ORF522 – Linear and Nonlinear Optimization

12. Introduction to nonlinear optimization

Bartolomeo Stellato — Fall 2020

Anonymous questionnaire

Best aspects

- Lecture slides
- Ed forum questions + answers at the beginning of the class

Things to improve

- **Instructor should go slower/write in some parts**
  - Please ask me to interrupt at any point over the lecture
  - You can also ask on Ed forum to repeat/go over some unclear parts
  - Happy to clarify concepts in the office hours
- **More worked examples/precepts that are different than the homework**
  - We could have some practice problems during AI office hours
  - Happy to review material from the lecture during my office hours

Today’s lecture

[Chapter 2-4 and 6, B and V] [Chapter A, B, H-H and L]

- Nonlinear optimization
- Examples
- Convex analysis review
- Convex optimization

References for this part of the course (all available online)

- J. Nocedal and S. Wright, “Numerical Optimization”
- S. Boyd and L. Vandenberghe, “Convex Optimization”
- L. Vandenberghe, “UCLA 236C lecture notes”
- Y. Nesterov, “Introductory lectures to convex optimization”
- J. B. Hiriart-Hrruty and C. Lemarechal, “Fundamentals of Convex analysis”
- A. Beck, “First-order methods in optimization”

What if the problem is no longer linear?

Nonlinear optimization

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

$x = (x_1, \dots, x_n)$ Variables

$f : \mathbf{R}^n \to \mathbf{R}$ Nonlinear objective function

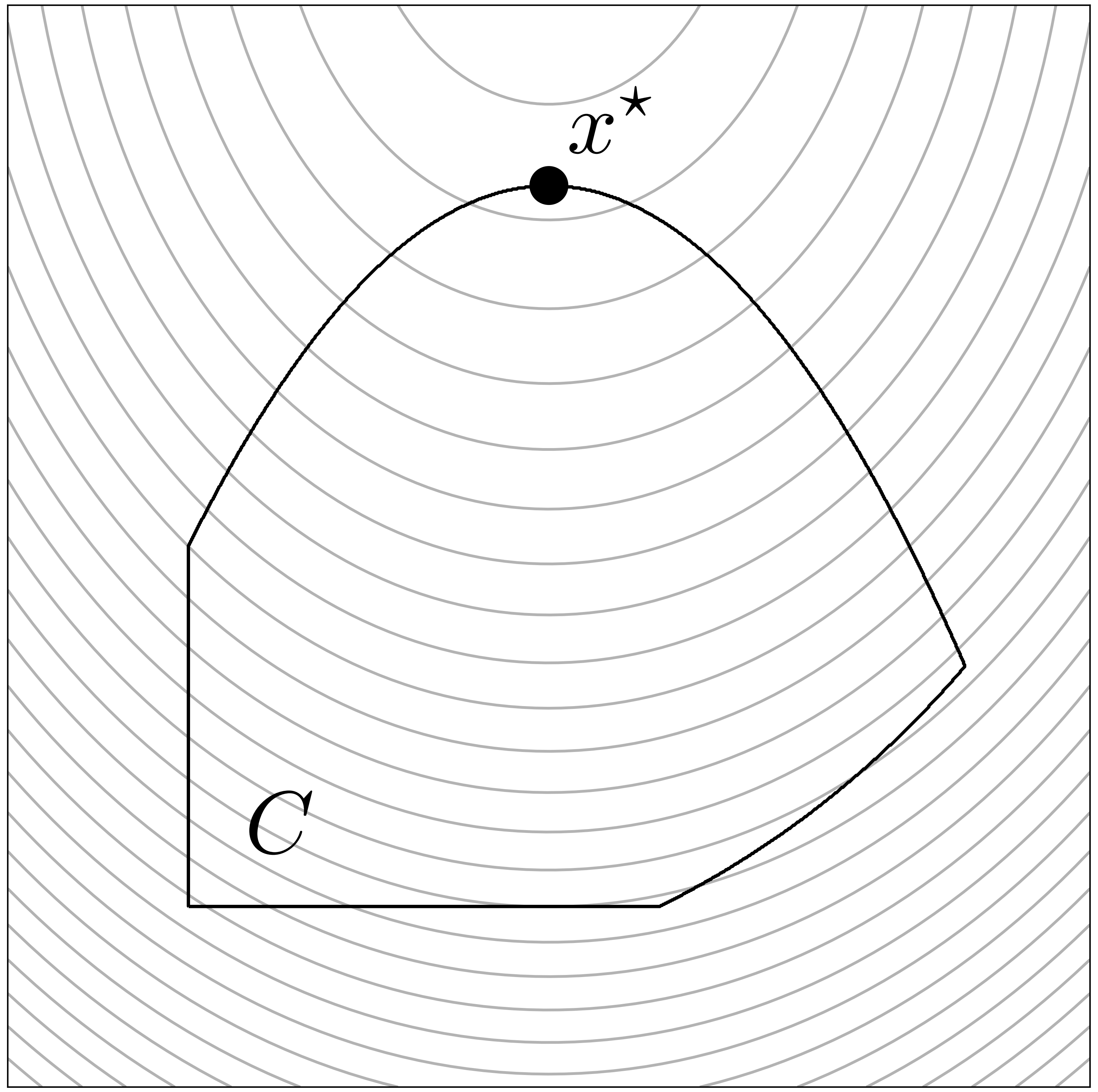
$g_i : \mathbf{R}^n \to \mathbf{R}$ Nonlinear constraints functions

Feasible set

$$C = \{x \mid g_i(x) \leq 0, \quad i = 1, \dots, m\}$$

Small example

$$\begin{array}{ll} \text{minimize} & 0.5x_1^2 + 0.25x_2^2 \\ \text{subject to} & e^{x_1} - 2 - x_2 \leq 0 \\ & (x_1 - 1)^2 + x_2 - 3 \leq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 1 \end{array}$$



Contour plot has curves (no longer lines)

Feasible set is no longer a polyhedron

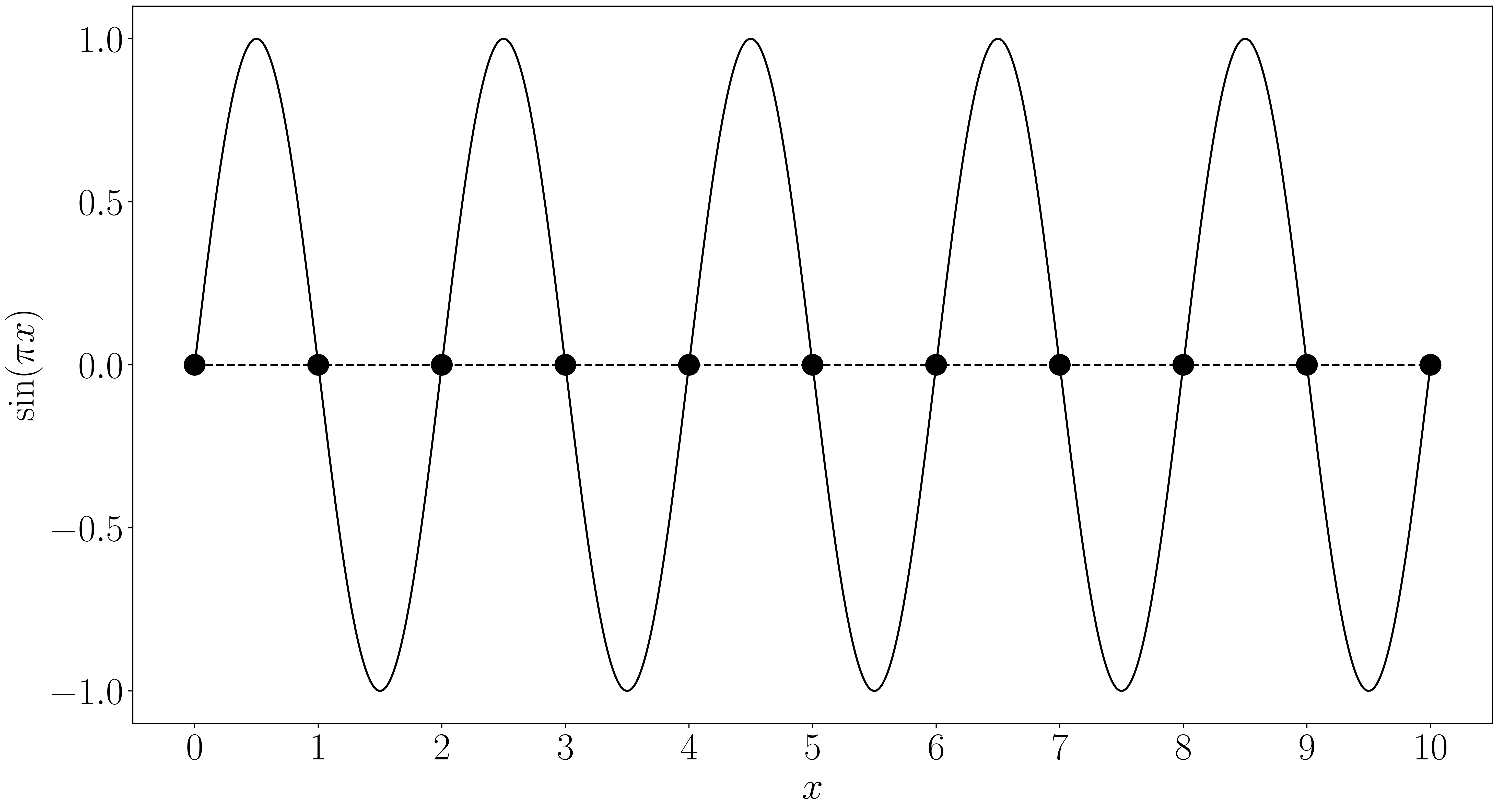
Integer optimization

It's still nonlinear optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \sin(\pi x) = 0 \end{array}$$

