ORF522 – Linear and Nonlinear Optimization

13. Optimality conditions for nonlinear optimization

Bartolomeo Stellato — Fall 2021

Ed forum

- Normal cone: What are the angles? Isn't it translated? (Answer in next slide)
- Midterm: for interior point methods, we can use the Sherman-Woodbury-Morrison to update the matrix factorization in each iteration quickly
  - False! $\boxed{ADA^T}$ for diagonal D, $D_{ii} = \frac{y_i}{s_i}$
  - Interior point methods are second order methods
    - Expensive, but high-quality iterations $O(n^3)$
    - Very few iterations - in practice, usually 25 or so

Interior Point Methods are 2nd Order Methods

find root $h(x)=0$

Newton's method $$h(x^k) + Dh(x^k)(x^{k+1} - x^k) = 0$$

Interior Point Methods are 2nd Order Methods

Newton's method $h(x^k) + Dh(x^k)(x^{k+1} - x^k) = 0$

Smoothed problems

$$\begin{array}{ll} \text{minimize} & c^T x - \tau \sum_{i=1}^m \log(s_i) \\ \text{subject to} & Ax + s = b \end{array}$$

Interior Point Methods are 2nd Order Methods

Newton's method $$h(x^k) + Dh(x^k)(x^{k+1} - x^k) = 0$$

Smoothed problems

$$\begin{array}{ll} \text{minimize} & c^T x - \tau \sum_{i=1}^m \log(s_i) \\ \text{subject to} & Ax + s = b \end{array}$$

Optimality conditions

$$\begin{array}{l} Ax + s = b \\ A^T y + c = 0 \\ s_i y_i = \tau \quad i = 1, \ldots, m \end{array} \qquad h(x, s, y) = 0$$

Interior Point Methods are 2nd Order Methods

Newton's method $h(x^k) + Dh(x^k)(x^{k+1} - x^k) = 0$

Smoothed problems

$$\begin{array}{ll} \text{minimize} & c^T x - \tau \sum_{i=1}^m \log(s_i) \\ \text{subject to} & Ax + s = b \end{array}$$

Optimality conditions

$$\begin{array}{l} Ax + s = b \\ A^T y + c = 0 \\ s_i y_i = \tau \quad i = 1, \ldots, m \end{array} \qquad h(x, s, y) = 0$$

- We will mostly focus on first order methods for the rest of the course

Upcoming Lectures

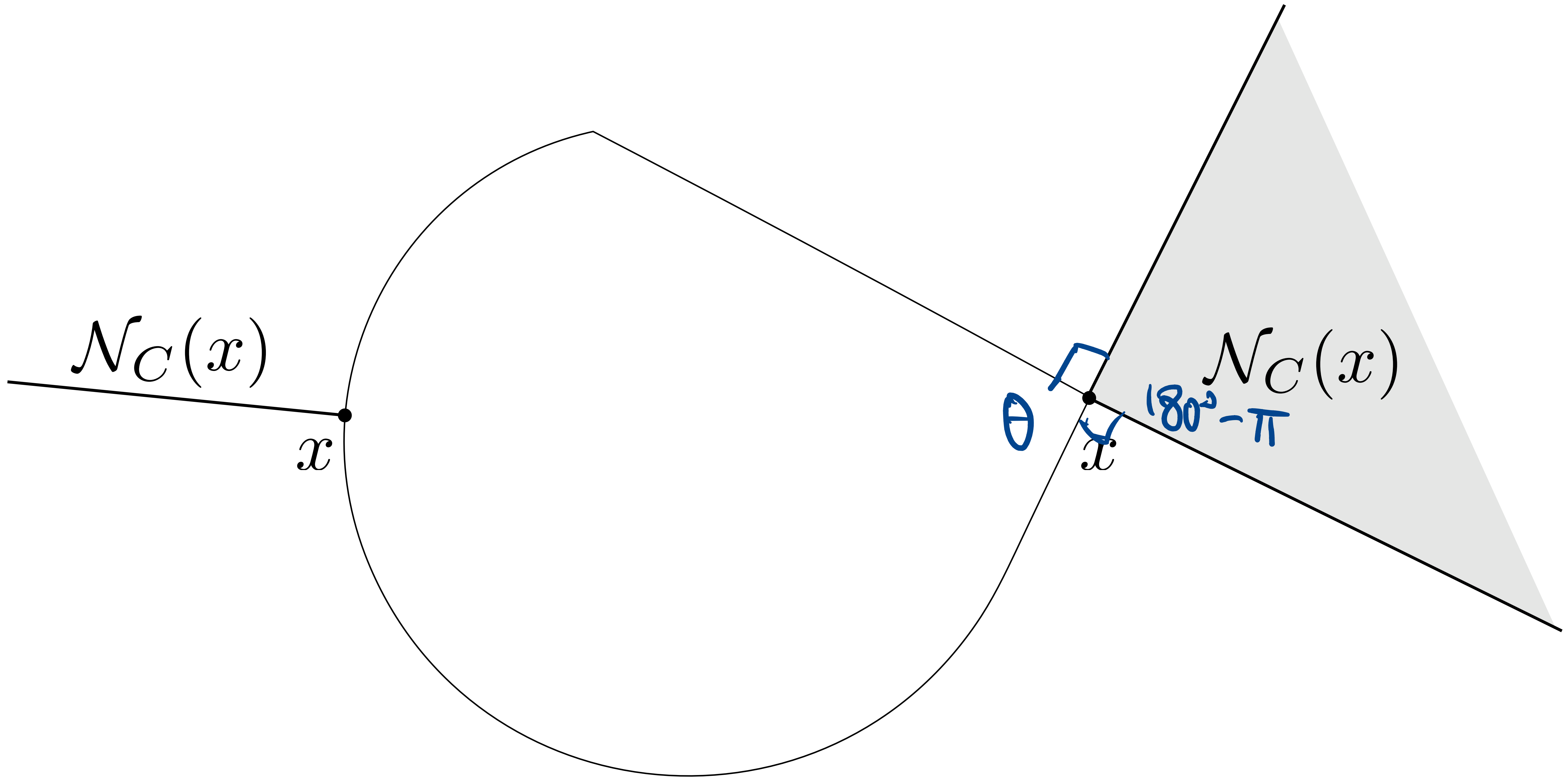
13	10/26	Optimality conditions	3 Out	[Ch 2 and 12, NO] [Ch 4 and 5, CO]
14	10/28	Gradient descent		[Ch 1 and 2, ICLO] [Ch 9, CO] [Ch 5, FMO]
15	11/02	Subgradient methods	3 Due	[Ch 3 and 8, FMO] [ee364b] [Ch 3, ILCO]
16	11/04	Proximal methods and intro to operator theory		[Ch 3 and 6, FMO] [PA] [PMO]
17	11/09	Operator theory	4 Out	[Ch 4, FMO] [PA] [PMO] [LSMO]
18	11/11	Operator splitting algorithms		[PMO] [PA] [LSMO] [ADMM]
19	11/16	Acceleration schemes	4 Due	[Ch 1, FMO] [Ch 2, ILCO] [Ch 3, COAC]

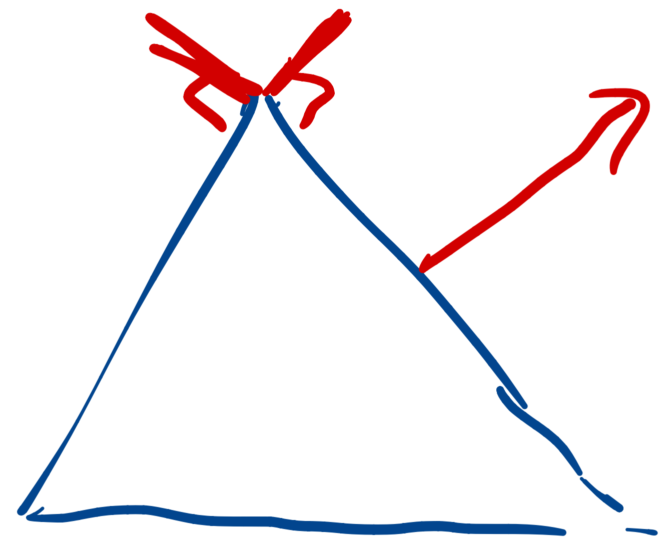
Recap

Normal cone

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

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





Gradient

Derivative

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

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

Example

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

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

First-order approximation

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

(affine function of y)

Hessian

$\nabla^2 f(x)$ is symmetric

– eigenvalues are real

$S = Q \Lambda Q^T$

$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$

Q orthonormal

$Q^T Q = I$

Hessian matrix (second derivative)

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

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

Example

$$f(x) = (1/2)x^T P x + q^T x$$
$$\nabla^2 f(x) = P$$

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]

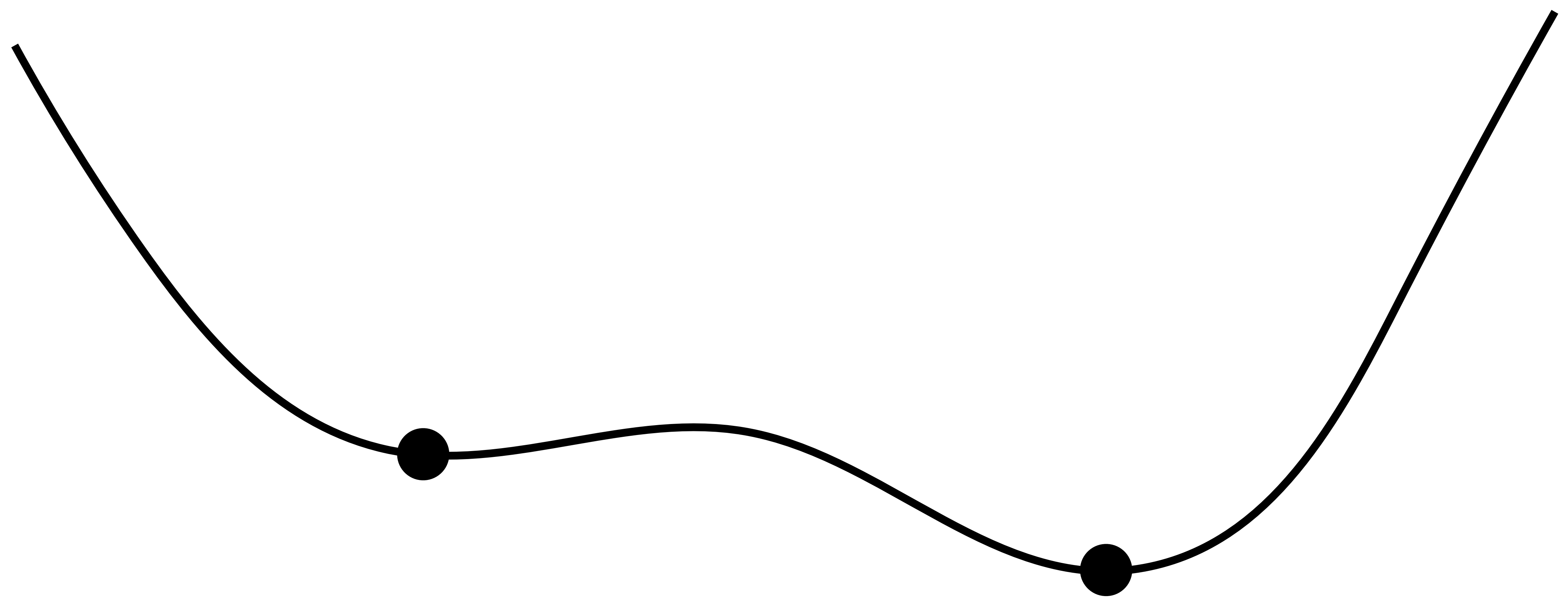
Optimality conditions for nonlinear optimization

