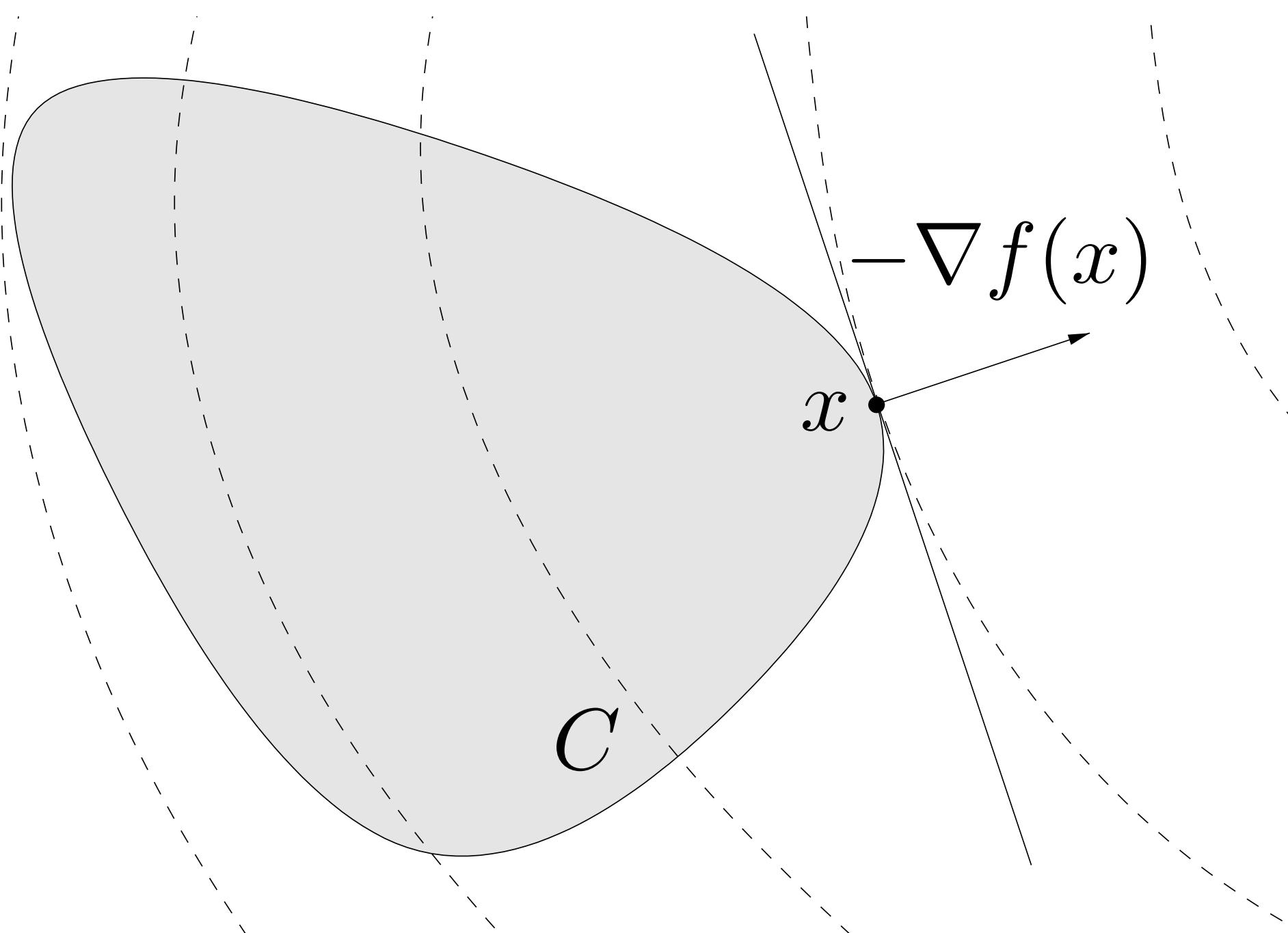
$Px^* + q + C^T y^* + A^T v^* = 0$	stationarity condition
$y^* \geq 0$	dual feasibility
$Ax - b = 0$	primal feasibility
$Cx - d \leq 0$	
$y_i(c_i^T x^* - d_i) = 0, \quad i = 1, \dots, m$	complementary slackness

Convex constrained nonconvex optimization

Minimization over convex set

[Section 3.7.3 and Example 3.74, A. Beck]