$\Updownarrow$

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



We cannot solve most nonlinear optimization problems

Examples of (solvable) nonlinear optimization

Regression

Fit affine function $f(z) = \alpha + \beta z$ to m points (z_i, y_i)

Approximation problem $Ax \approx b$ where $$A = \begin{bmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_m \end{bmatrix}, \quad x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Goal

minimize $\quad \|Ax - b\|$

1-norm or ∞-norm $\quad \Longrightarrow \quad$ **linear optimization**

2-norm $\quad \Longrightarrow \quad$ **least-squares** $\quad \|Ax - b\|_2^2 = \sum_i (f(z_i) - y_i)^2$

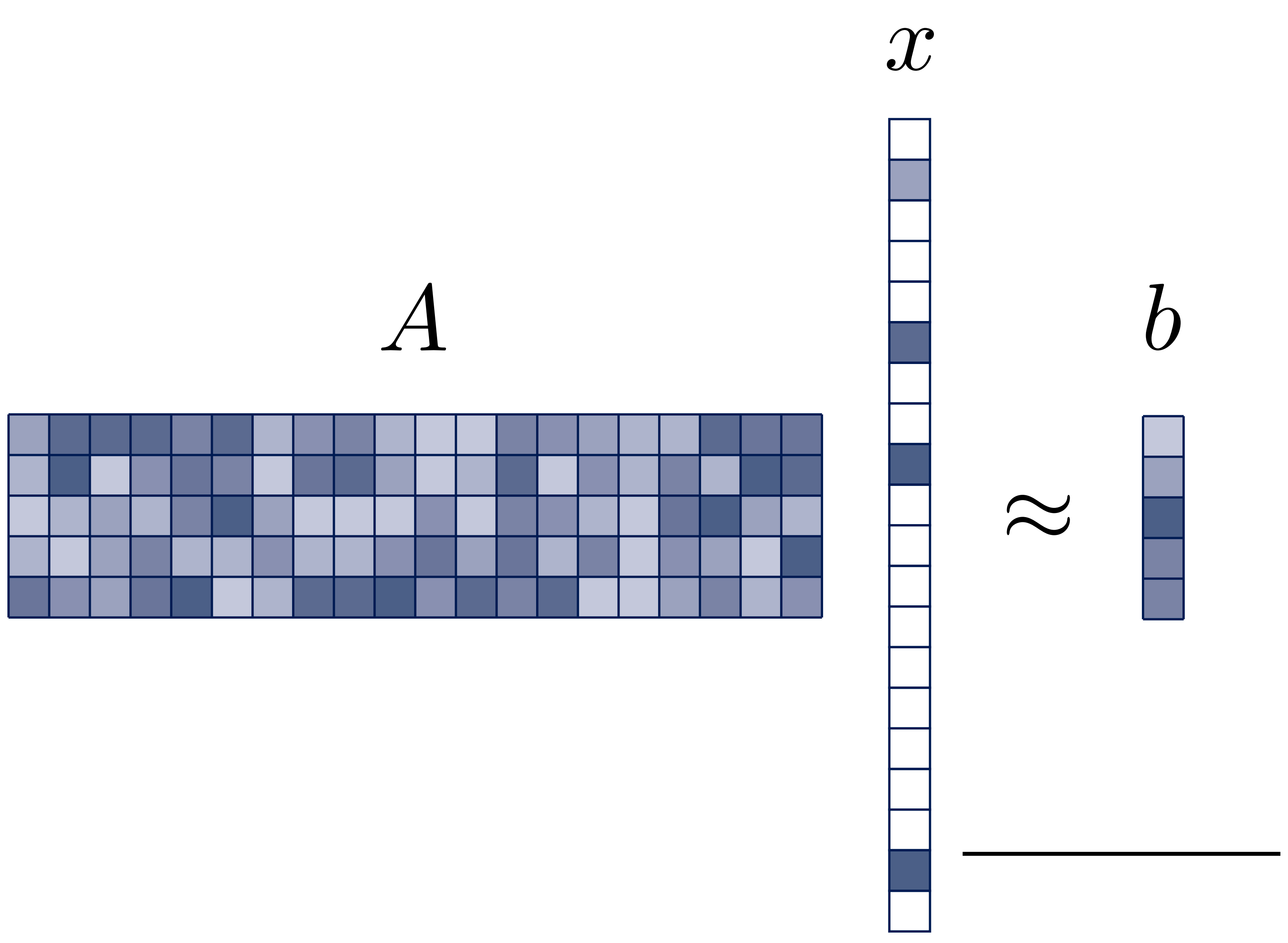
Sparse regression

Regressor selection

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

subject to $\mathbf{card}(x) \leq k$

(very hard)



sparse x is
more robust

Add regularization to the objective

Regularized regression (ridge)

minimize $\|Ax - b\|_2^2 + \gamma\|x\|_2^2$

Regularized regression (lasso)

minimize $\|Ax - b\|_2^2 + \gamma\|x\|_1$

Lasso vs ridge regression

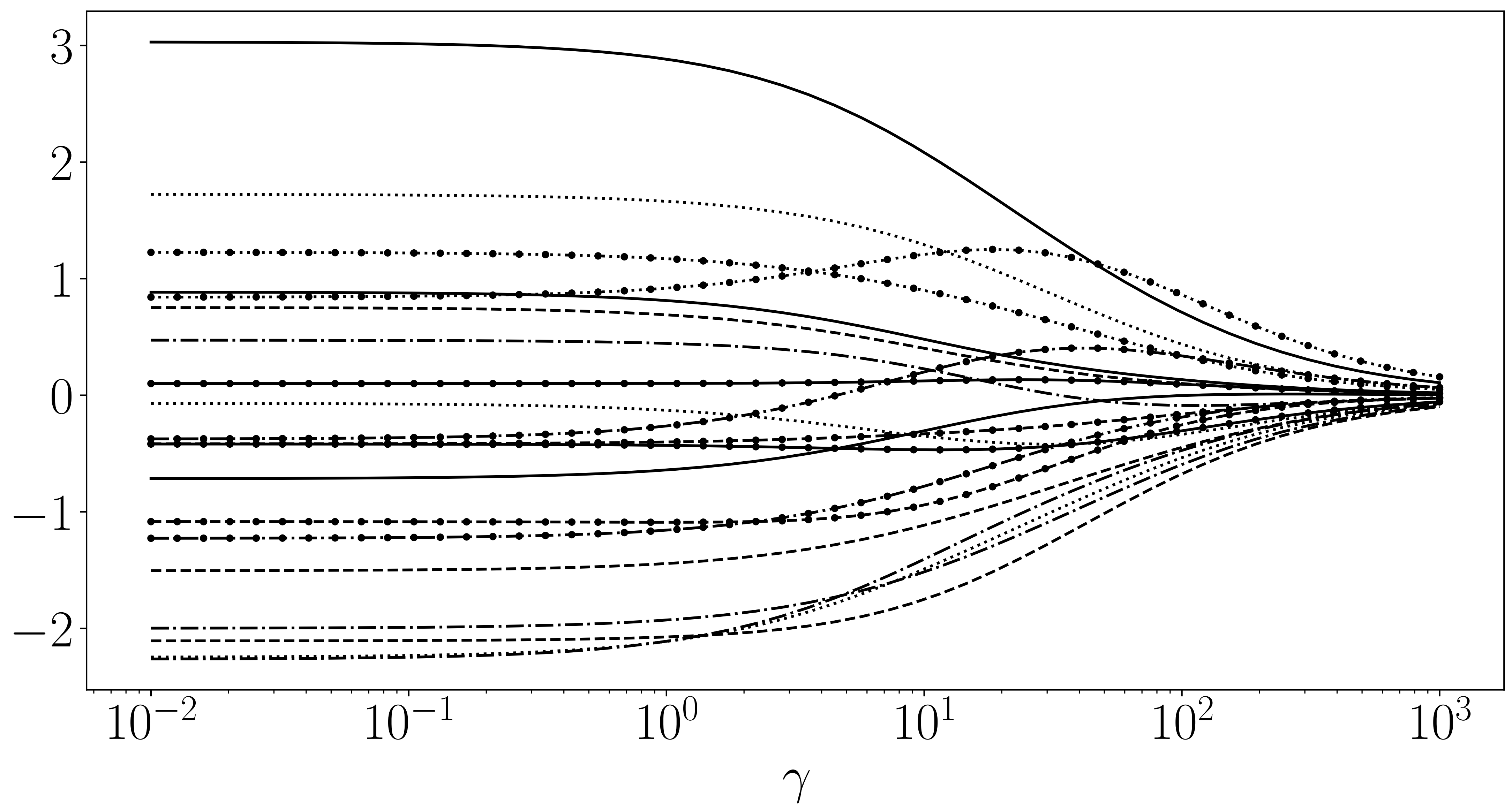
Regularized regression (ridge)