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

Unconstrained optimization

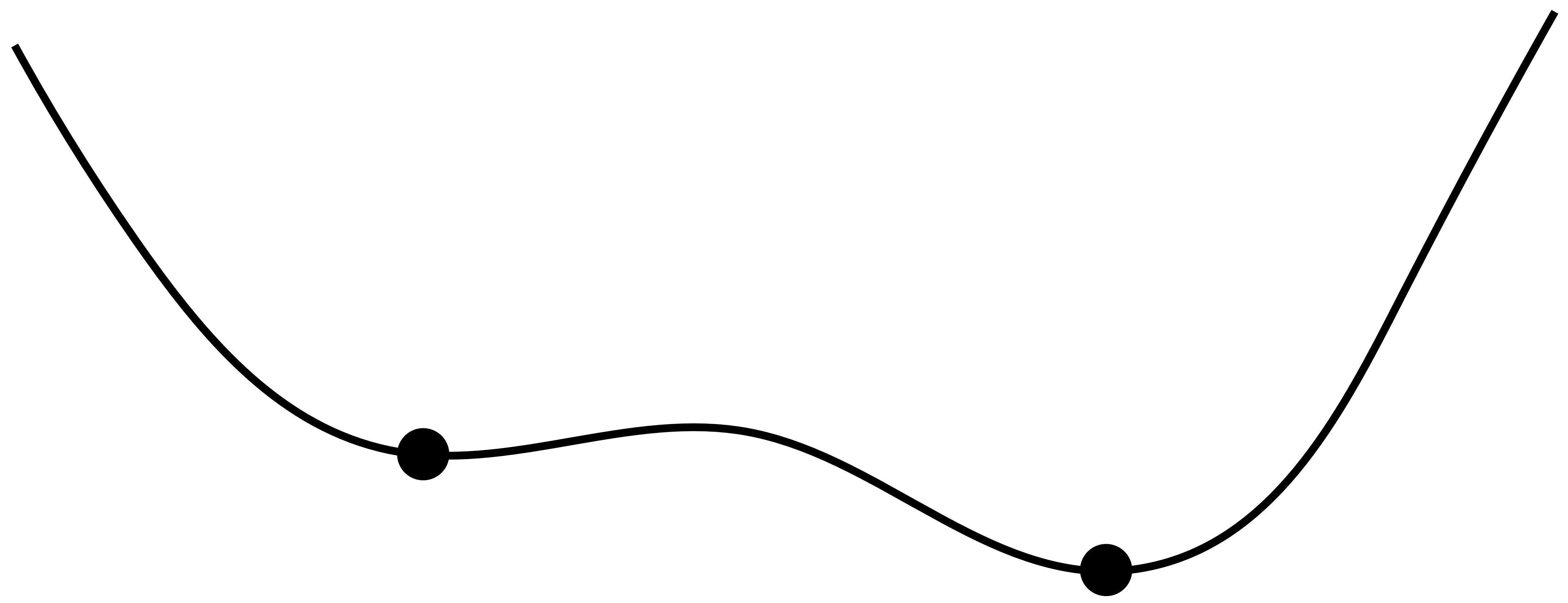
First-order necessary conditions

Fermat's Theorem



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,

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

First-order necessary condition

Proof (contraposition)

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

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

$y = x^\star + td$

Then, by Taylor approximation

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

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

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

$\blacksquare$

Example: least-squares

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

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

Example: least-squares

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

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

First-order optimality condition

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

[Matrix Cookbook, https://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf]

Example: least-squares

minimize $\|Ax - b\|_2^2$

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

First-order optimality condition

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

$\longrightarrow$

Normal-equations

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

(they always have a solution)

[Matrix Cookbook, https://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf]

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

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

$\longrightarrow$

Normal-equations

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

(they always have
a solution)

Explicit solution

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

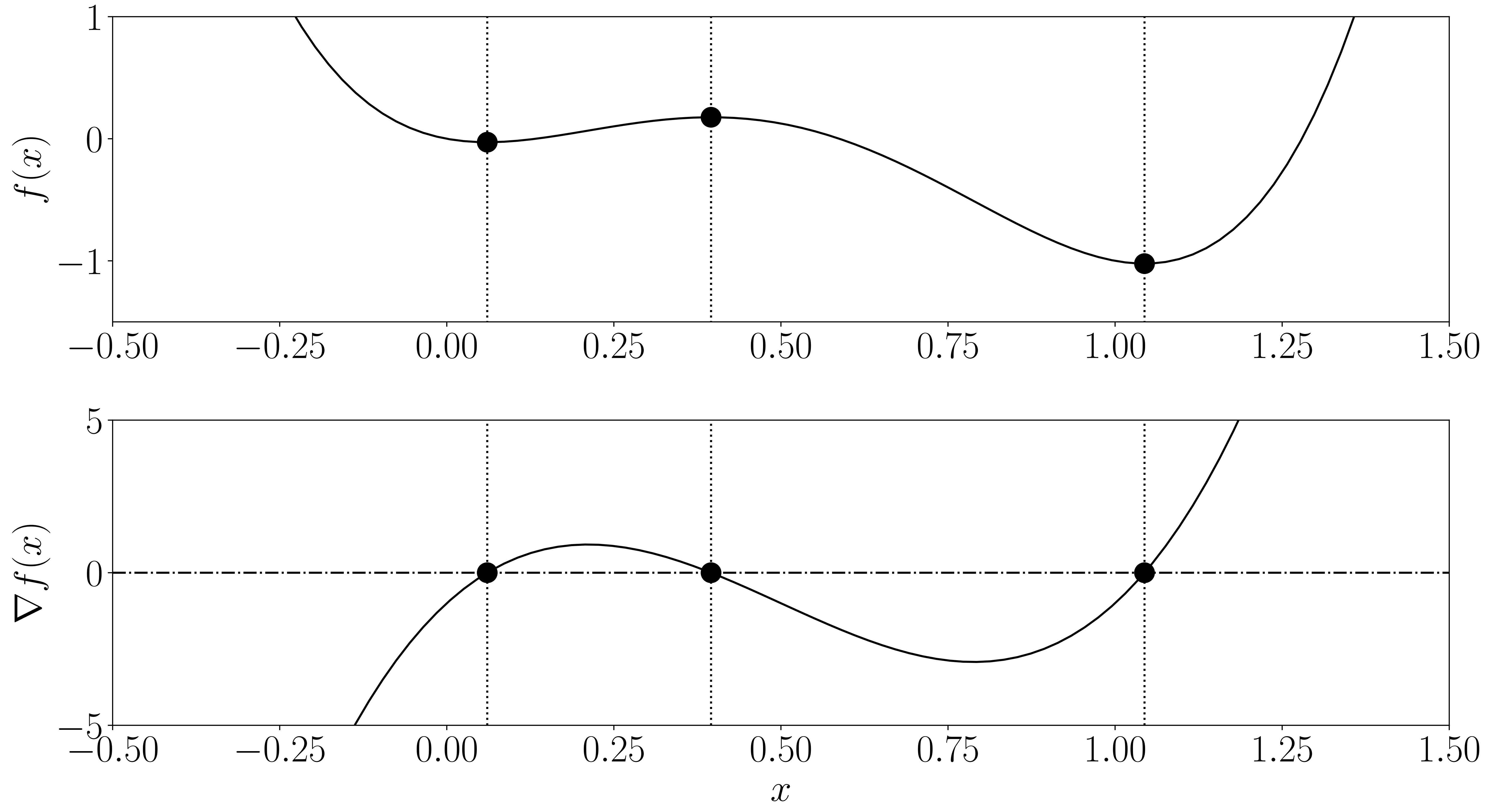
(pseudoinverse)

[Matrix Cookbook, https://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf]

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

$v^T A^T A v = \|Av\|_2^2 \geq 0$

H PSD means $v^T H v \geq 0 \quad \forall v \in \mathbb{R}^n$

$H = A^T A$

Theorem

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

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

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

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

Second-order necessary condition

x^* local min $\Rightarrow \nabla f(x^*)=0,\ \nabla^2 f(x^*) \succeq 0$

$\Leftarrow$?

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 $\blacksquare$

Least-squares continued

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

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

First-order optimality condition

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

Explicit solution

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

Least-squares continued

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

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

First-order optimality condition

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

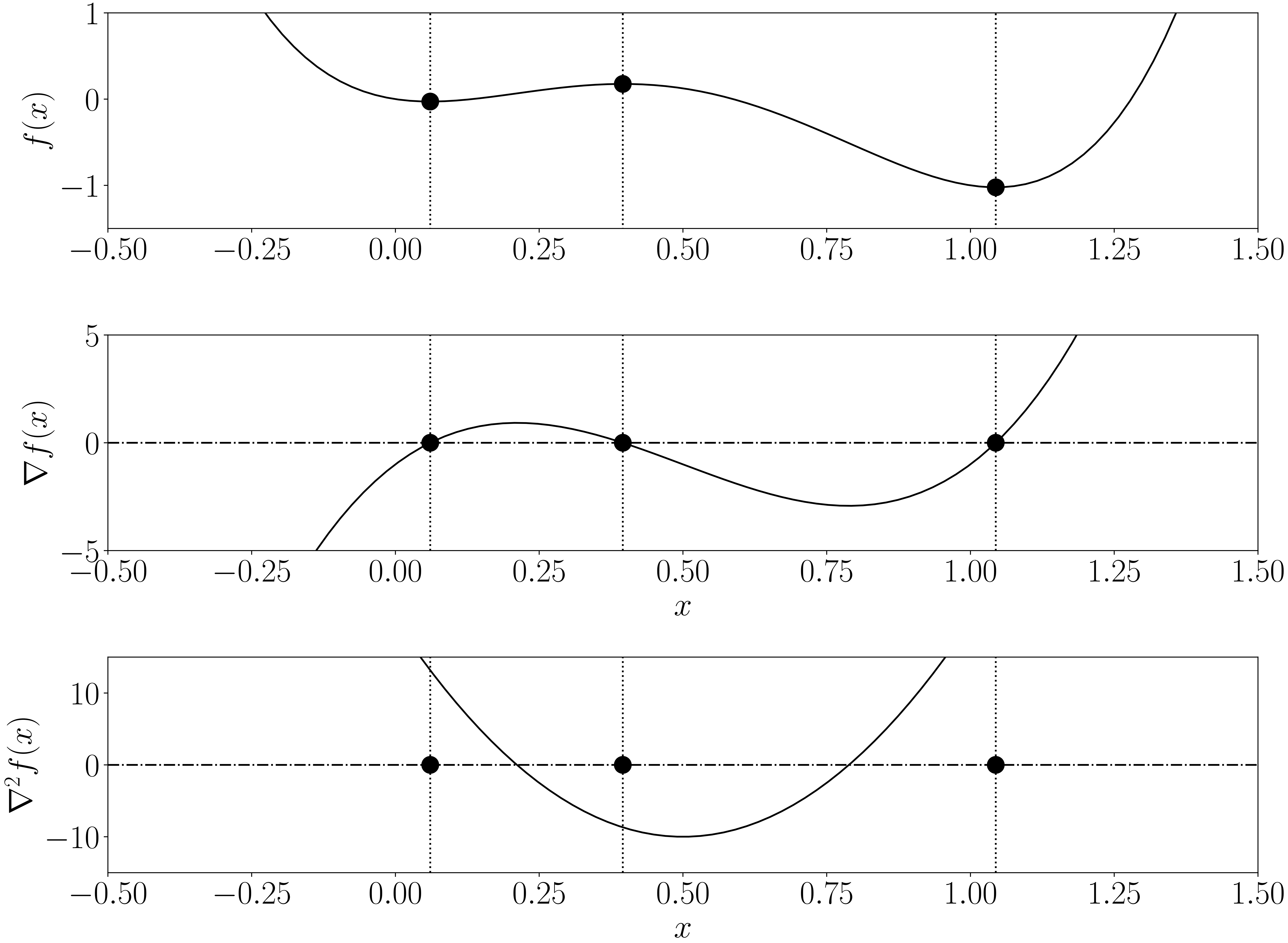
Explicit solution

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

Second-order optimality condition

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

Example fixed



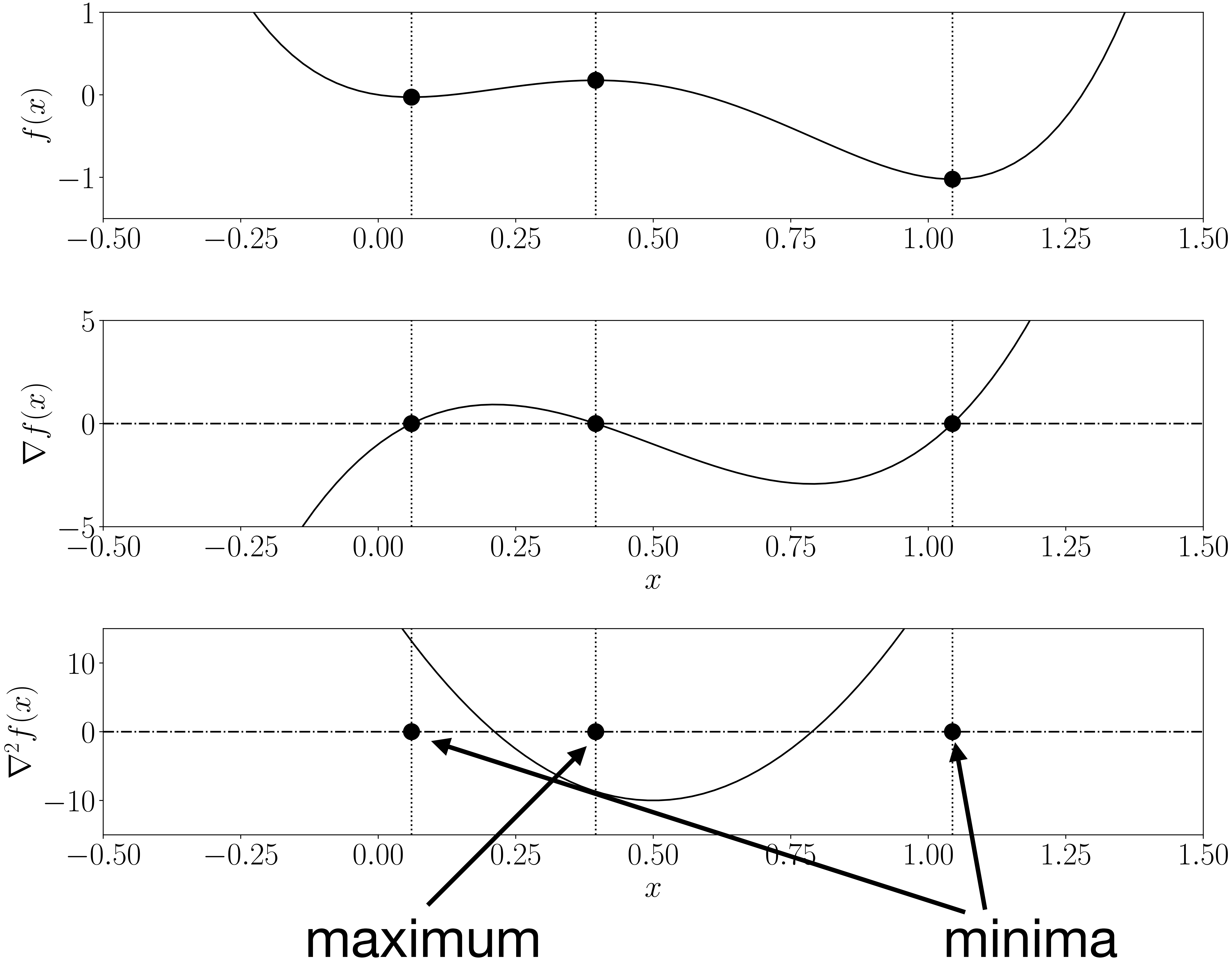
$$f(x) = 10x^2(1-x)^2 - x$$

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

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

Converse counterexample?