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \xleftarrow{\hspace{1cm}} \text{convex set} \end{array}$$



First-order optimality condition

If x^* is a local minimum, then

$$\nabla f(x^*)^T(y - x^*) \geq 0, \quad \forall y \in C$$

(f can be nonconvex)

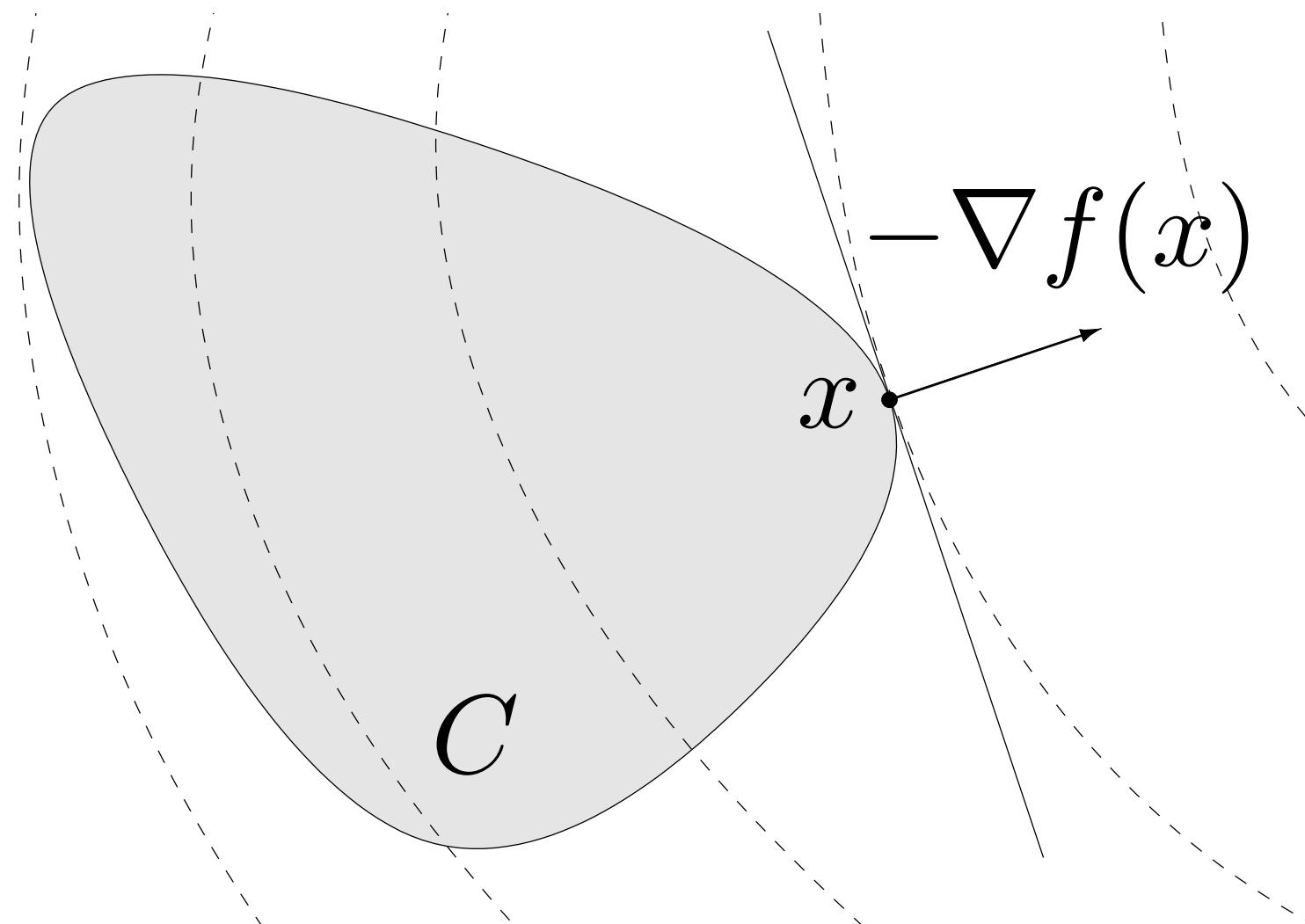
Why do you need a convex set?