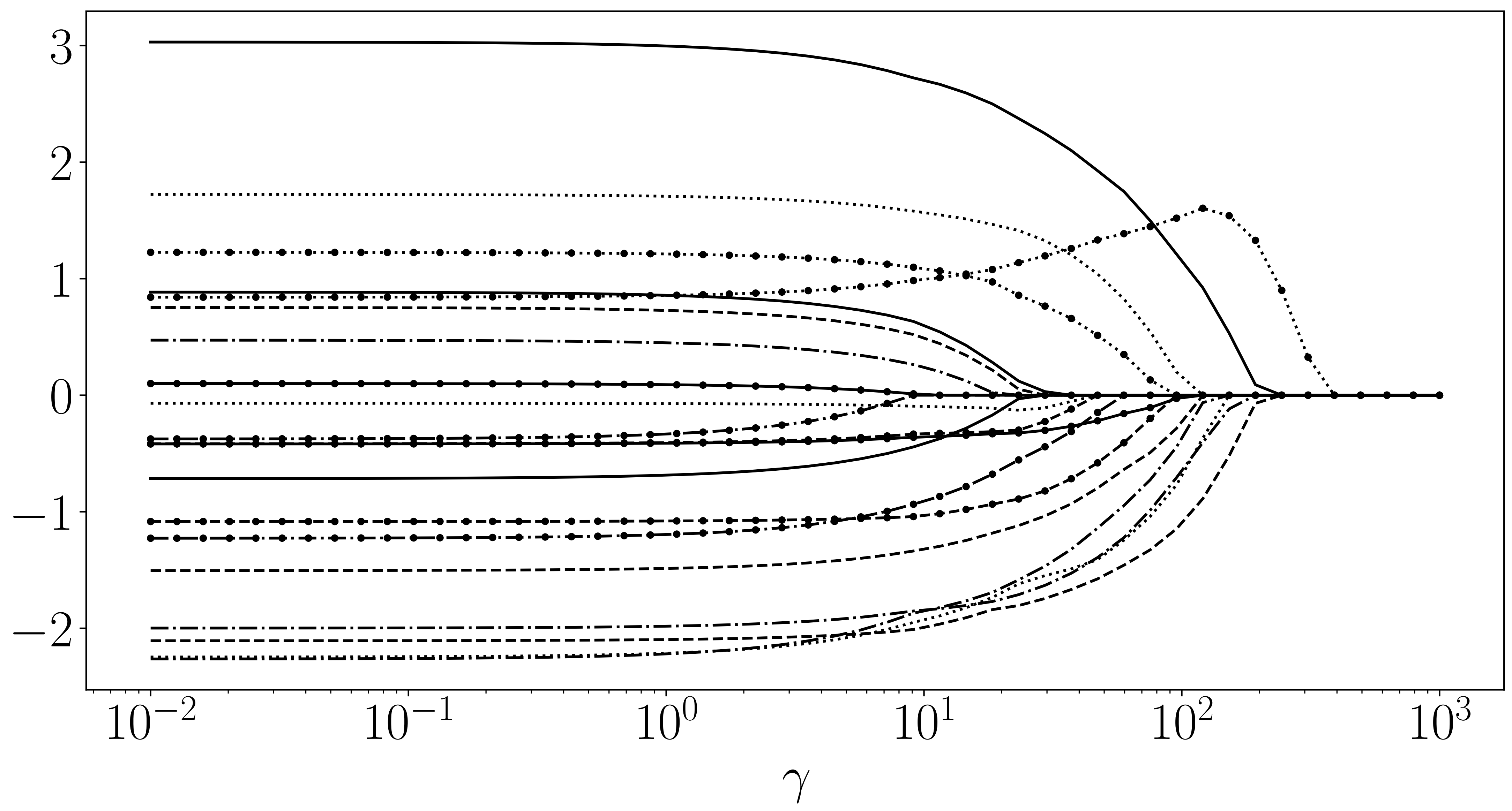
minimize $\quad \|Ax - b\|_2^2 + \gamma \|x\|_2^2$

Regularized regression (lasso)

minimize $\quad \|Ax - b\|_2^2 + \gamma \|x\|_1$

Regularization paths





Portfolio optimization

We have a total of n assets

x_i is fraction of money invested in asset i
p_i is the relative price change of asset i → **Returns** $p^T x$

p random variable: mean μ, covariance Σ

Portfolio optimization

$$\begin{array}{ll} \text{maximize} & \mu^T x - \gamma x^T \Sigma x \\ \text{subject to} & \mathbf{1}^T x = 1 \\ & x \geq 0 \end{array}$$

Expected return → $\mu^T x$

Risk → $x^T \Sigma x$

Risk-aversion parameter → γ

Convex analysis review

Extended real-value functions

$f(x)$ on $\mathbf{dom} f$

Extended-value extension

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom} f \\ \infty & x \notin \mathbf{dom} f \end{cases}$$

No need to explicitly define domain

$$\mathbf{dom} \tilde{f} = \{x \mid \tilde{f}(x) < \infty\}$$

Indicator functions

Indicator function

$$\mathcal{I}_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

Constrained form

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

$\longleftrightarrow$

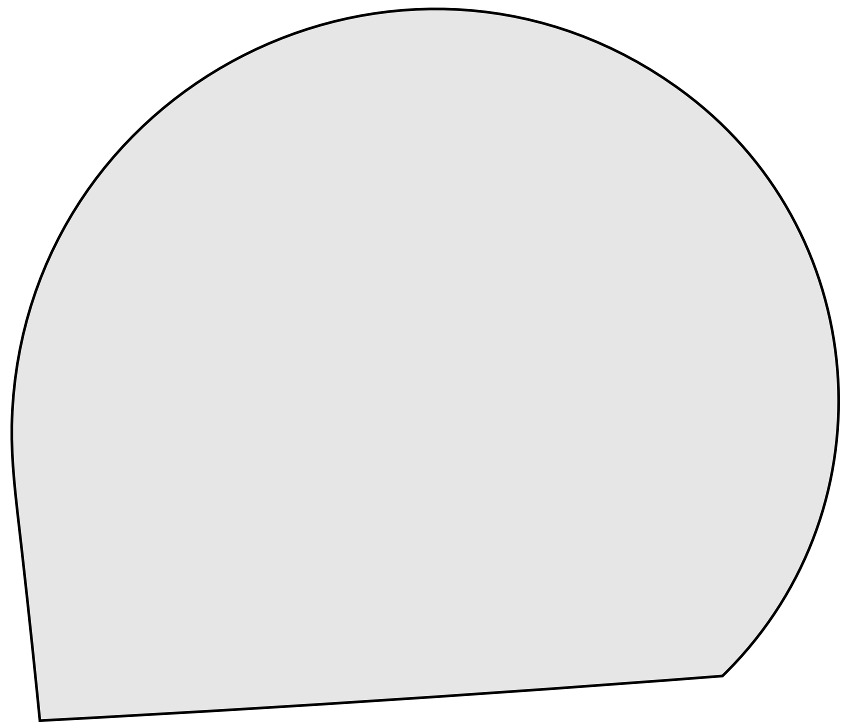
Unconstrained form

$$\text{minimize} \quad f(x) + \mathcal{I}_C(x)$$

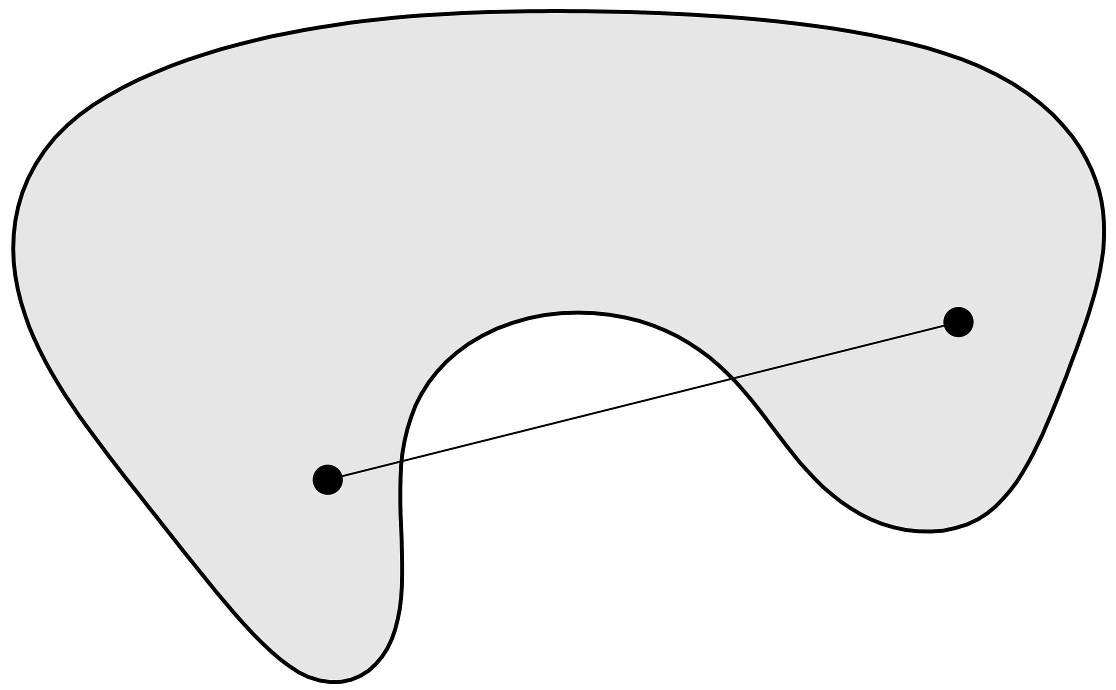
Convex set

Definition

For any $x, y \in C$ and any $\alpha \in [0, 1]$ $\qquad \alpha x + (1-\alpha)y \in C$