Example fixed



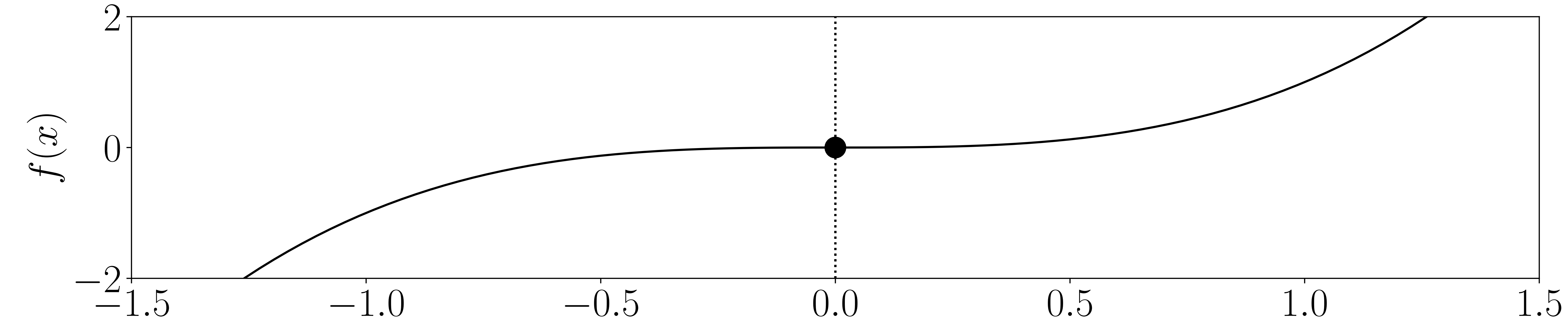
$$f(x) = 10x^2(1-x)^2 - x$$

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

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

Converse counterexample?

Second-order necessary condition is not sufficient

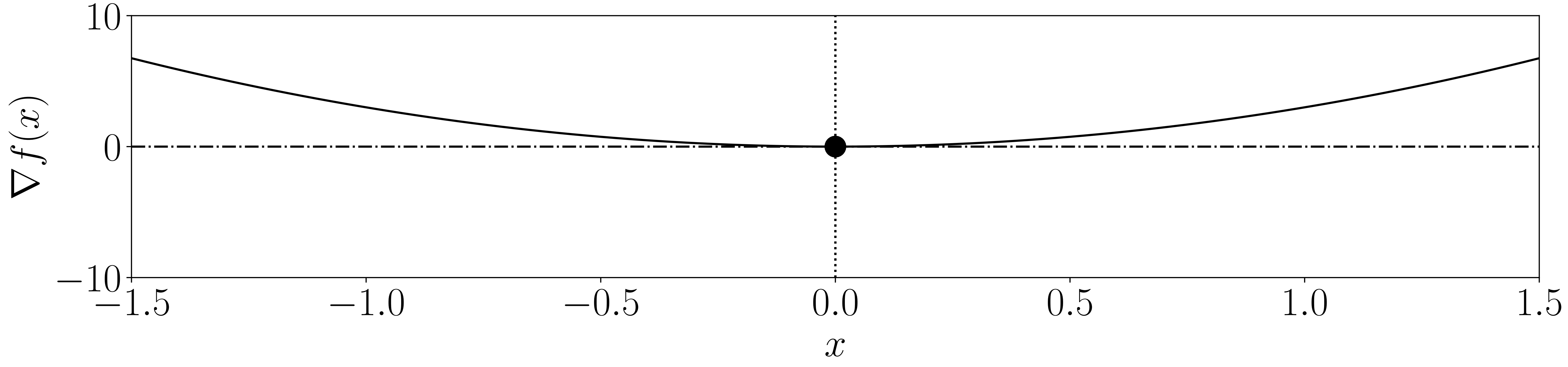


Cubic function

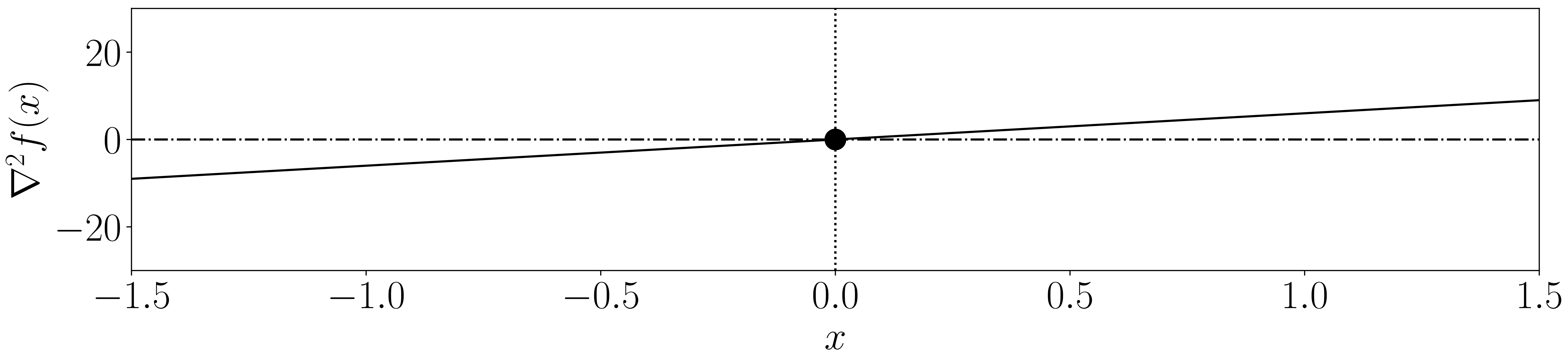
$$f(x) = x^3$$

at $x = 0$,

$\nabla f(x) = 0$, $\nabla^2 f(x) = 0$

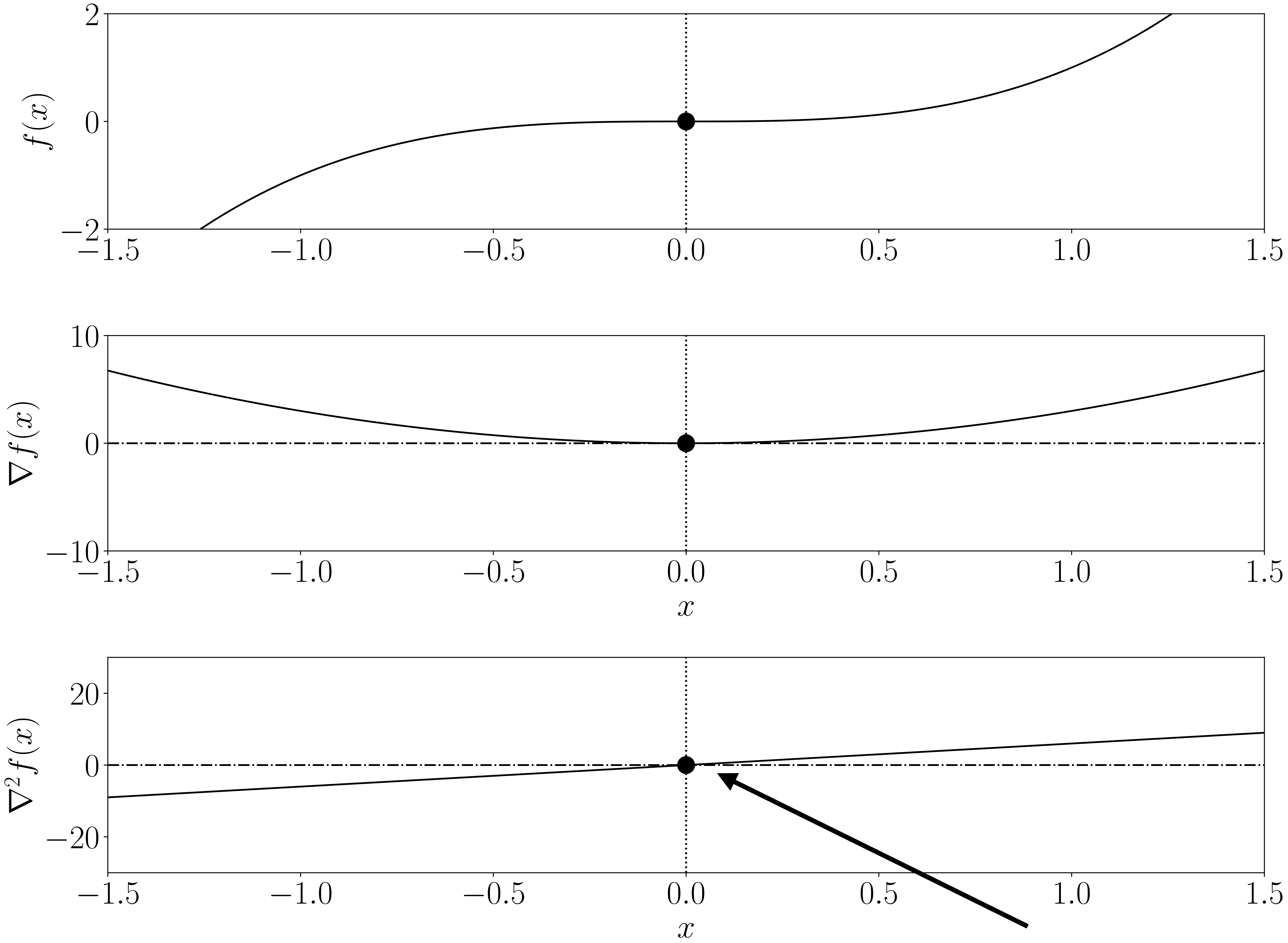


$$\nabla f(x) = 3x^2$$



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

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

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

$\exists \lambda > 0 \quad \text{s.t.} \quad \nabla^2 f(x^*) - \lambda I \succeq 0$

Proof

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

Second-order sufficient condition

$\nabla f(x^*)=0$, $\nabla^2 f(x^*)\succ 0 \Rightarrow x^*$ strict local min

$\Leftarrow$?