First-order necessary optimality condition

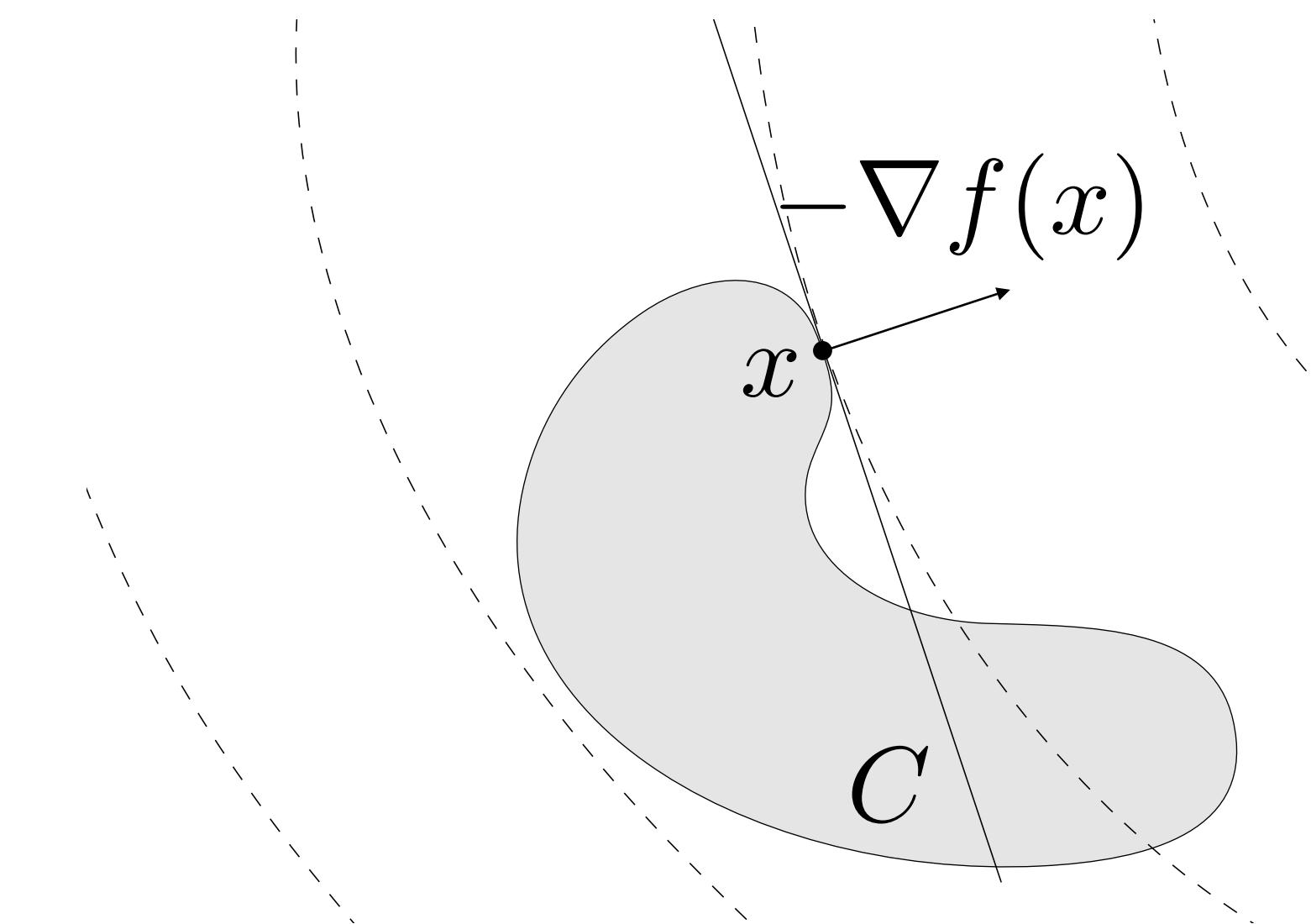
If x^* is a local minimum, then

$$\nabla f(x^*)^T(y - x^*) \geq 0, \quad \forall y \in C$$

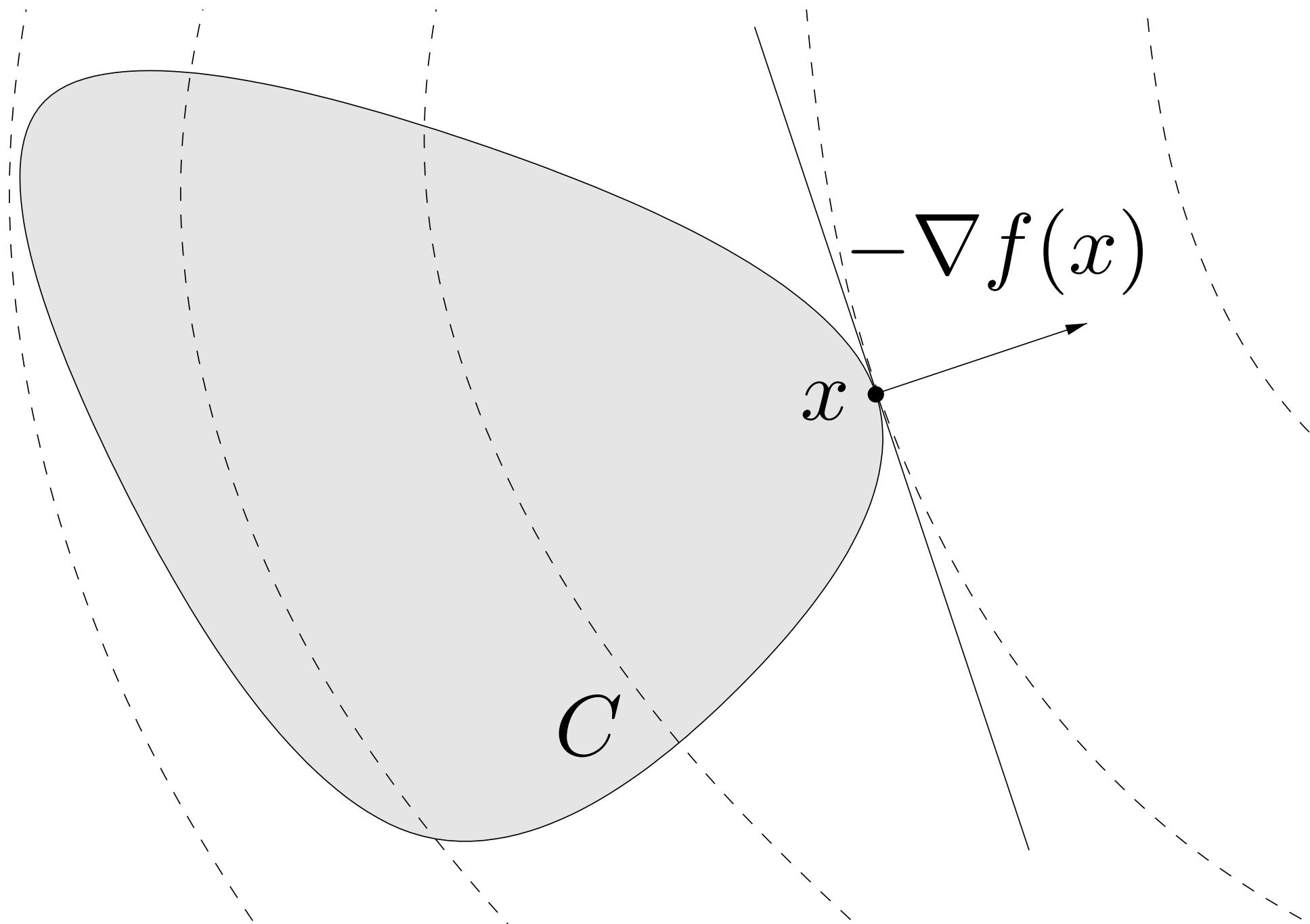
Convex set



Nonconvex set



Normal cone condition



First-order necessary optimality condition

If x^* is a local minimum, then

$$\nabla f(x^*)^T(y - x^*) \geq 0, \quad \forall y \in C$$

Normal cone

$$\mathcal{N}_C(x) = \{g \mid g^T(y - x) \leq 0, \quad \text{for all } y \in C\}$$

