ORF307 – Optimization

8. Piecewise linear optimization

Bartolomeo Stellato — Spring 2023

Ed Forum

- What are exactly x_i^+ and x_i^-?
- What does it mean to eliminate free/unconstrained variables?

Recap

Standard form

Definition

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- Minimization
- Equality constraints
- Nonnegative variables

- Matrix notation for **theory**
- Standard form for **algorithms**

Standard form

Transformation tricks

Change objective

If “maximize”, use $-c$ instead of c and change to “minimize”.

Standard form

Transformation tricks

Change objective

If "maximize", use $-c$ instead of c and change to "minimize".

Eliminate inequality constraints

If $Ax \leq b$, define s and write $Ax + s = b, \; s \geq 0$.

If $Ax \geq b$, define s and write $Ax - s = b, \; s \geq 0$.

s are the **slack variables**

Standard form

Transformation tricks

Change objective

If "maximize", use $-c$ instead of c and change to "minimize".

Eliminate inequality constraints

If $Ax \leq b$, define s and write $Ax + s = b,\ s \geq 0$.

If $Ax \geq b$, define s and write $Ax - s = b,\ s \geq 0$.

s are the **slack variables**

Change variable signs

If $x_i \leq 0$, define $y_i = -x_i$.

Standard form

Transformation tricks

Change objective

If “maximize”, use $-c$ instead of c and change to “minimize”.

Eliminate inequality constraints

If $Ax \leq b$, define s and write $Ax + s = b, \ s \geq 0$.

If $Ax \geq b$, define s and write $Ax - s = b, \ s \geq 0$.

s are the **slack variables**

Change variable signs

If $x_i \leq 0$, define $y_i = -x_i$.

Eliminate “free” variables

If x_i unconstrained, define $x_i = x_i^+ - x_i^-$, with $x_i^+ \geq 0$ and $x_i^- \geq 0$.

Standard form

Transformation example

$$\begin{array}{lrcrl} \text{minimize} & 2x_1 & + & 4x_2 & \\ \text{subject to} & x_1 & + & x_2 & \geq 3 \\ & 3x_1 & + & 2x_2 & = 14 \\ & x_1 & & & \geq 0 \end{array}$$

$$\downarrow$$

$$\begin{array}{lrcrcrcrl} \text{minimize} & 2x_1 & + & 4x_2^+ & - & 4x_2^- & & & \\ \text{subject to} & x_1 & + & x_2^+ & - & x_2^- & - & x_3 & = 3 \\ & 3x_1 & + & 2x_2^+ & - & 2x_2^- & & & = 14 \\ & x_1, & & x_2^+, & & x_2^-, & & x_3 & \geq 0. \end{array}$$

Today's lecture

Piecewise linear optimization

- Vector norms
- Piecewise linear optimization
- Turning vector norm problems as LPs
- Support vector machines

Vector norms

Vector norms

Euclidean norm

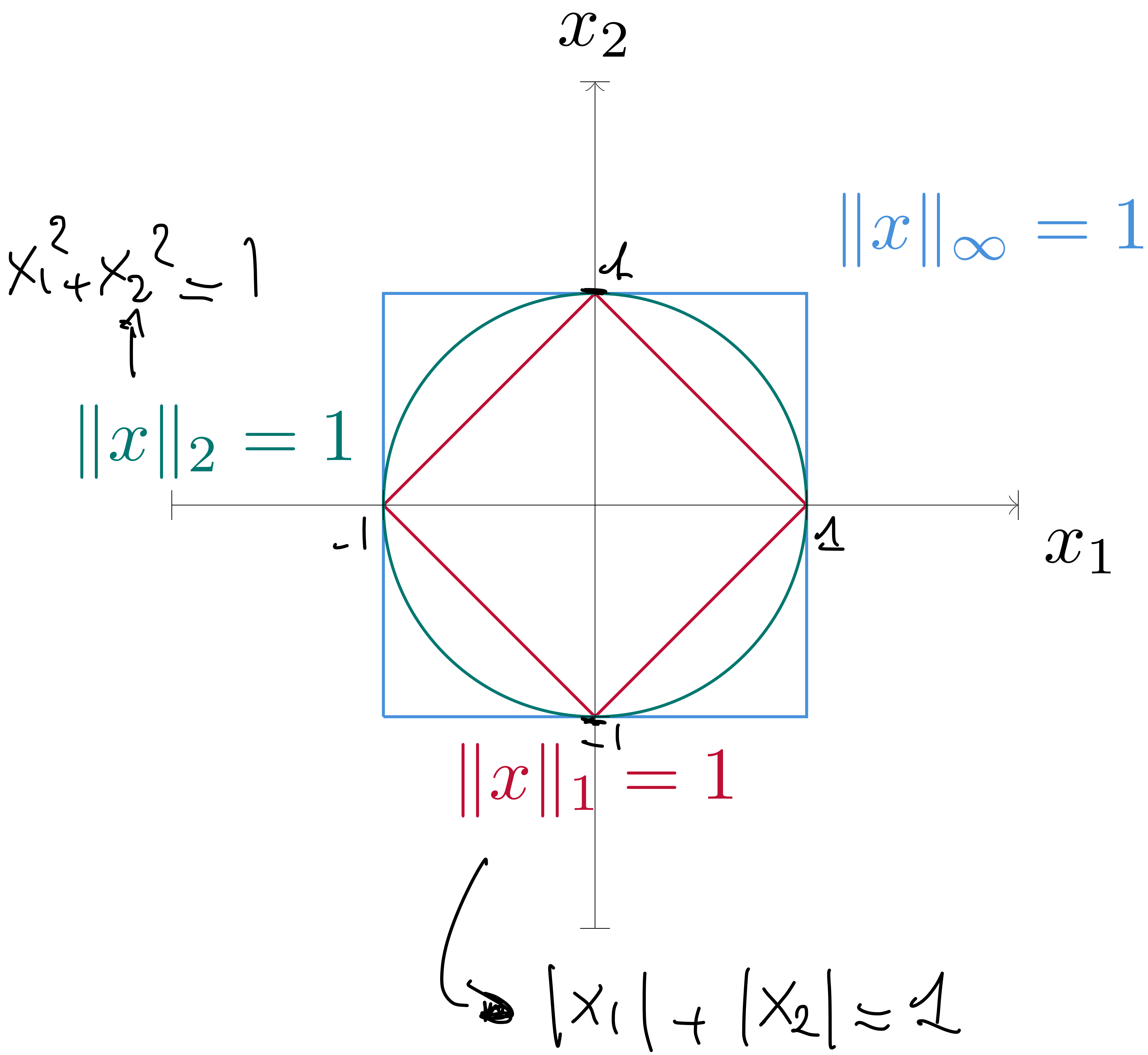
$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