Convex



Not convex

Examples

- $\mathbf{R}^n$
- Hyperplanes
- Hyperspheres
- Polyhedra

Examples

- Cardinality
- $\mathbf{Z}^n$
- Any disjoint set

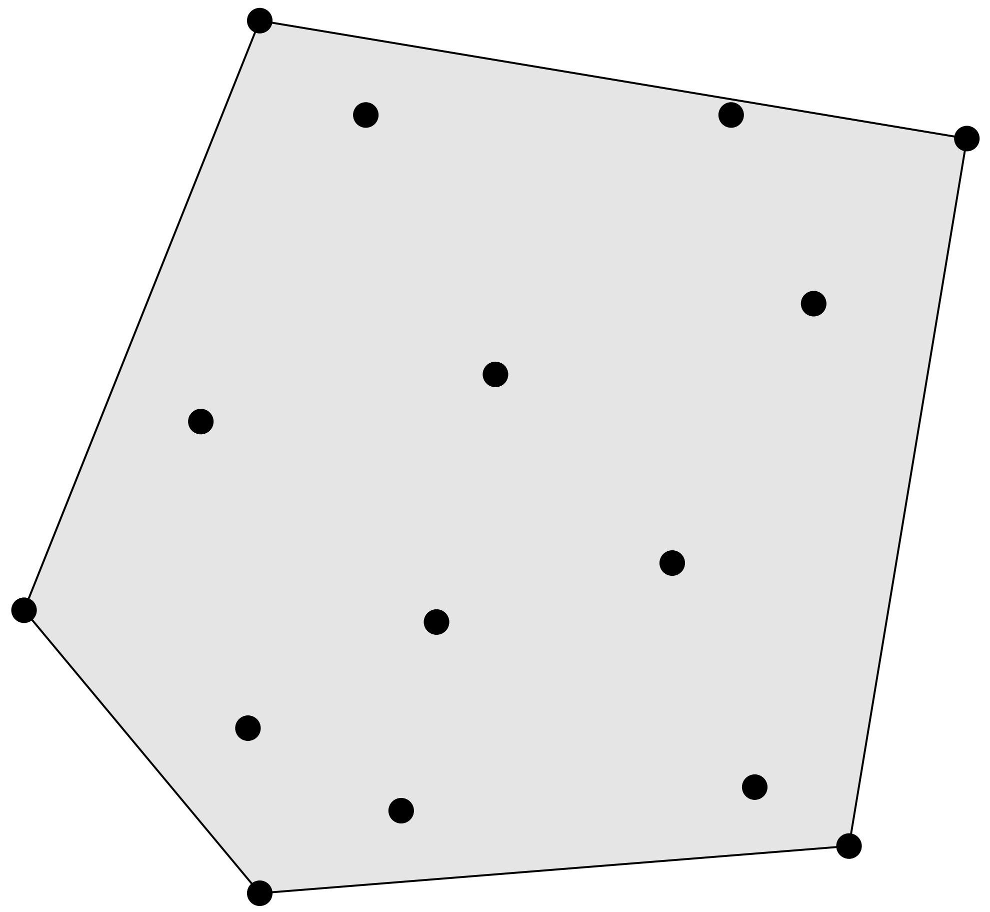
Convex combinations

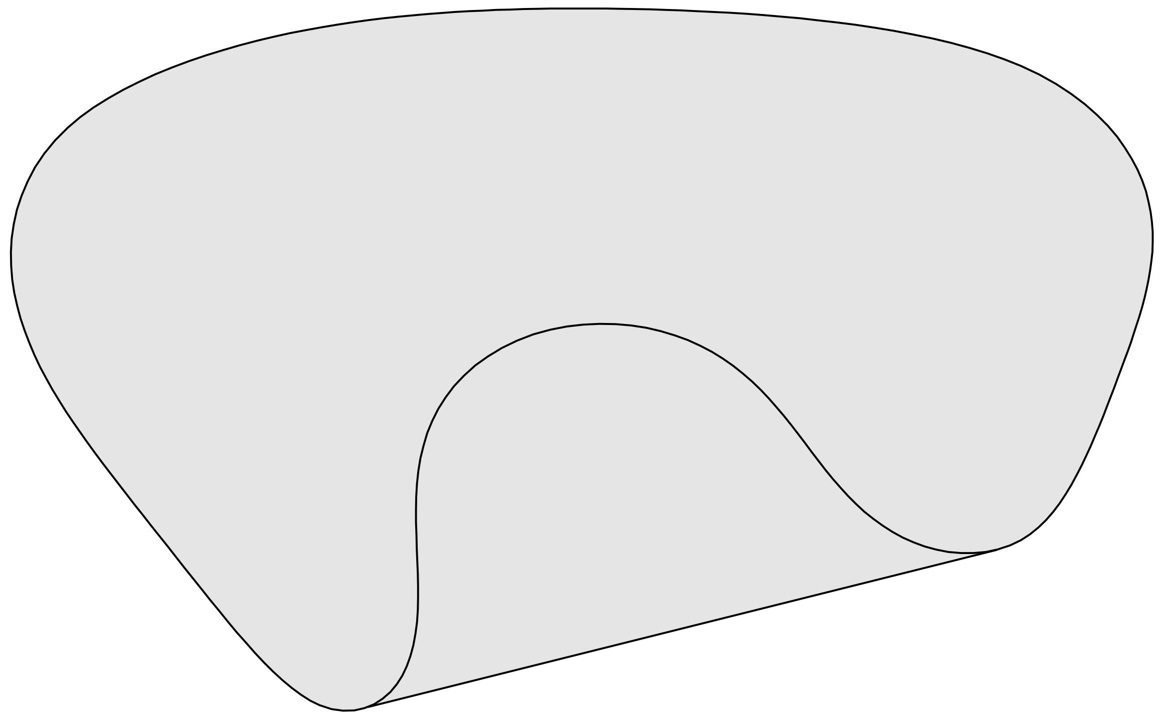
Convex combination

$\alpha_1 x_1 + \cdots + \alpha_k x_k$ for any $x_1, \ldots, x_k$ and $\alpha_1, \ldots, \alpha_k$ such that $\alpha_i \geq 0,\ \sum_{i=1}^{k} \alpha_i = 1$

Convex hull

$$\mathbf{conv}\, C = \left\{ \sum_{i=1}^{k} \alpha_i x_i \mid x_i \in C, \quad \alpha_i \geq 0, \quad i = 1, \ldots, k, \quad \mathbf{1}^T \alpha = 1 \right\}$$