Reformulated condition

$$-\nabla f(x^*) \in \mathcal{N}_C(x^*)$$

Remark

If f and C are convex, then it is

necessary and sufficient

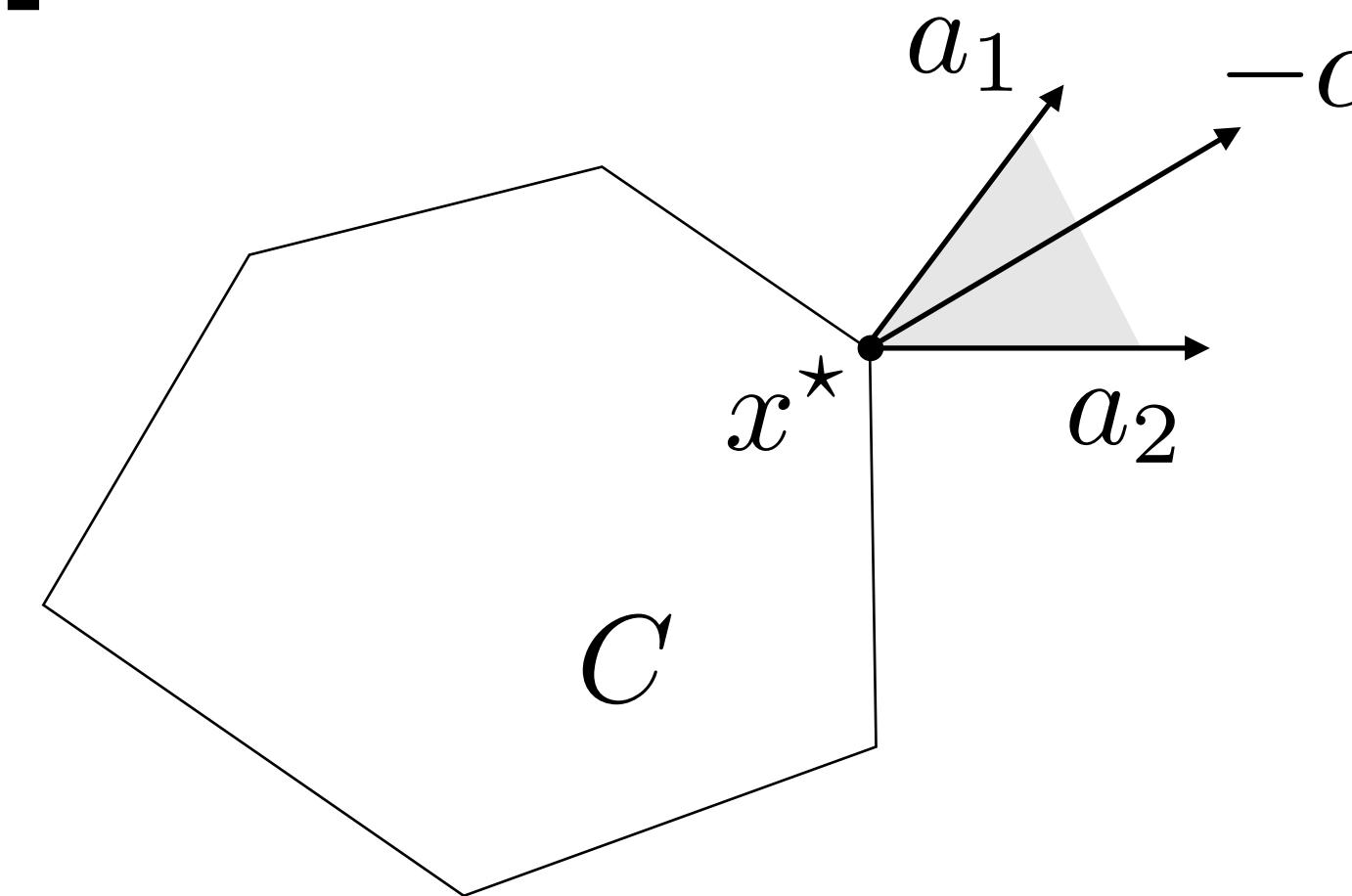
[Section 4.2.3, B and V]

Normal cone condition

Linear program example

minimize $c^T x$

subject to $Ax \leq b$



Recap from Lecture 8

Two active constraints at optimum: $a_1^T x^* = b_1$, $a_2^T x^* = b_2$

Optimal dual solution y satisfies:

$$A^T y + c = 0, \quad y \geq 0, \quad y_i = 0 \text{ for } i \neq \{1, 2\}$$

In other words, $-c = a_1 y_1 + a_2 y_2$ with $y_1, y_2 \geq 0$

Normal cone to polyhedron

$$-c \in \mathcal{N}_{\{Ax \leq b\}}(x^*) = \{A^T y \mid y \geq 0 \text{ and } y_i(a_i^T x^* - b_i) = 0\}$$

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 using Farkas lemma
- **Derive** optimality conditions for constrained optimization using Lagrangian
- **Apply** normal cone to derive necessary first-order conditions for nonconvex optimization over convex set

Next lecture

- Optimization algorithms: iteratively solve first-order optimality conditions