1-norm (Manhattan norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

∞-norm (max-norm)

$$\|x\|_\infty = \max_i |x_i|$$

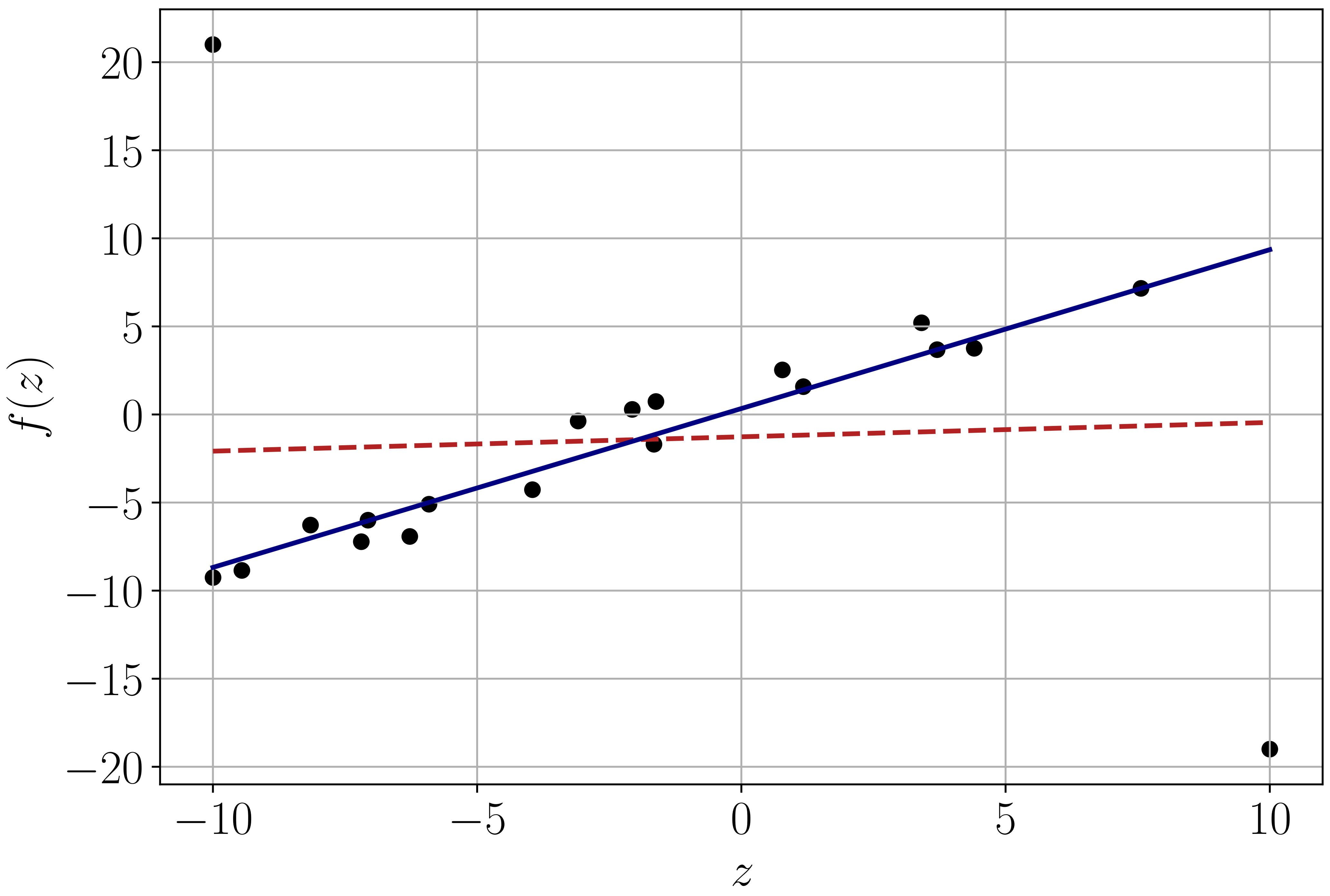


Data-fitting example

Fit a linear function $f(z) = a + bz$ to m data points (z_i, f_i):