$f(x)=x^4$

$\nabla^2 f(x)=12x^2$

$\nabla f(x)=4x^3$

at zero

zero is a strict local min

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 strict 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

$$f(\underbrace{x^\star + td}_{y}) = f(x^\star) + t^2 (1/2) \underbrace{d^T \nabla^2 f(x^\star) d}_{>0} + o(t^2) > f(x^\star)$$

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)

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

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

$2A^T A \succ 0$ 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

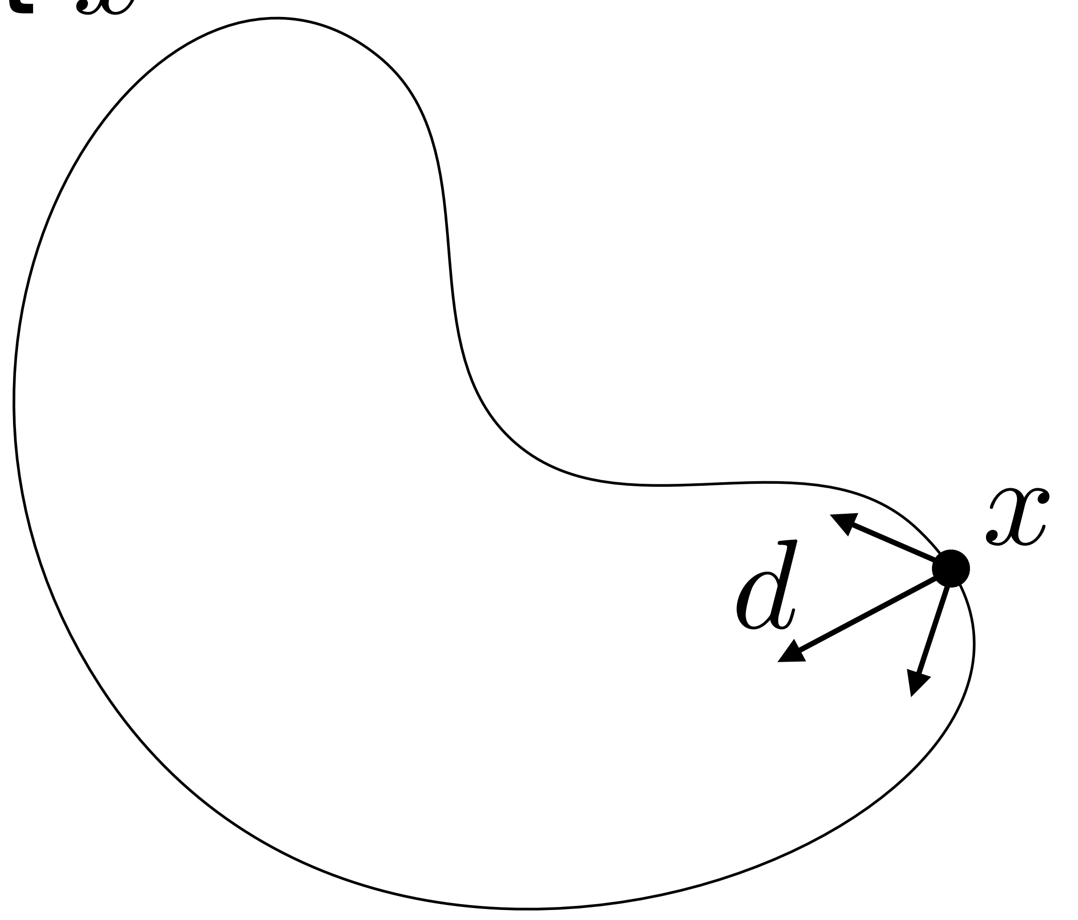
Feasible direction

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

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



$A(x+td) = b$

$b + tAd = b$ $\quad Ad = 0$

$g_i(x) = x^2$

look at 0

$\nabla g_i(x) = 2x$

$\nabla g_i(0) = 0$

Examples

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

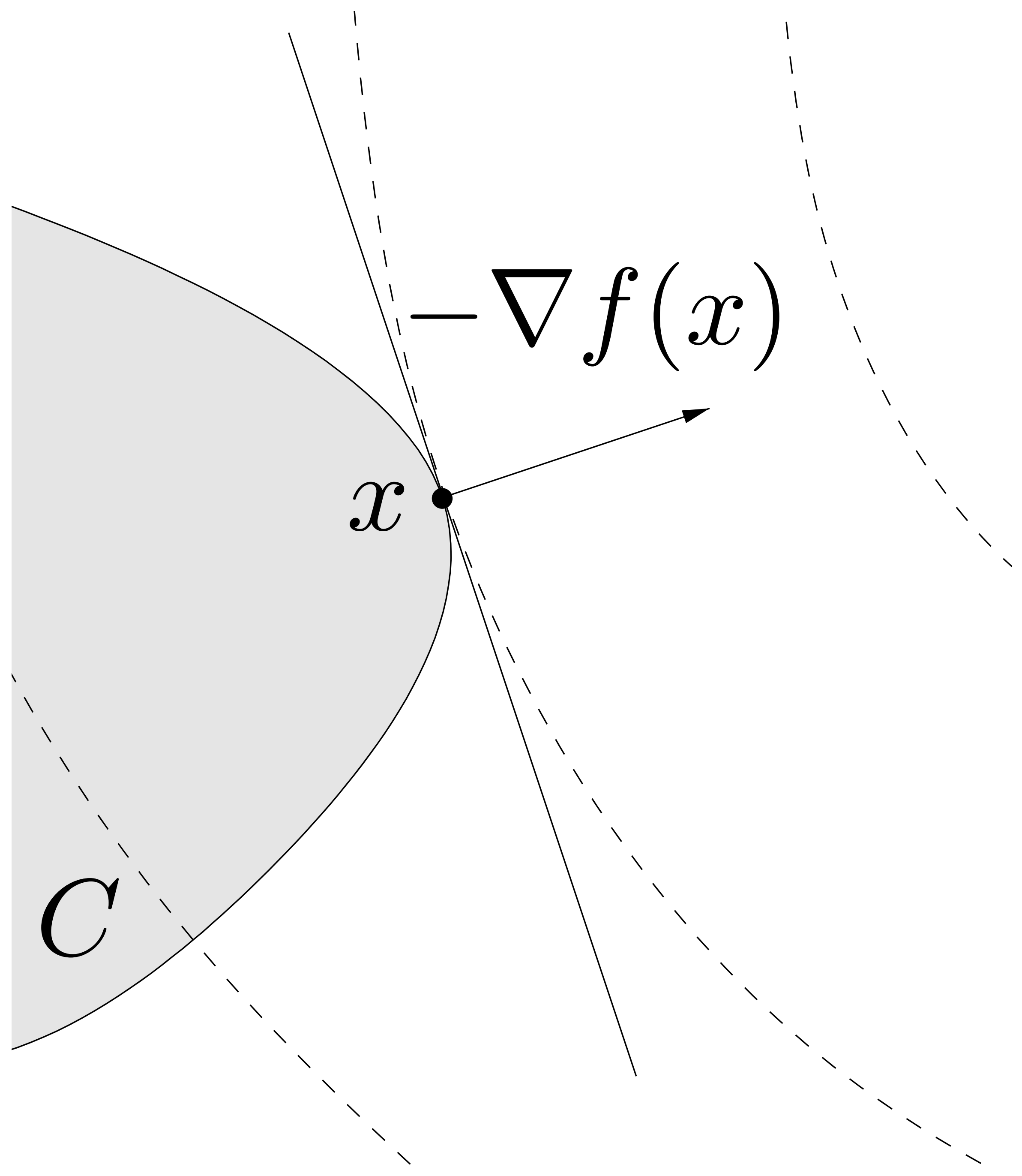
$C = \{Ax \leq b\} \implies F(x) = \{d \mid a_i^T d \leq 0 \quad \text{if } a_i^T x = b_i\}$

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

First-order necessary optimality condition

All feasible directions do not decrease the cost

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



Theorem

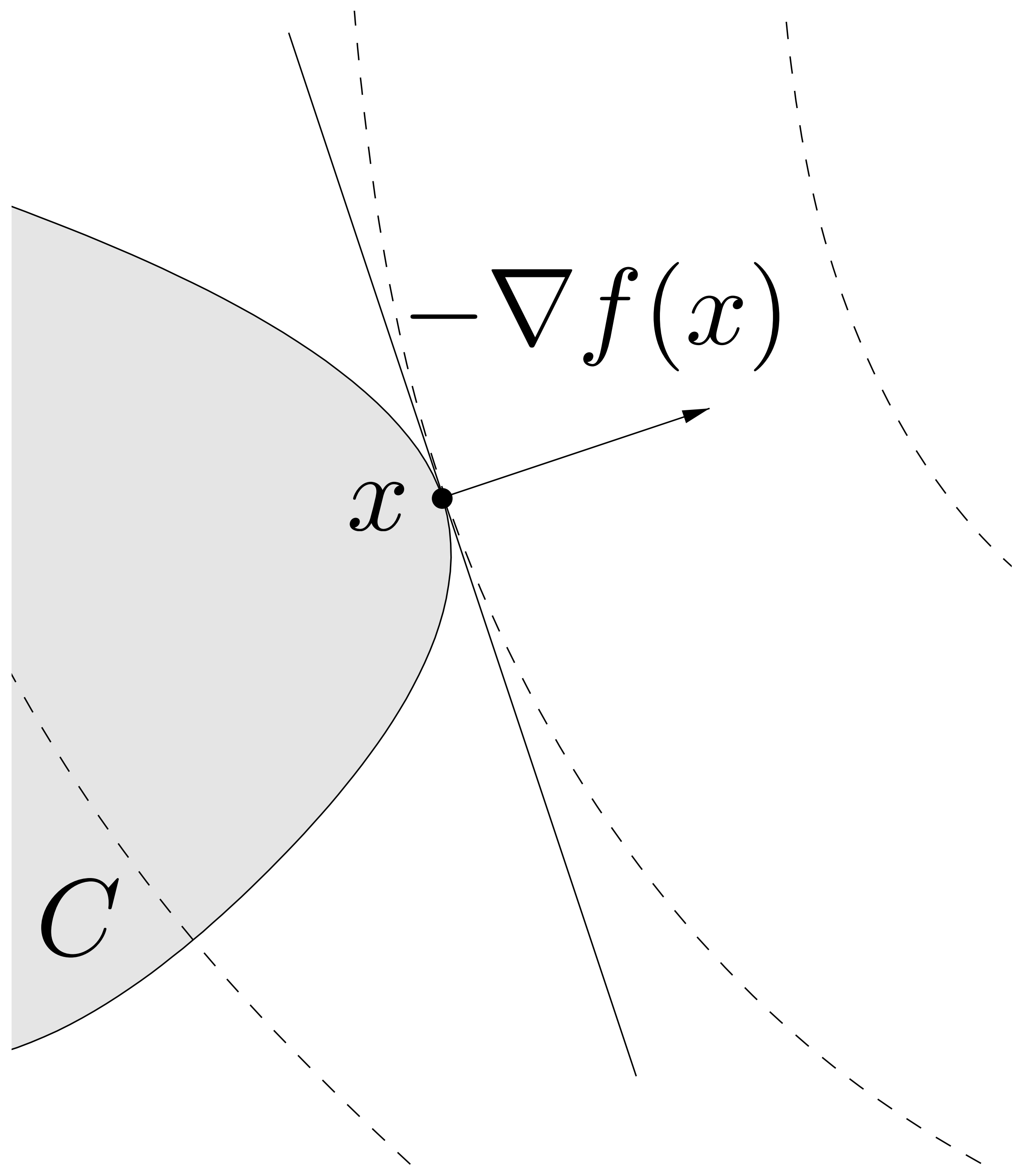
If $x^\star$ is a local minimum, then

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

First-order necessary optimality condition

All feasible directions do not decrease the cost

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



Theorem

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

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

Unconstrained case

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

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**

Descent direction

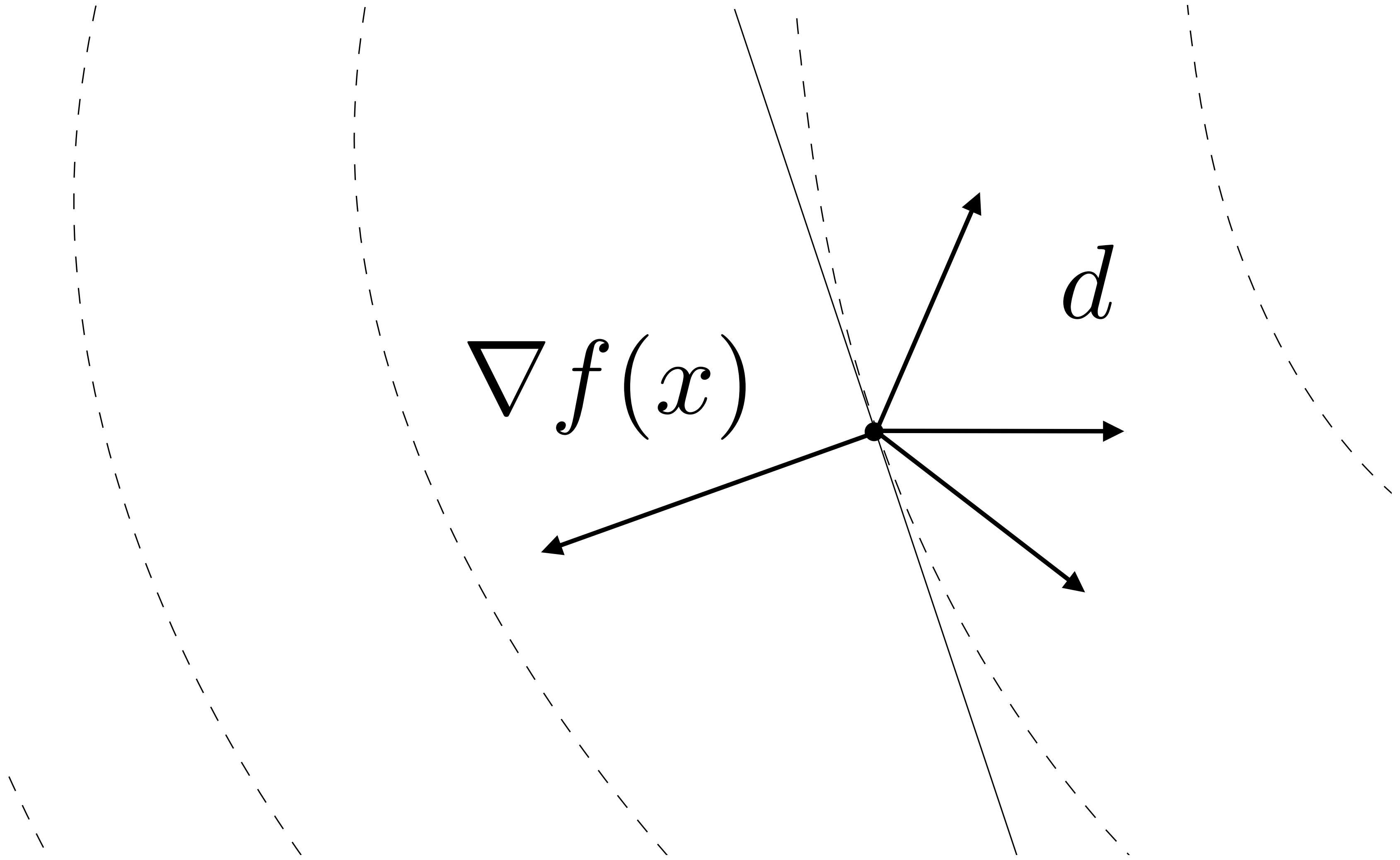
Given continuously differentiable f, we call d a **descent direction** at x if there exists $\bar{t}$ such that