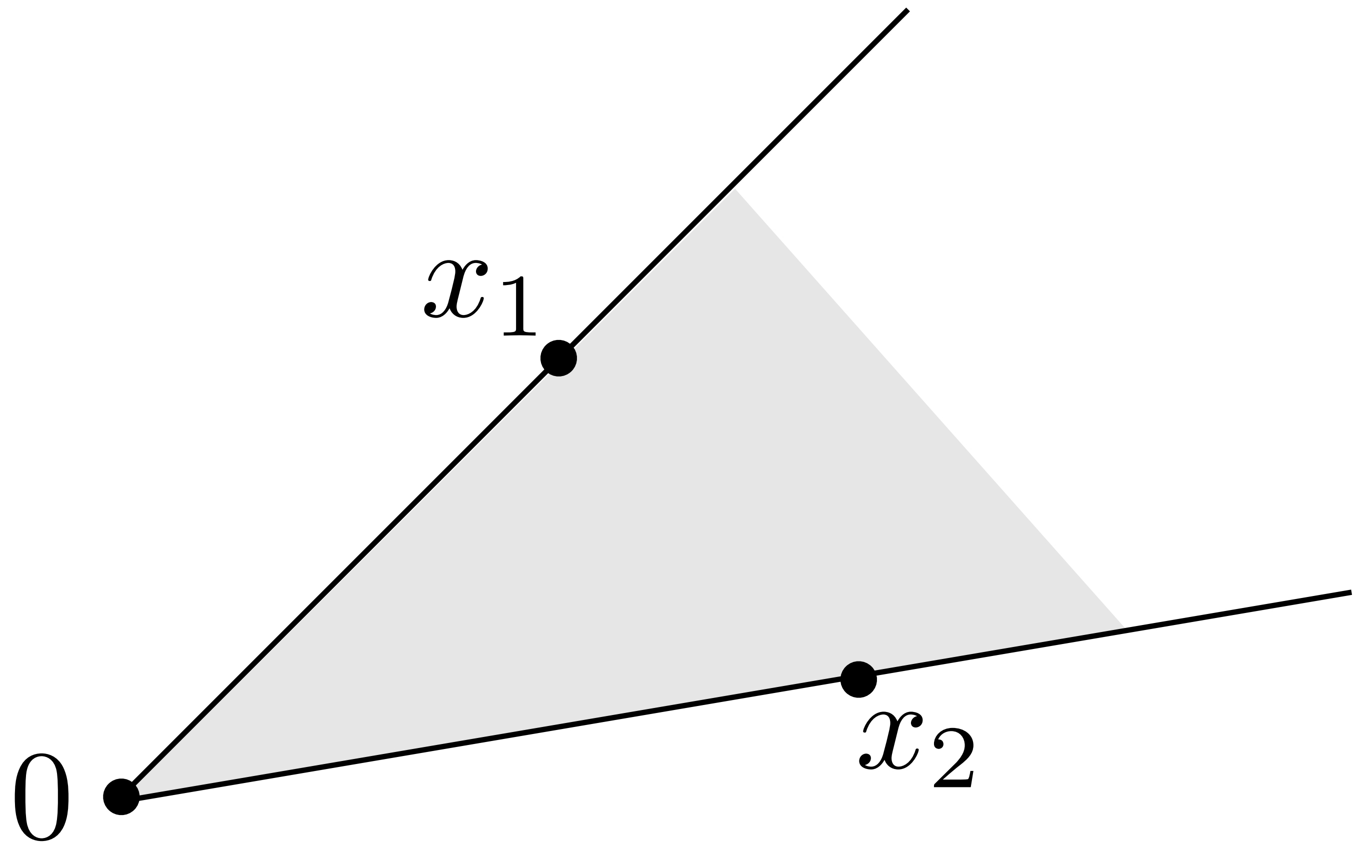
Cones

Cone

$x \in C \quad \implies \quad tx \in C$ for all $t \geq 0$

Convex cone

$x_1, x_2 \in C \quad \implies \quad t_1 x_1 + t_2 x_2 \in C$ for all $t_1, t_2 \geq 0$



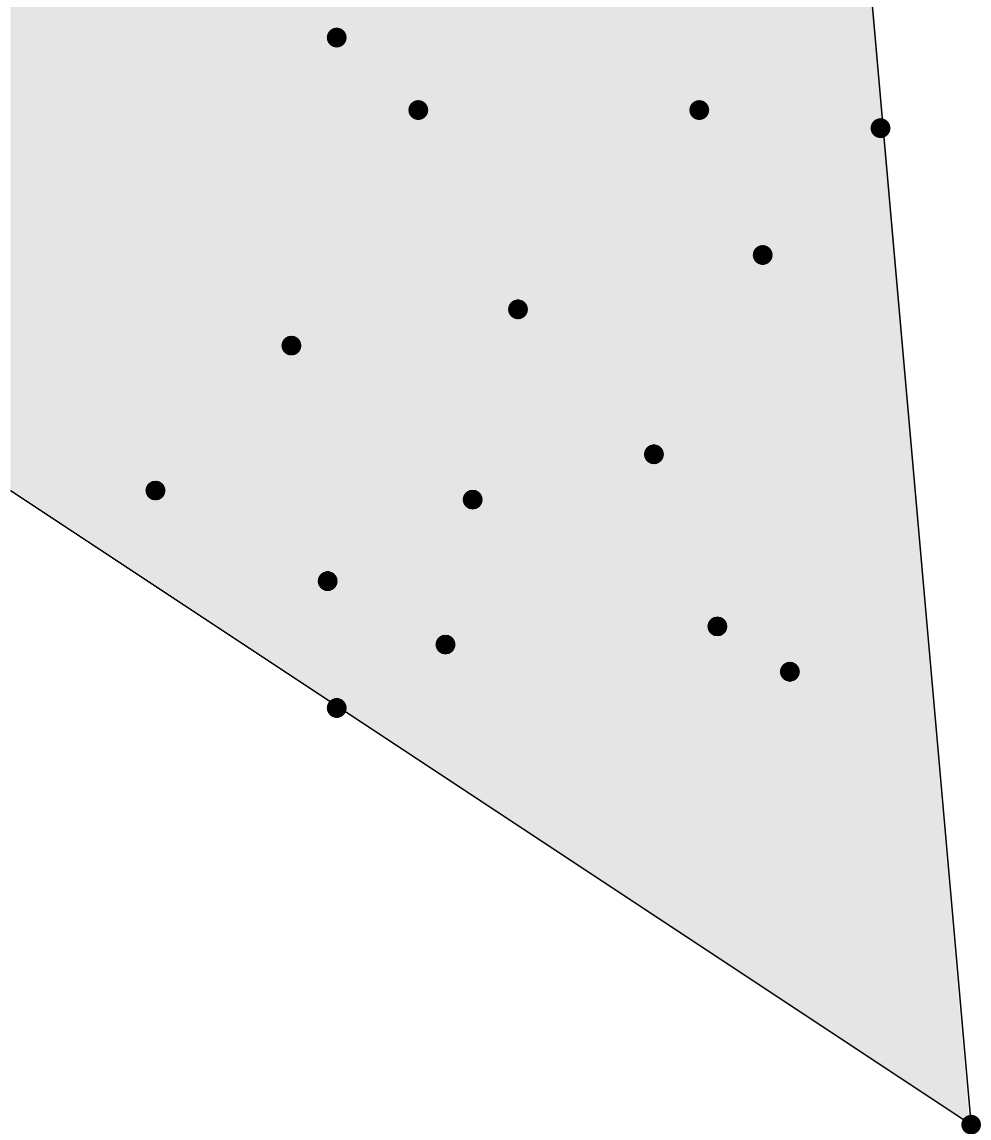
Conic combinations

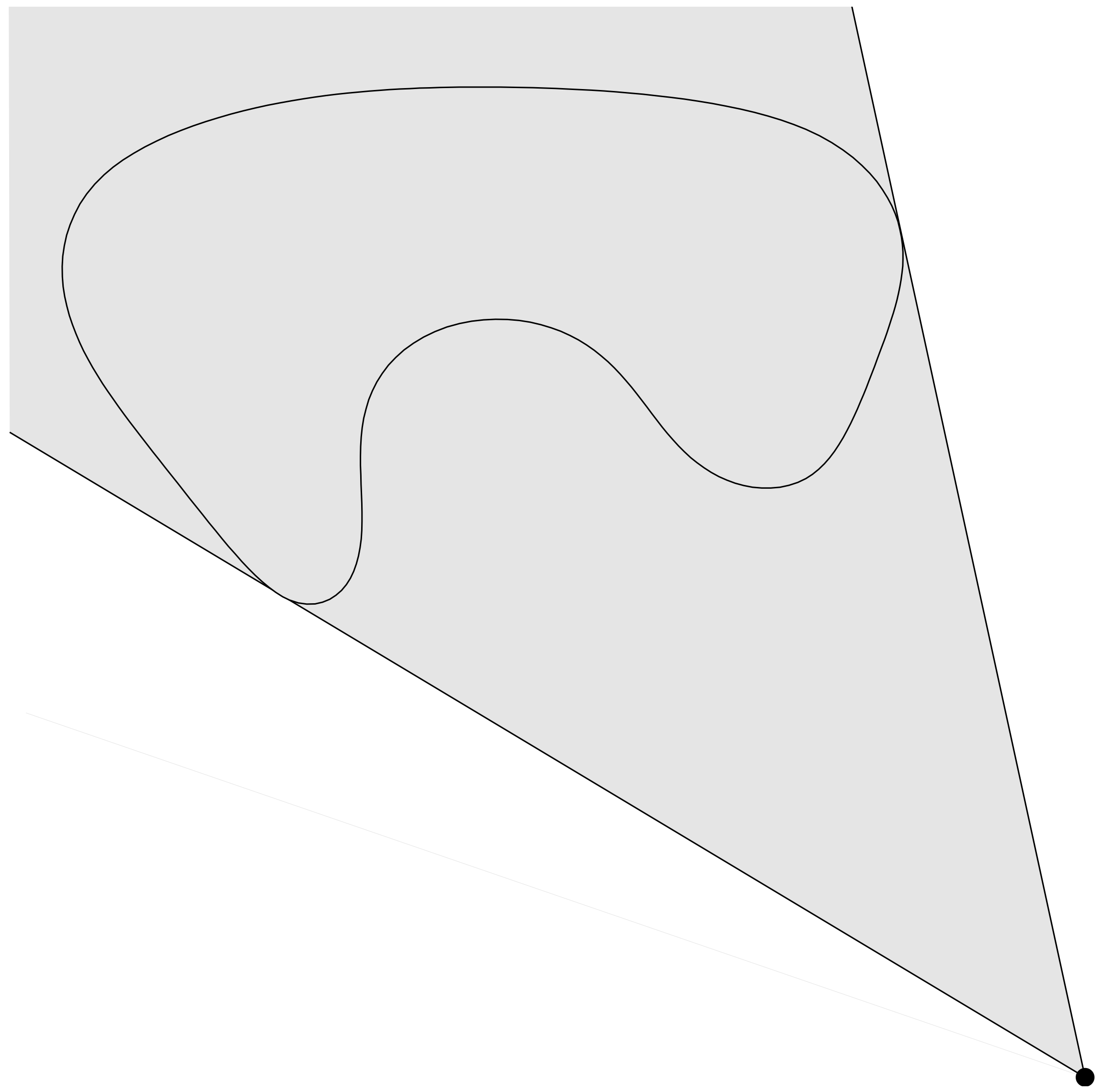
Conic combination

$\alpha_1 x_1 + \cdots + \alpha_k x_k$ for any $x_1, \ldots, x_k$ and $\alpha_1, \ldots, \alpha_k$ such that $\alpha_i \geq 0$