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

$D(x)$ 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

$$\downarrow$$

There is no feasible descent direction

Converse false

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

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

Proof

Feasible directions

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

Descent directions

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

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

Let $y = -\nu$

$\blacksquare$

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

Proof

$\nu = \nu^+ - \nu^-, \ \nu^+ \geq 0, \ \nu^- \geq 0$

$A^T(\nu^+ - \nu^-) = \nabla f(x^*)$

$(A^T \ -A^T)\begin{pmatrix} \nu^+ \\ \nu^- \end{pmatrix} = \nabla f(x^*)$

Feasible directions

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

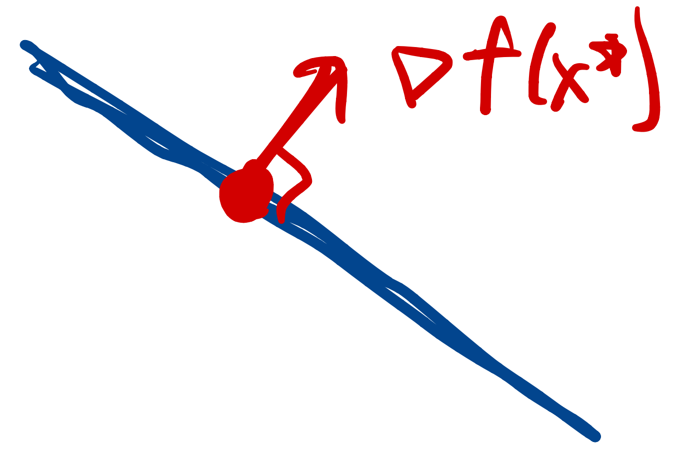
Descent directions

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

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

Let $y = -\nu$ ∎

$Ax = 0$



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

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

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

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

Optimality conditions

Feasibility $\quad Cx = d$

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

Example: constrained least squares

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

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

Optimality conditions

Feasibility $\quad Cx = d$

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

$\longrightarrow$

Linear system solution

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

Regularity conditions for invertibility

Necessary conditions for smooth nonlinear optimization

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

($g_i(x)$ nonlinear)

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

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

Necessary conditions for smooth nonlinear optimization

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

($g_i(x)$ 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

$\{\nabla g_i(x), \quad i \in \mathcal{A}(x)\}$ is **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 $u \geq 0$

Let's show they they are alternatives:

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

where $B = \begin{bmatrix} A & \mathbf{1} \end{bmatrix}$, $c = (0, \ldots, 0, 1)$ and $\tilde{d} = (d, \epsilon)$

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

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 they are alternatives:

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

where $B = \begin{bmatrix} A & \mathbf{1} \end{bmatrix}$, $c = (0, \ldots, 0, 1)$ and $\tilde{d} = (d, \epsilon)$

By Farkas lemma, we have the alternative $B^T u = c, \; u \geq 0$, 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

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

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

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

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

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

$\longrightarrow$

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 $\longrightarrow$ $\exists u \geq 0$ such that $A^T u = 0$ and $\mathbf{1}^T u = 1$

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

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

$\longrightarrow$ Infeasible system

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

Farkas lemma variation $\longrightarrow$ $\exists u \geq 0$ 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)

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

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

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).

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

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

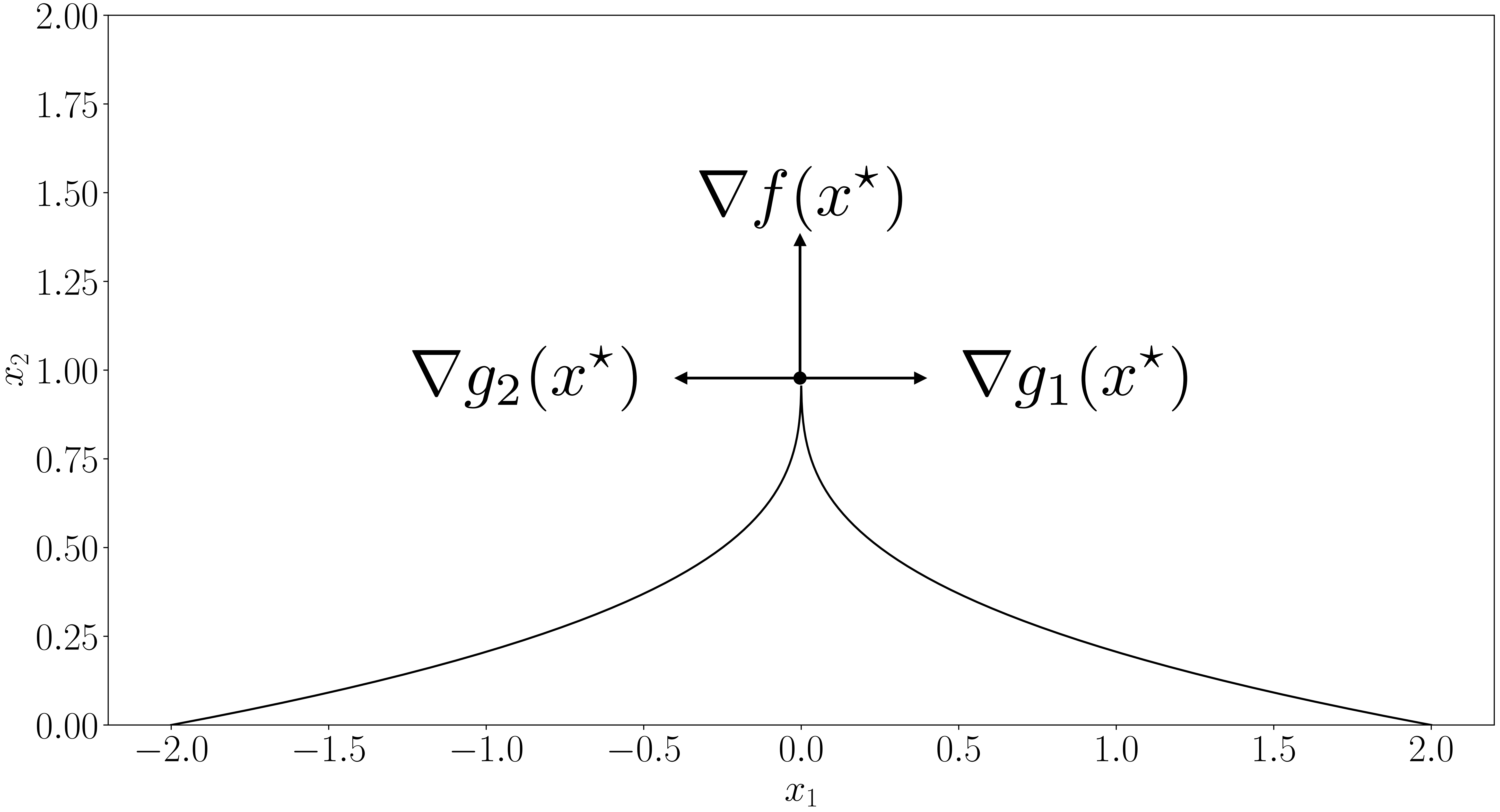
Which can be rewritten as $\nabla f(x^\star) + \sum_{i=1}^{m} y_i \nabla g_i(x^\star) = 0$

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

What happens if LICQ fails?

$$
\begin{array}{ll}
\text{minimize} & -x_2 \\
\text{subject to} & x_1 - 2(1 - x_2)^3 \leq 0 \\
& -x_1 - 2(1 - x_2)^3 \leq 0 \\
& x \geq 0
\end{array}
$$

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



Lagrangian function and duality

Lagrangian

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

Optimal cost

$f(x^\star) = p^\star$

Lagrangian

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

Optimal cost

$$f(x^\star) = p^\star$$

Lagrange multipliers

$$\begin{array}{lll} g_i(x) \leq 0 & \Longrightarrow & y_i \geq 0 \\ h_i(x) = 0 & \Longrightarrow & v_i \end{array}$$

Lagrangian

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

Optimal cost

$$f(x^\star) = p^\star$$

Lagrange multipliers

$$g_i(x) \leq 0 \quad \Longrightarrow \quad y_i \geq 0$$
$$h_i(x) = 0 \quad \Longrightarrow \quad v_i$$

Lagrangian

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

$(y \geq 0)$

Lagrangian

Interpretation

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

Lagrangian

remember $y \geq 0$

Interpretation

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

Proof

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

Lagrangian

Interpretation

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

Proof

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

Dual function

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

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

Lagrange dual problem

Finding the best lower bound

Always concave (-convex) problem

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

Dual problem

$$d^\star = \max_{y \geq 0, v} \underbrace{\min_x L(x, y, v)}_{g(y,v)}$$

Lower bound condition always holds

Weak duality

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

Stationarity condition

minimize $\quad f(x)$

subject to $\quad g_i(x) \leq 0, \quad i = 1, \ldots, m$

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

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

$\ell(x)$

Min-max formulation

$$p^\star = \min_x \left[\max_{y \geq 0, v} \; L(x, y, v) \right]$$ (minimize unconstrained version)

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

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

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

Theorem

If $x^\star$ is a local minimizer and 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**

$$\begin{aligned} g_i(x^\star) &\leq 0, \quad i = 1, \dots, m \\ h_i(x^\star) &= 0, \quad i = 1, \dots, p \end{aligned}$$ **primal feasibility**

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

Strong duality theorem

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

Theorem

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

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

Slater's condition

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

Strong duality theorem

For LPs, a sufficient condition for strong duality is

- primal feas.
- dual feas.

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

Theorem

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

$$\begin{array}{ll} g_i(x) < 0, \quad i = 1, \ldots, m, & \text{(for non-affine } g_i) \\ h_i(x) = 0, \quad i = 1, \ldots, p & \end{array}$$

Slater's condition

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

Converse is false

Counterexample $\min_x 0$

s.t. $x^2 \leq 0$

$x^* = 0$, strong duality holds

- Slater's does not hold

Remarks

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

KKT for convex problems

Always sufficient

For $x^\star, y^\star, v^\star$ that satisfy the KKT conditions

$$f(x^\star) = f(x^\star) + \underbrace{\sum_{i=1}^{m} y_i^\star g_i(x^\star)}_{\text{comp. slackness}} + \sum_{i=1}^{p} v_i^\star \underbrace{h_i(x^\star)}_{=0} = L(x^\star, y^\star, v^\star)$$

$g(y,v) = \min_x L(x,y,v) \leftarrow$ unconstrained

KKT for convex problems

Always sufficient

For $x^\star, y^\star, v^\star$ that satisfy the KKT conditions

f cvx
g_i cvx
h_i affine

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

$f(x^*)$
$\parallel$

$$\nabla f(x^\star) + \sum_{i=1}^{m} y_i^\star \nabla g_i(x^\star) + \sum_{i=1}^{p} v_i^\star \nabla h_i(x^\star) = 0 \quad \Rightarrow \quad g(y^\star, v^\star) = L(x^\star, y^\star, v^\star) \quad \text{[Convexity]}$$

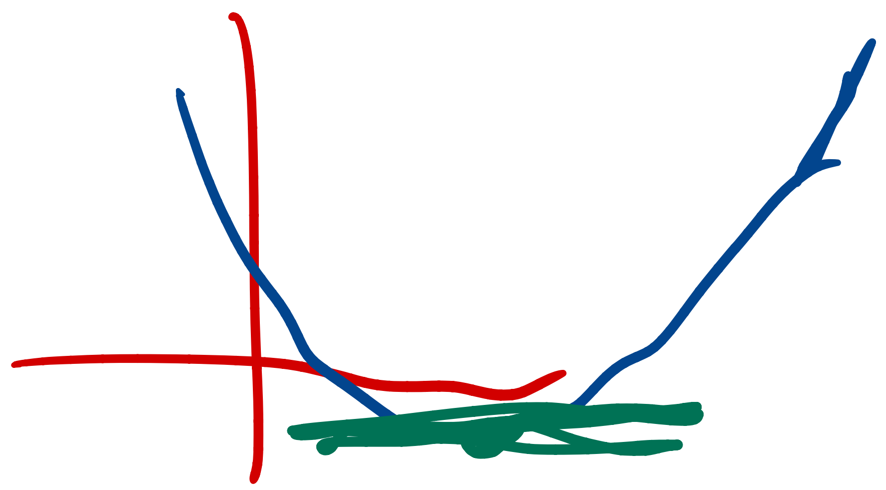
f convex, differentiable $\implies$ [$\nabla f(x) = 0 \iff x$ is a global min]

$\xRightarrow{}$ cvx

for any cvx fn f: $f(y) - f(x) \geq \cancel{\nabla f(x)^T (y-x)} \quad \forall x, y$

if $\nabla f(x) = 0$

then $f(y) \geq f(x) \quad \forall y$



Given cvx prob. Diff, constraints KKT holds $\overset{?,\Leftarrow}{\Rightarrow}$ [strong duality holds and x^*, y^*, v^* optimal]

KKT for convex problems

$f(x^*) = g(y^*, v^*)$

Always sufficient

For $x^\star, y^\star, v^\star$ that satisfy the KKT conditions

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

$$\nabla f(x^\star) + \sum_{i=1}^{m} y_i^\star \nabla g_i(x^\star) + \sum_{i=1}^{p} v_i^\star \nabla h_i(x^\star) = 0 \quad \Rightarrow \quad g(y^\star, v^\star) = L(x^\star, y^\star, v^\star)$$ [Convexity]

f convex, differentiable $\implies$ [$\nabla f(x) = 0 \iff x$ is a global min]

Therefore, $f(x^\star) = g(y^\star, v^\star)$ and $x^\star, y^\star, v^\star$ are primal-dual optimal

Necessary when constraint qualifications (Slater's) condition holds

If $x^\star$ strictly primal feasible (Slater's), then strong duality $f(x^\star) = g(y^\star, v^\star)$

Therefore, dual optimum attained and KKT conditions satisfied

KKT remarks

History

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

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 $\nabla f(x) = 0$

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

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 strong duality holds, KKT conditions are **necessary and sufficient**

Example: KKT conditions for convex QP

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

$P \succeq 0$

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

Generalizes LPs

$\min_x q^T x$
s.t. $Cx \leq d$

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

KKT Optimality conditions

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

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

$\cancel{Ax - b = 0}$
$Cx - d \leq 0$ **primal feasibility**

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

Convex constrained nonconvex optimization

Minimization over convex set

[Section 3.7.3 and Example 3.74, A. Beck]

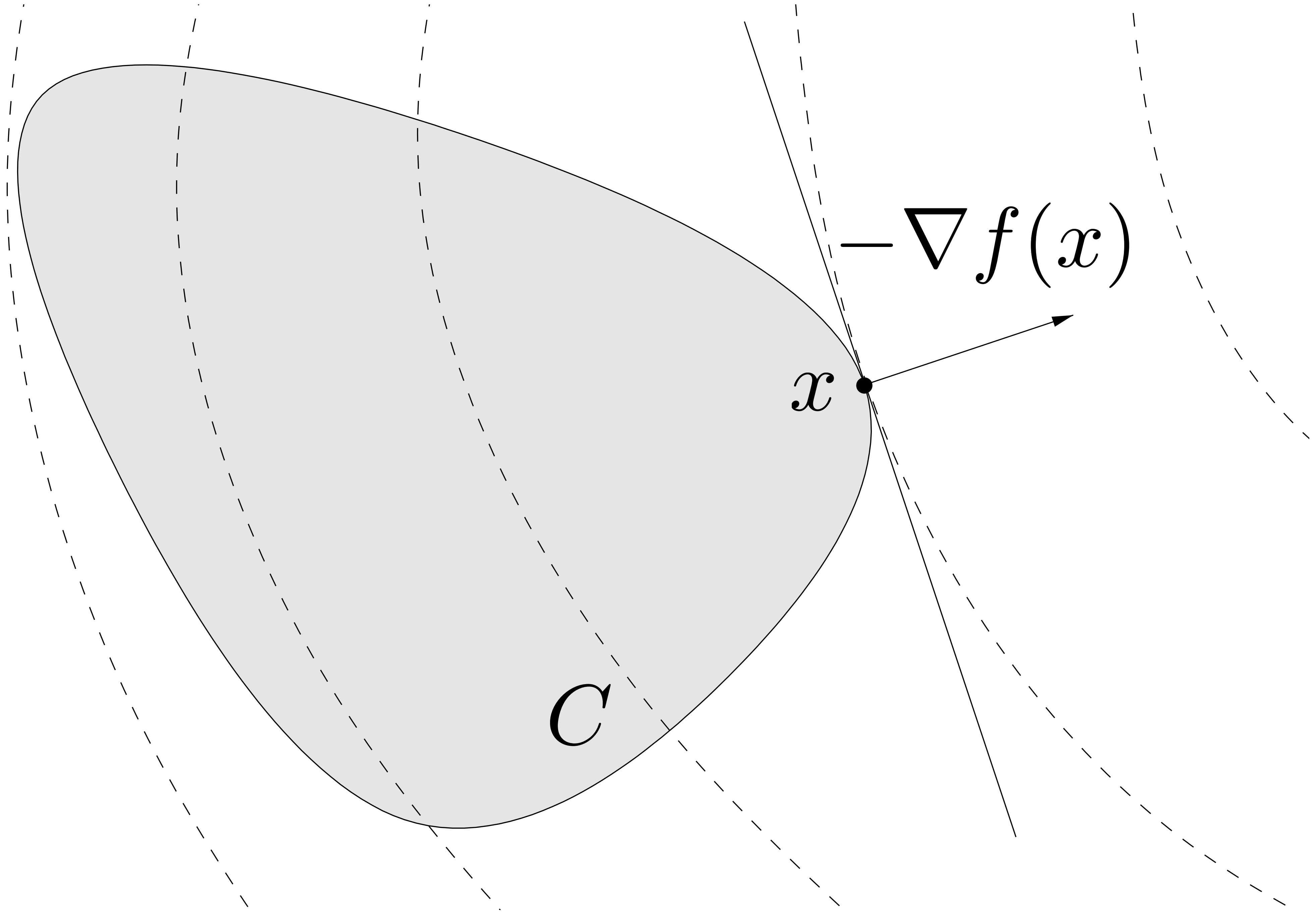
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array}$$

convex set

Minimization over convex set

[Section 3.7.3 and Example 3.74, A. Beck]

$$\begin{array}{ll} \text{minimize} & f(x) \leftarrow \text{non-convex} \\ \text{subject to} & x \in C \longleftarrow \text{convex set} \end{array}$$



First-order optimality condition

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

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

(f can be nonconvex)

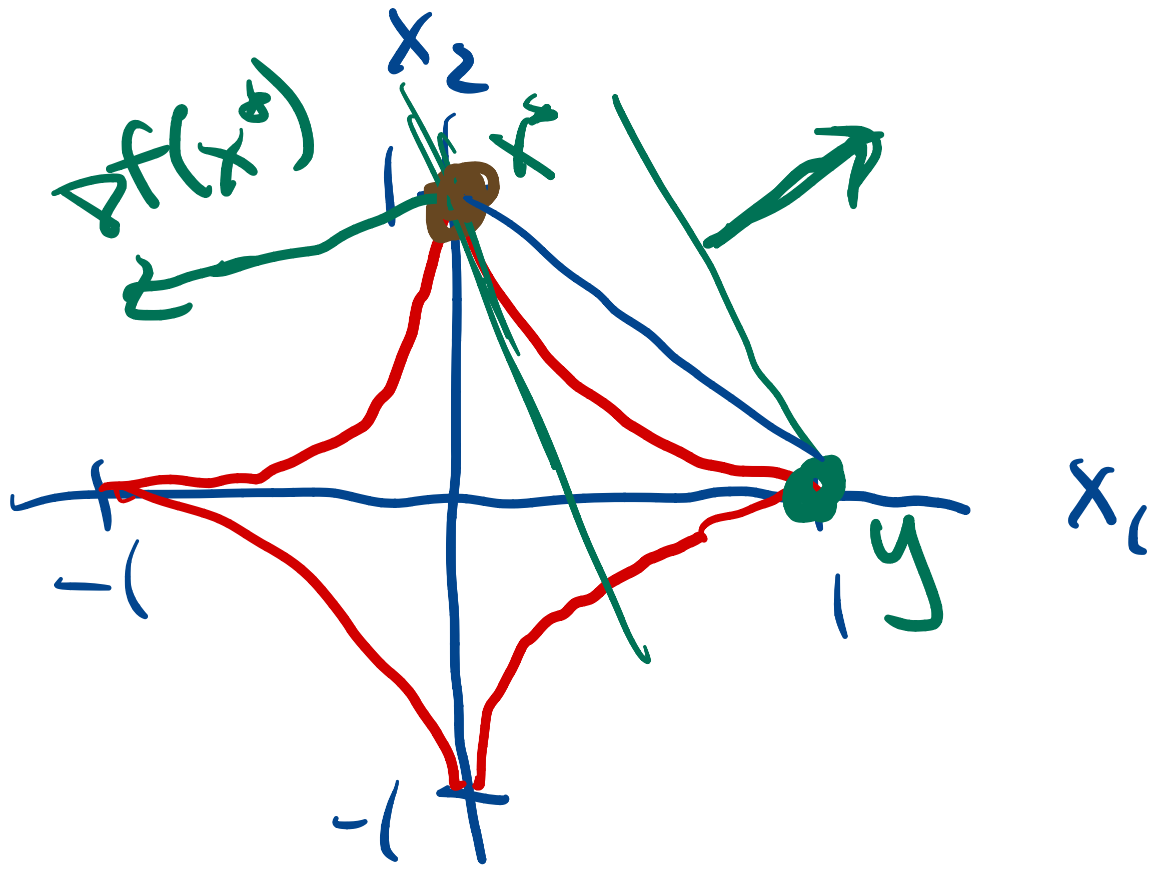
Why do you need a convex set?