Conic hull

$$\left\{ \sum_{i=1}^{k} \alpha_i x_i \mid x_i \in C, \quad \alpha_i \geq 0, \quad i = 1, \ldots, k \right\}$$





Cones

Examples

Nonnegative orthant

$$\mathbf{R}_+^n = \{x \in \mathbf{R}^n \mid x \geq 0\}$$

Norm-cone

$\{(x, t) \mid \|x\| \leq t\}$ (if 2-norm, second-order cone)

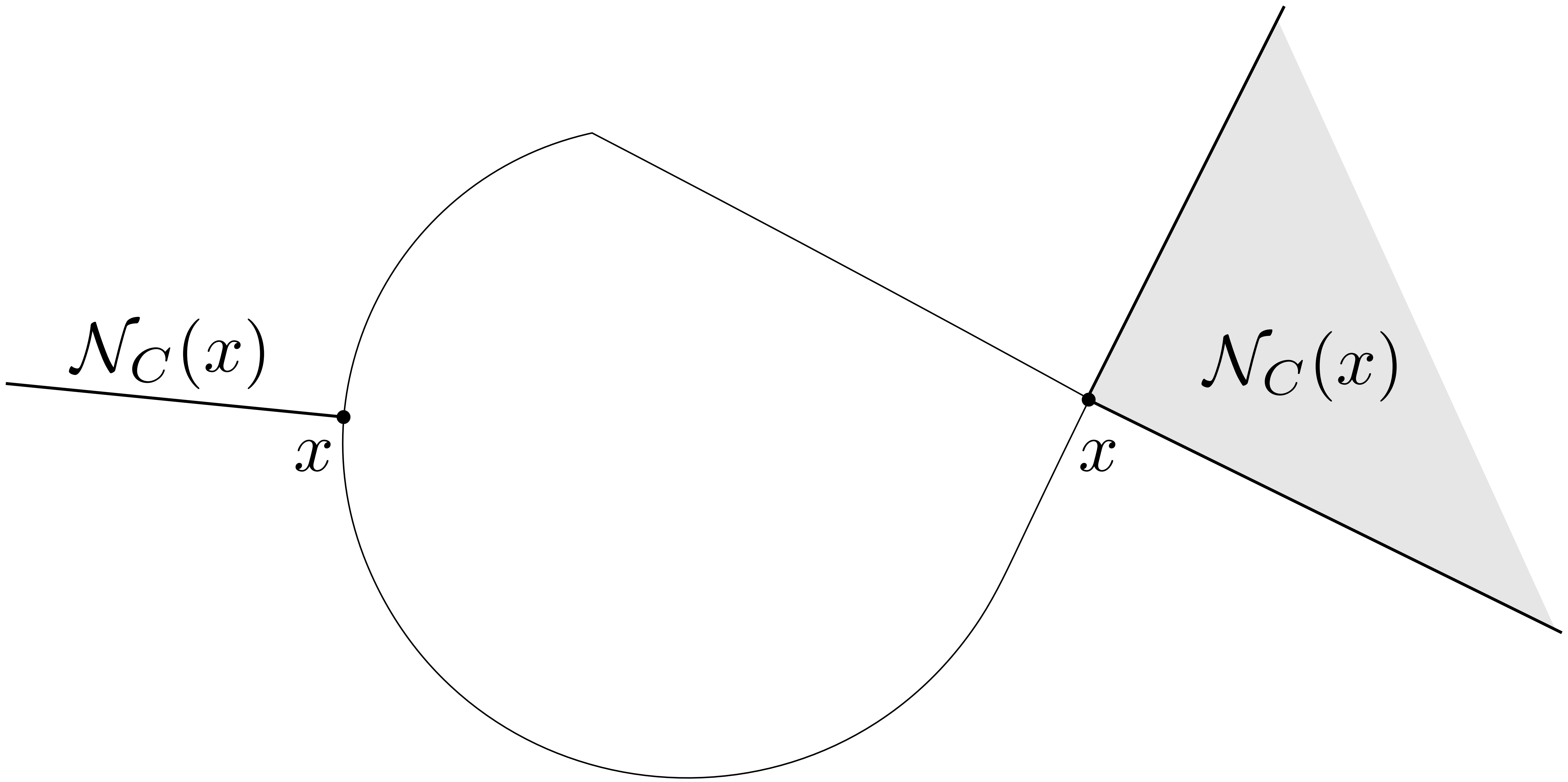
Positive semidefinite cone

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n \mid z^T X z \geq 0, \quad \text{for all} \quad z \in \mathbf{R}^n\}$$

Normal cone

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

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



$\mathcal{N}_C(x)$ is **always convex**

What if $x \in \mathbf{int} S$?

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

Hessian matrix (second derivative)

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

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

Example

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

Second-order approximation

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

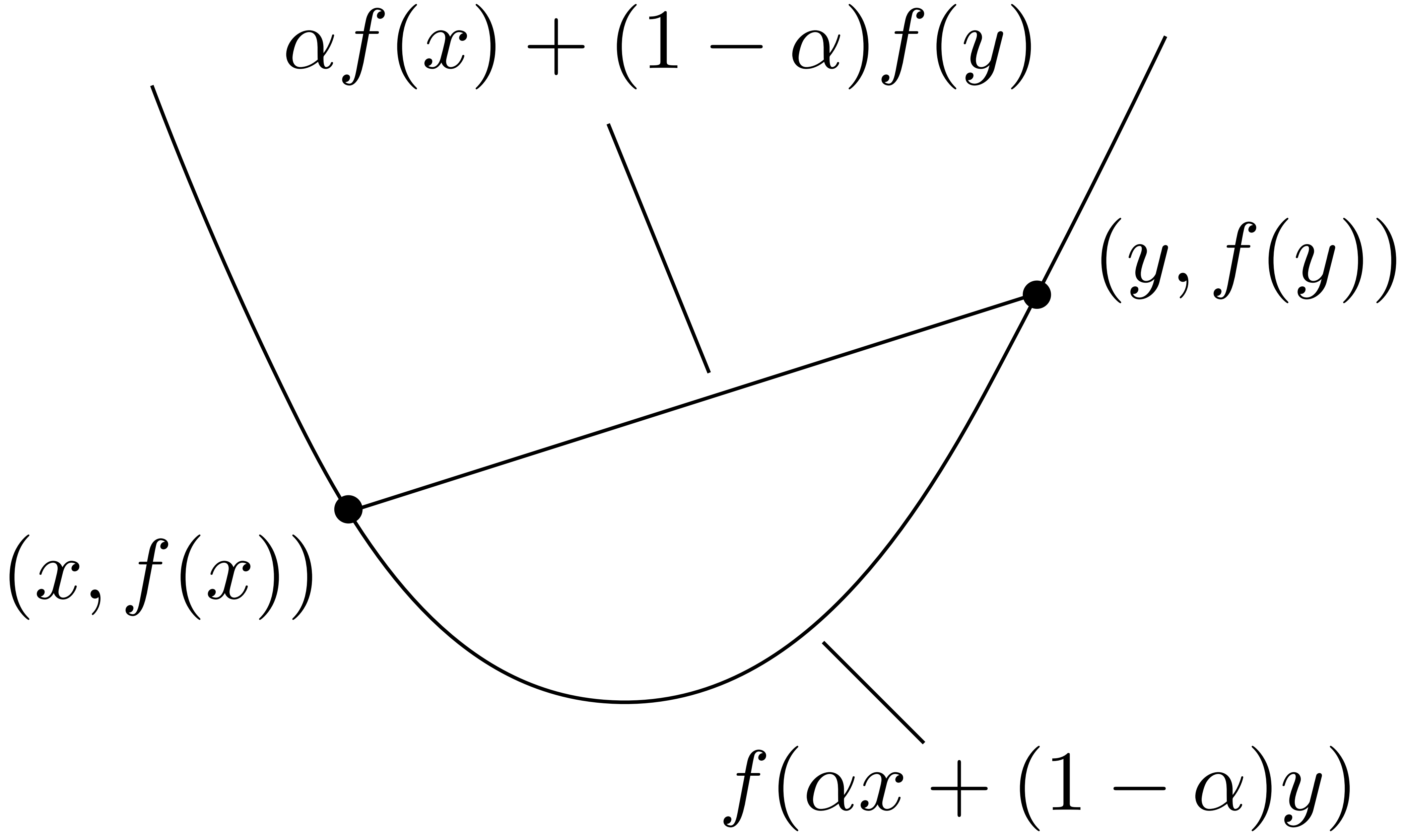
(quadratic function of y)