First-order necessary optimality condition

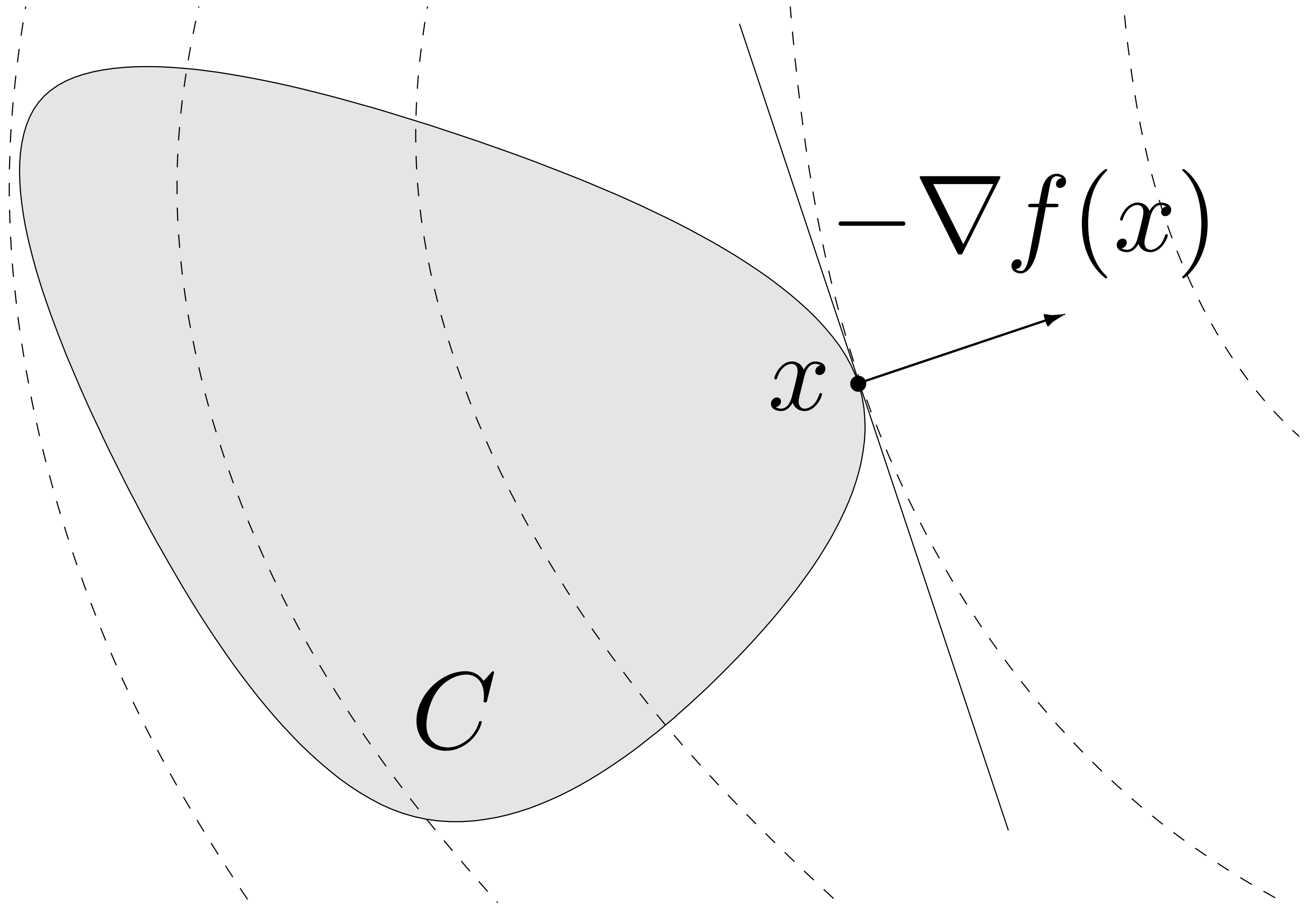
If $x^\star$ is a local minimum, then

$$\underline{\nabla f(x^\star)^T(y - x^\star)} \geq 0, \quad \forall y \in C$$

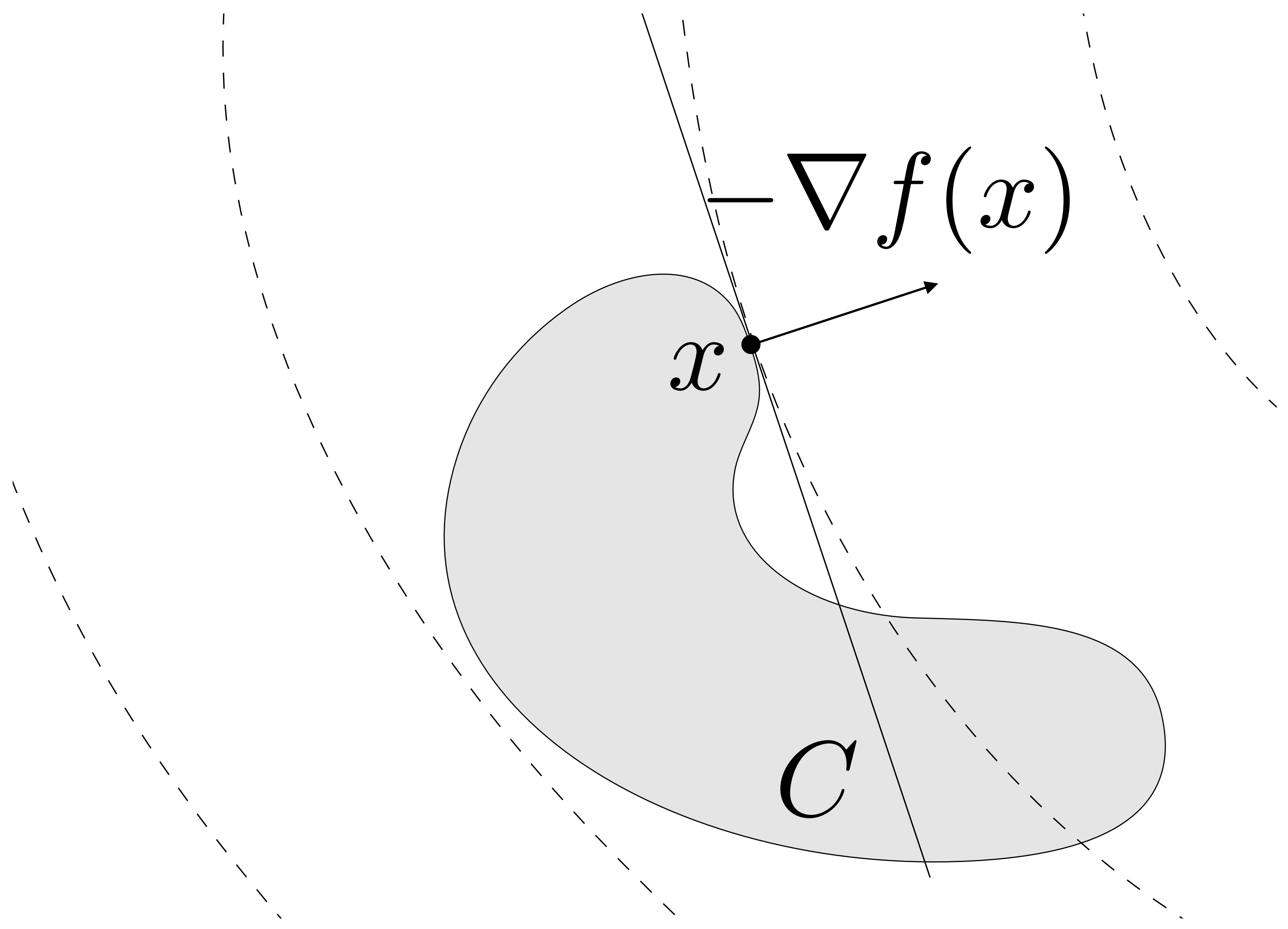
$$\min_x \; -2x_1 - x_2$$
$$\text{s.t. } \|x\|_{\frac{1}{2}} \leq 1$$