Convex optimization

Convex functions

Convex function

For every $x, y \in \mathbf{R}^n,\ \alpha \in [0,1]$ $$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$



Concave function

f is concave if and only if $-f$ is convex

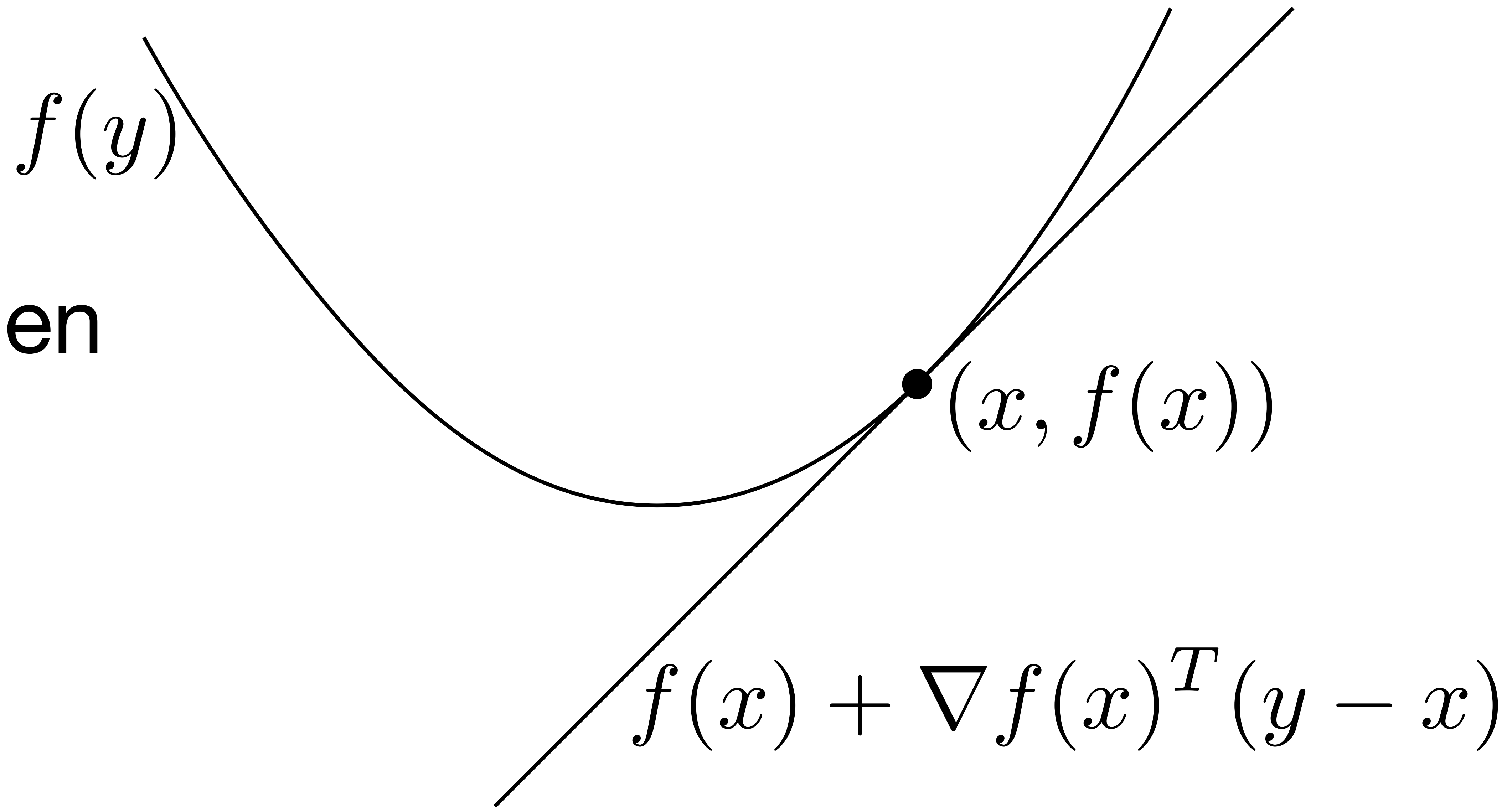
Convex conditions

First-order

Let f be a continuous differentiable function, then it is convex if and only if $\mathbf{dom} f$ is convex and

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

for all $x, y \in \mathbf{dom} f$



Second-order

If f is twice differentiable, then f is convex if and only if $\mathbf{dom} f$ is convex and

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

for all $x \in \mathbf{dom} f$

Verifying convexity

Basic definition (inequality)

First and second order conditions (gradient, hessian) ⟶ **Hard!**

Convex calculus (directly construct convex functions)

- Library of basic functions that are convex/concave
- Calculus rules or transformations that preserve convexity

⟶ **Easy!**

Disciplined Convex Programming

Convexity by construction

General composition rule

$h(f_1(x), f_2(x), \ldots, f_k(x))$ is convex when h is convex and for each i

Only sufficient condition

- h is nondecreasing in argument i and f_i is convex, or
- h is nonincreasing in argument i and f_i is concave, or
- f_i is affine

Check your functions at `https://dcp.stanford.edu/`

More details and examples in ORF523

Convex optimization problems

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

$f : \mathbf{R}^n \rightarrow \mathbf{R}$ Convex objective function

$g_i : \mathbf{R}^n \rightarrow \mathbf{R}$ Convex constraints functions

Convex feasible set

$$C = \{x \mid g_i(x) \leq 0, \quad i = 1, \ldots, m\}$$

Modelling software for convex optimization

Modelling tools simplify the formulation of convex optimization problems

- **Construct problems** using library of basic functions
- **Verify convexity** by general composition rule
- Express the problem in input format required by a specific solver

Examples

- CVX, YALMIP (Matlab)
- CVXPY (Python)
- Convex.jl (Julia)

Solving convex optimization problems

CVXPY

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2 \\ \text{subject to} & 0 \leq x \leq 1 \end{array}$$

```
x = cp.Variable(n)
objective = cp.Minimize(cp.norm(A*x - b))
constraints = [0 <= x, x <= 1]
problem = cp.Problem(objective, constraints)

# The optimal objective value is returned by `problem.solve()`.
result = problem.solve()

# The optimal value for x is stored in `x.value`.
print(x.value)
```

Local vs global minima (optimizers)