Convex set

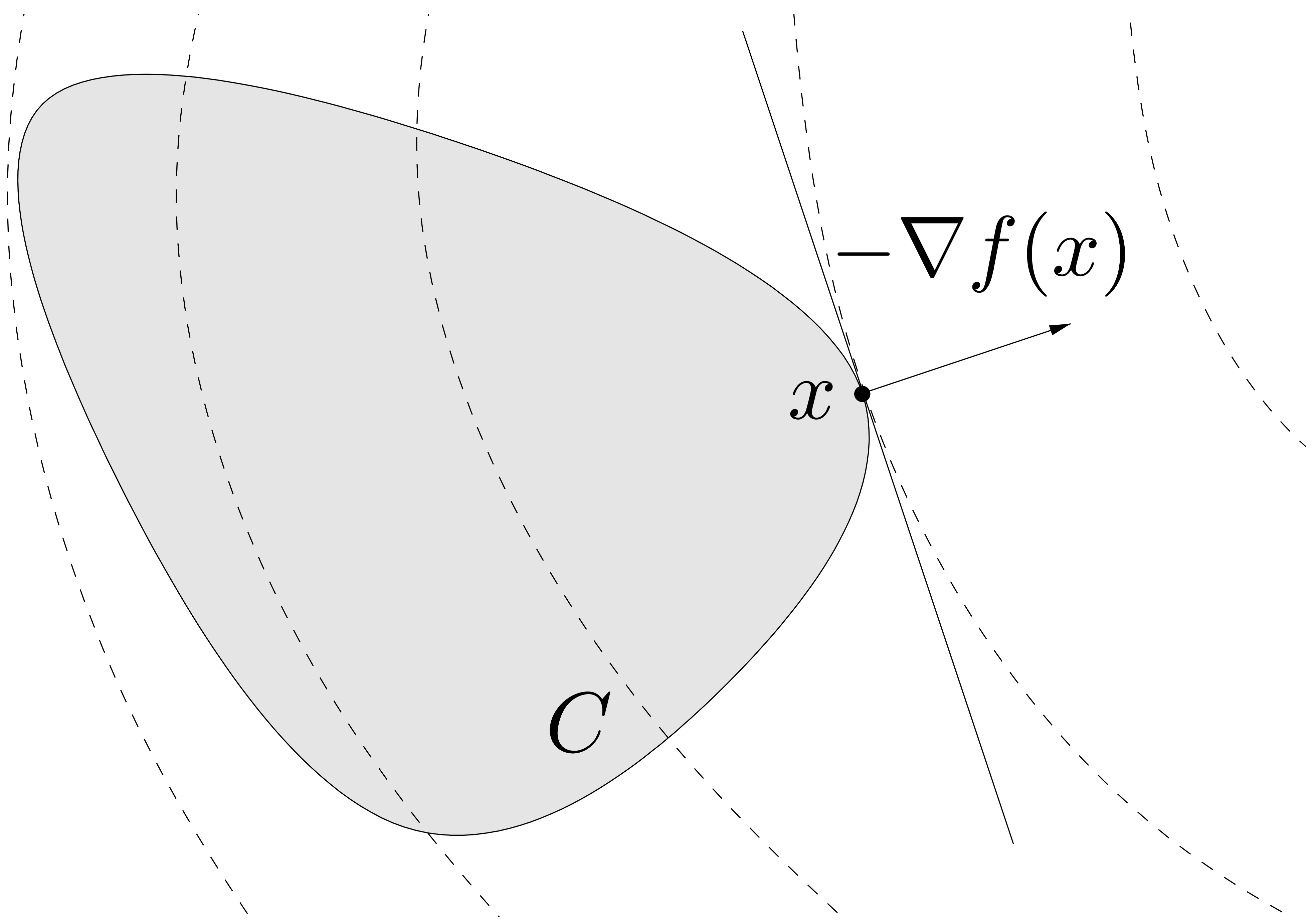


Nonconvex set



• local min

Normal cone condition

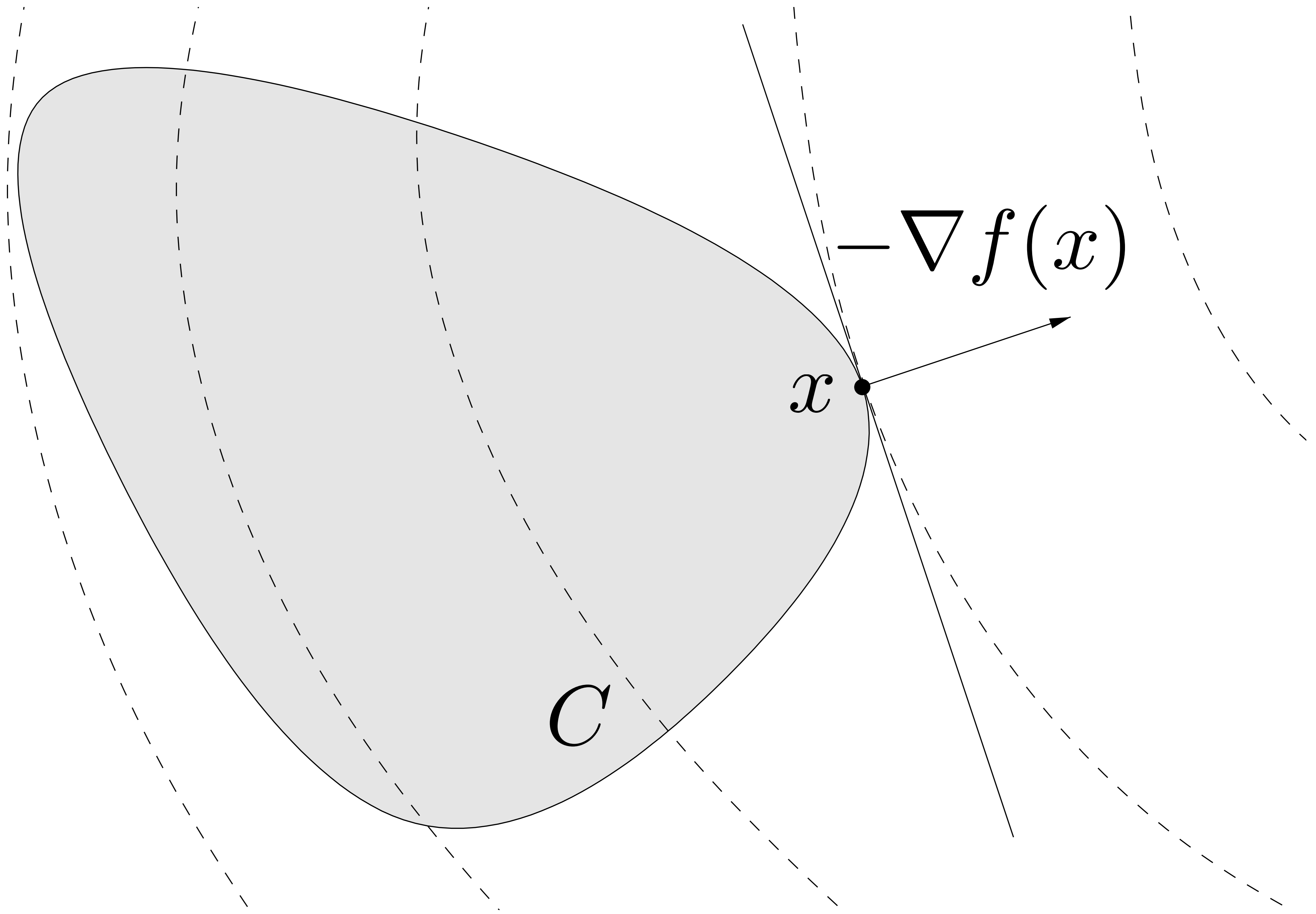


First-order necessary optimality condition

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

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

Normal cone condition



First-order necessary optimality condition

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

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

$$-\nabla f(x^\star)^T(y - x^\star) \leq 0 \quad \forall y \in C$$

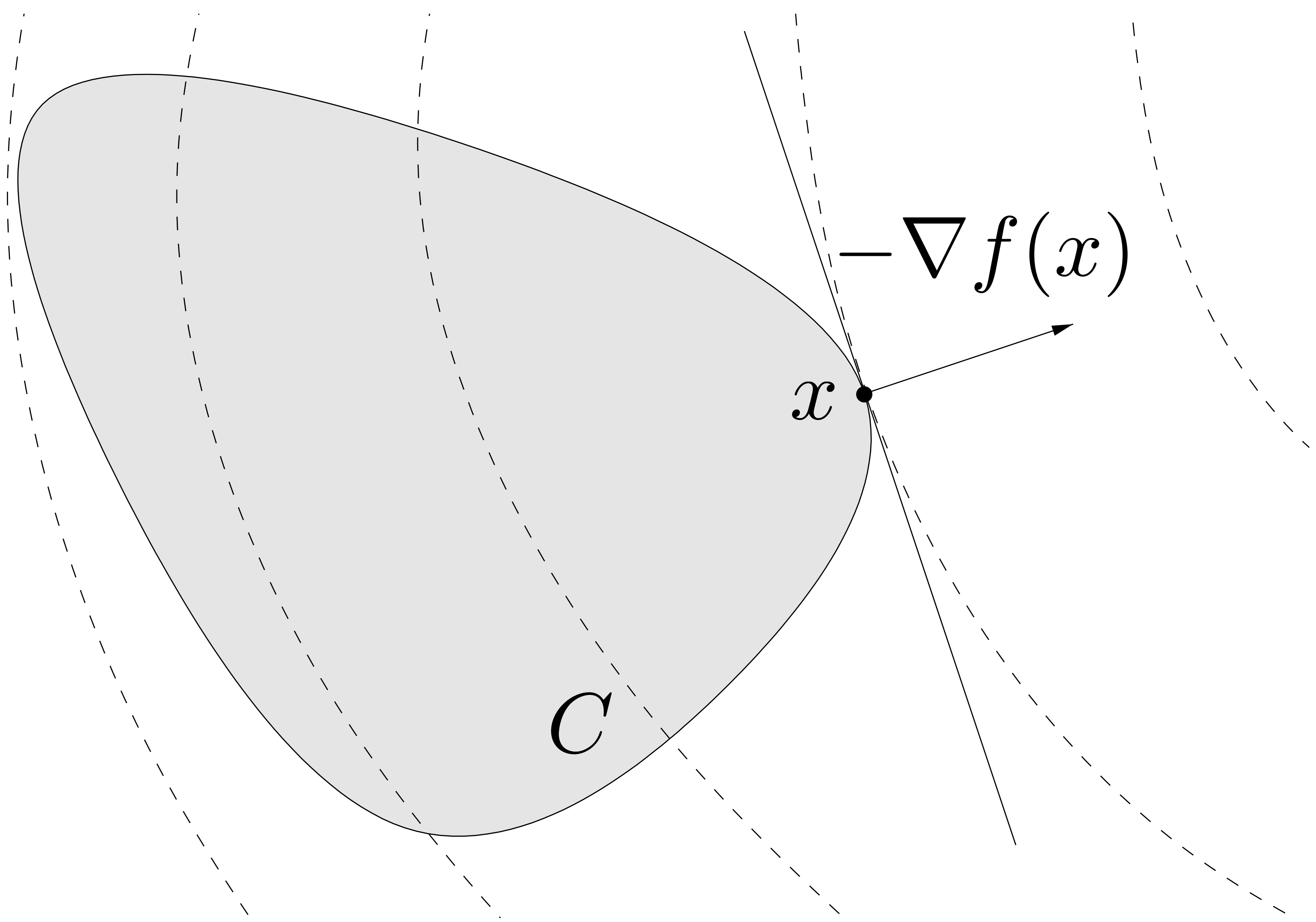
Normal cone

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

Reformulated condition

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

Normal cone condition



First-order necessary optimality condition

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

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

Normal cone

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

Reformulated condition

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

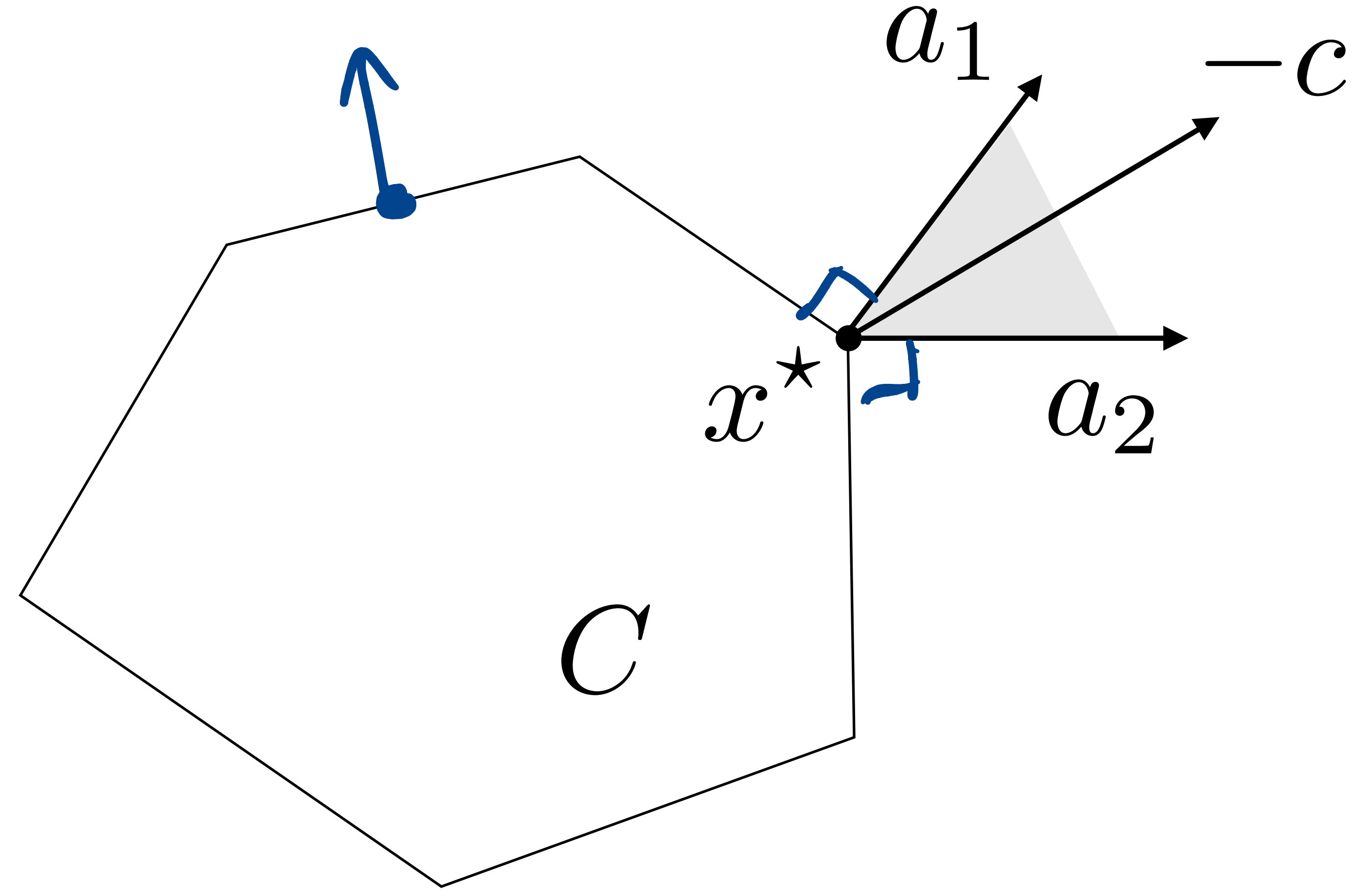
Remark
If f and C are convex, then it is
necessary and sufficient
[Section 4.2.3, B and V]

Normal cone condition

Linear program example

minimize $\quad c^T x$

subject to $\quad Ax \leq b$



Recap from Lecture 8

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

Optimal dual solution y satisfies:

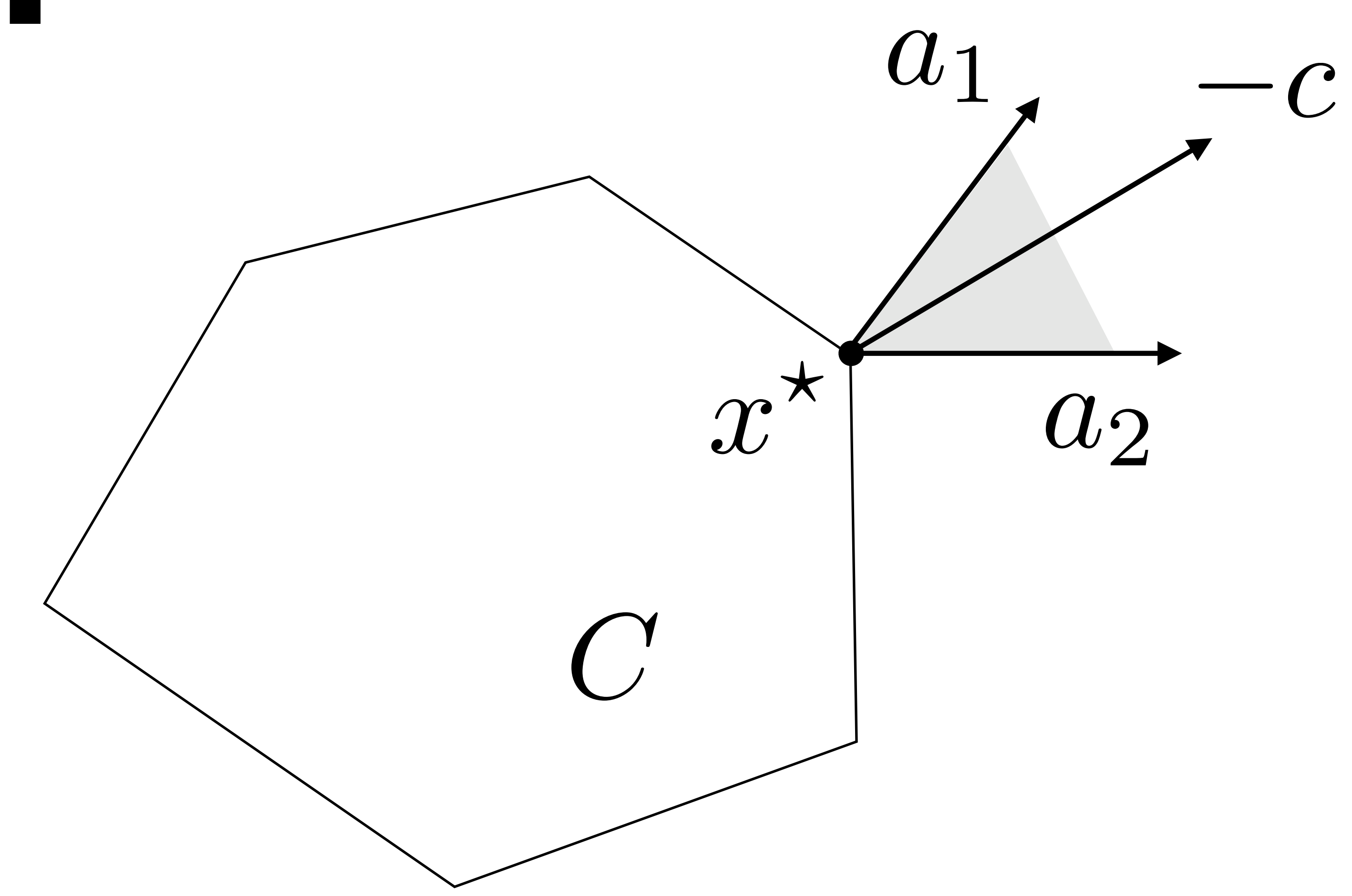
$$A^T y + c = 0, \quad y \geq 0, \quad y_i = 0 \text{ for } i \neq \{1, 2\}$$

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

Normal cone condition

Linear program example

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$



Recap from Lecture 8

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

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

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

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