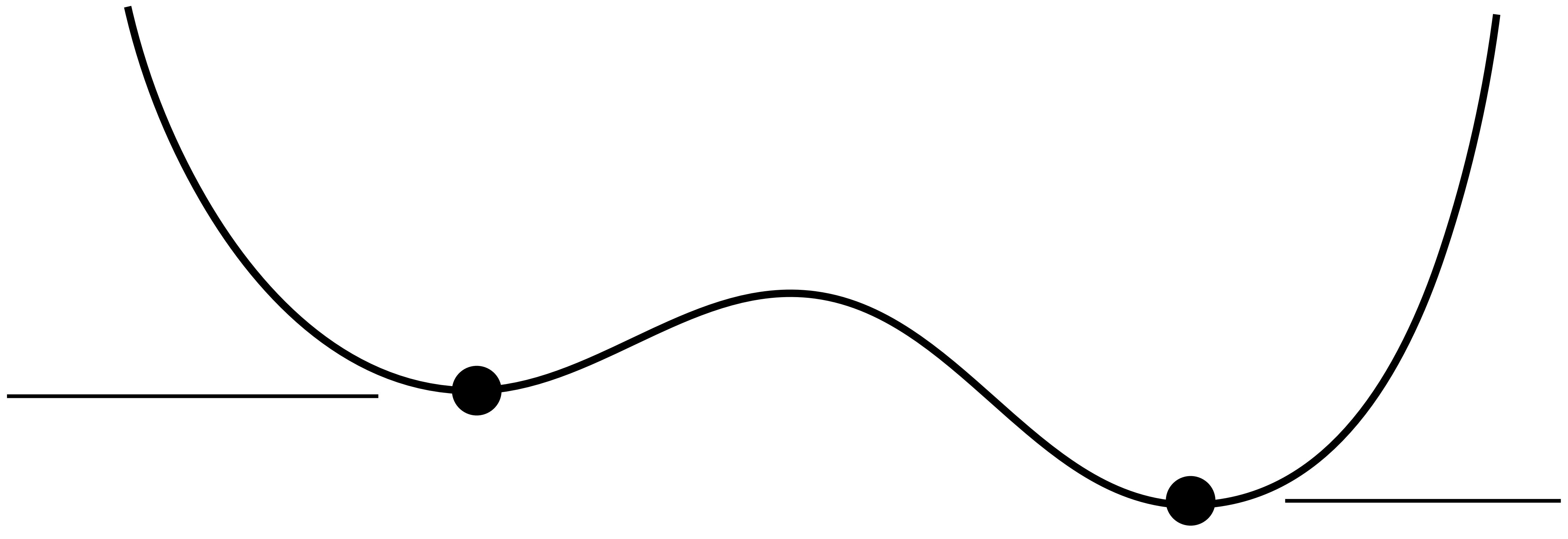
minimize $\quad f(x)$

subject to $\quad x \in C$

Local optimizer x

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

such that $\|x - y\|_2 \leq R$



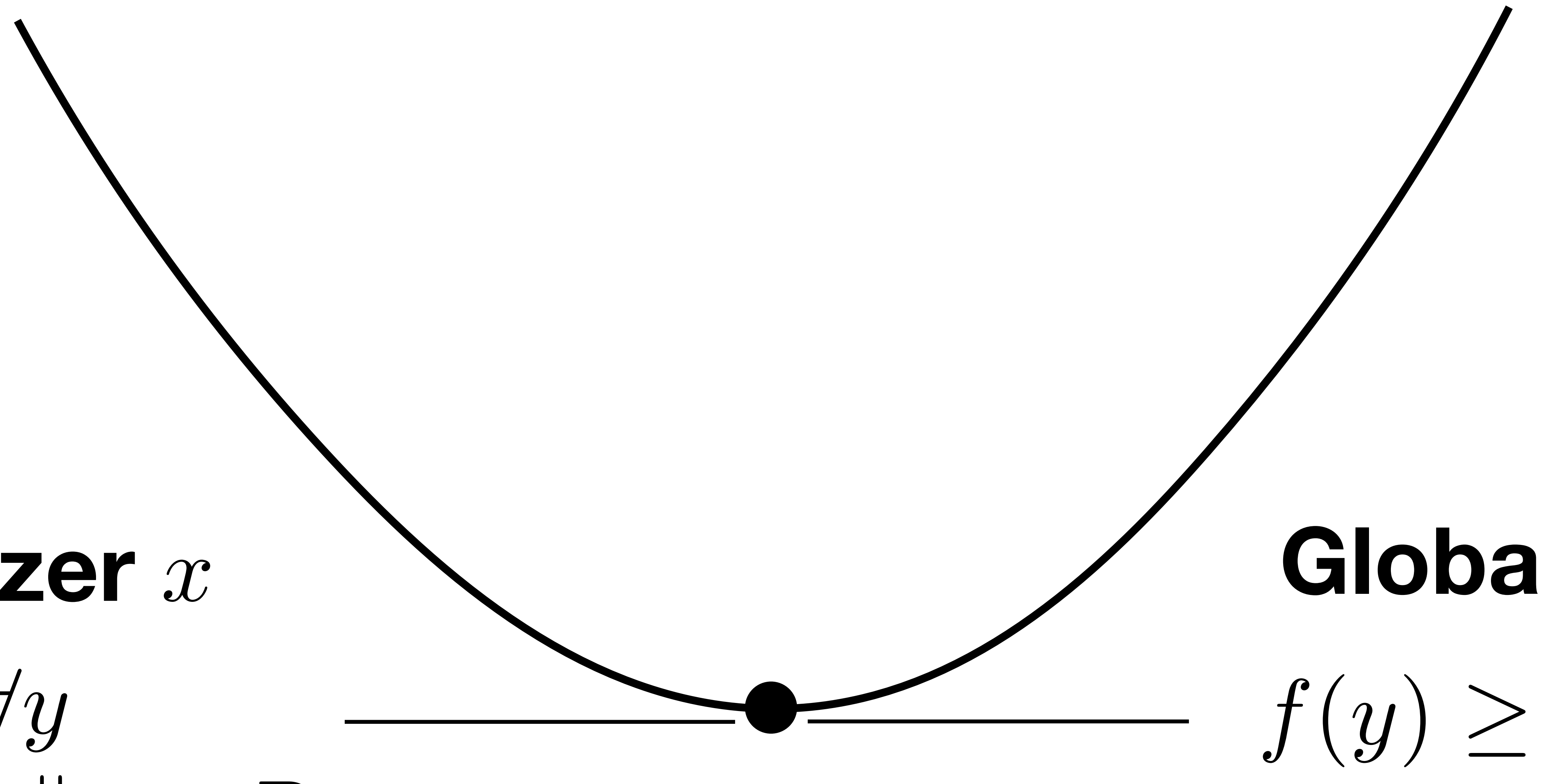
Global optimizer x

$f(y) \geq f(x), \quad \forall y \in C$

Optimality and convexity

Theorem

For a convex optimization problem, any **local minimum** is a **global minimum**



Local optimizer x

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

such that $\|x - y\|_2 \leq R$

Global optimizer x

$f(y) \geq f(x), \quad \forall y \in C$

Optimality and convexity

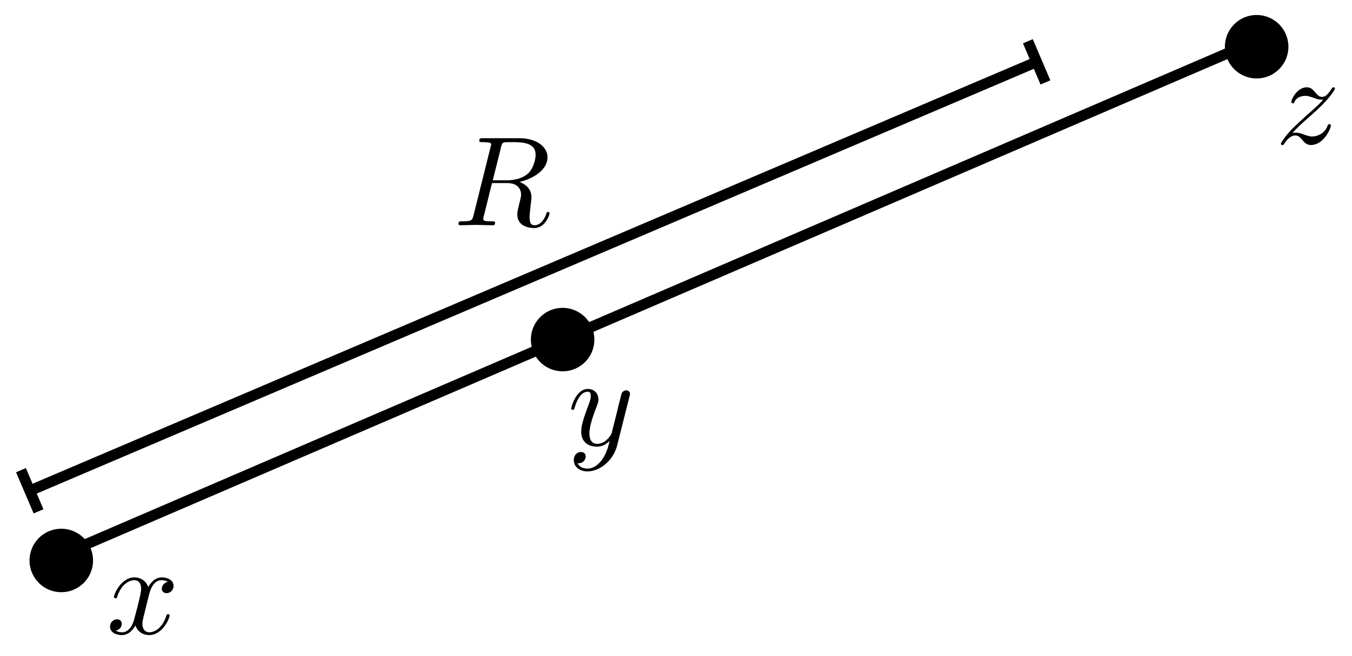
Proof (contradiction)

Suppose that f is convex and x is a local (not global) minimum for f, i.e.,

$$f(y) \geq f(x), \quad \forall y \text{ such that } \|x - y\|_2 \leq R.$$

Therefore, there exists a feasible z such that $\|z - x\| > R$ and $f(z) < f(x)$.

Consider $y = (1-\alpha)x + \alpha z$ with $\alpha = \dfrac{R}{2\|z - x\|_2}$.



Then, $\|y - x\|_2 = \alpha\|z - x\|_2 = R/2 < R$, and
by convexity of the feasible set, y is feasible.

By convexity of f we have $f(y) \leq (1-\alpha)f(x) + \alpha f(z) < f(x)$,
which contradicts the local optimum definition.

Therefore, x is globally optimal. ∎

"...in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

R. Tyrrell Rockafellar, in SIAM Review, 1993

Nonlinear optimization

Topics of this part of the course

Conditions to **characterize** minima

Algorithms to **find** (local) minima

(if applied to **convex problems,** they find **global minima**)

Introduction to nonlinear optimization

Today, we learned to:

- **Define** nonlinear optimization problems
- **Understand** convex analysis fundamentals (sets, cones, functions, and gradients)
- **Verify** convexity and **construct** convex optimization problems
- **Define** convex optimization problems in CVXPY
- **Understand** the importance of *convexity vs nonconvexity* in optimization

Next lecture

- Optimality conditions in nonlinear optimization