Approximation problem $Ax \approx b$ where

$$\underbrace{\begin{bmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_m \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{x} \approx \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{b}$$



Recall our regression problem:

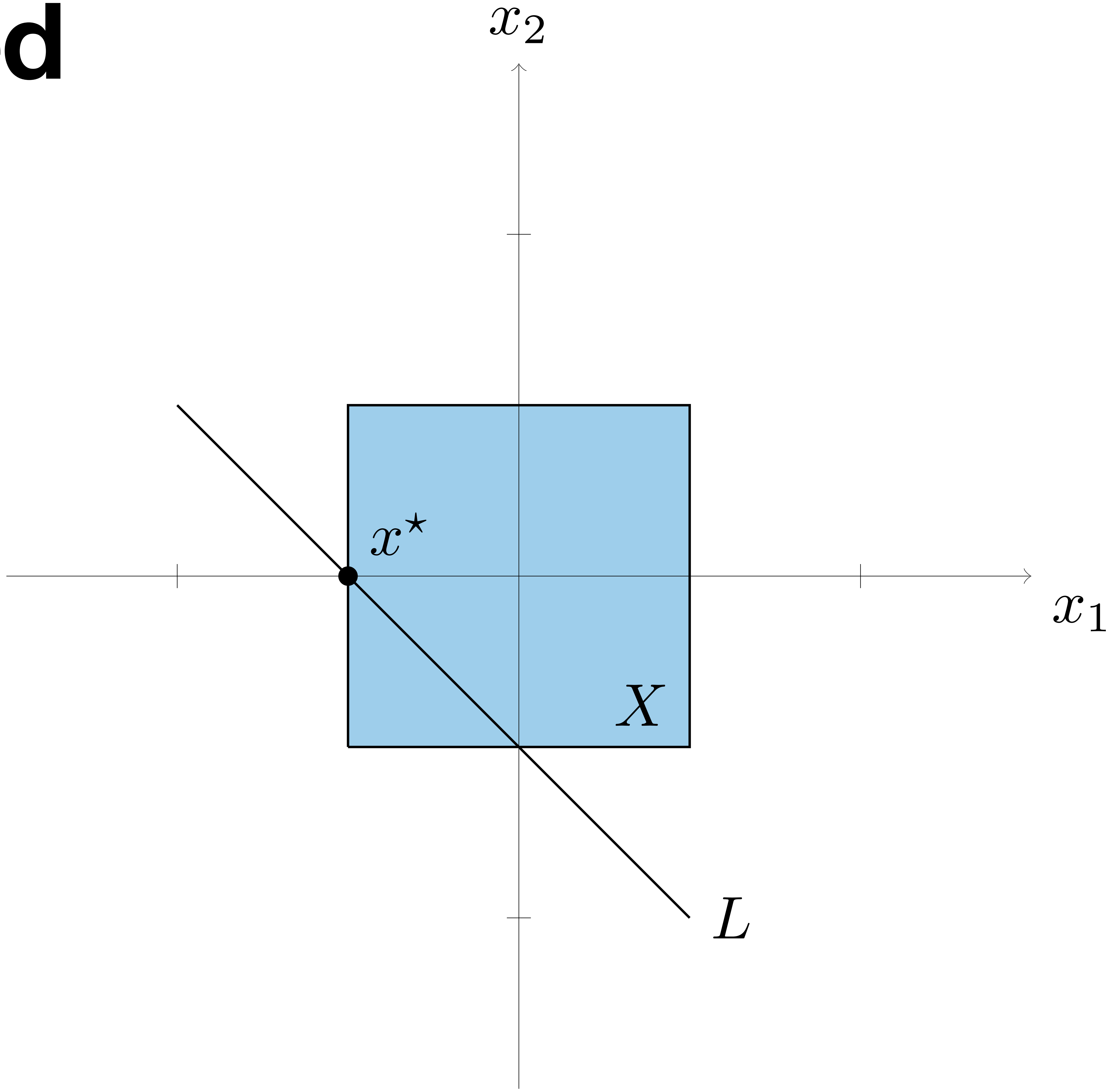
minimize $\displaystyle\sum_{i=1}^{m} |Ax - b|_i = \|Ax - b\|_1$

Why is it a linear program?

Simple example revisited

Goal find point as far left as possible,
in the unit box X,
and restricted to the line L

$$\begin{array}{ll} \text{minimize} & x_1 \\ \text{subject to} & \|x\|_\infty \leq 1 \\ & x_1 + x_2 = -1 \end{array}$$



Simple example revisited

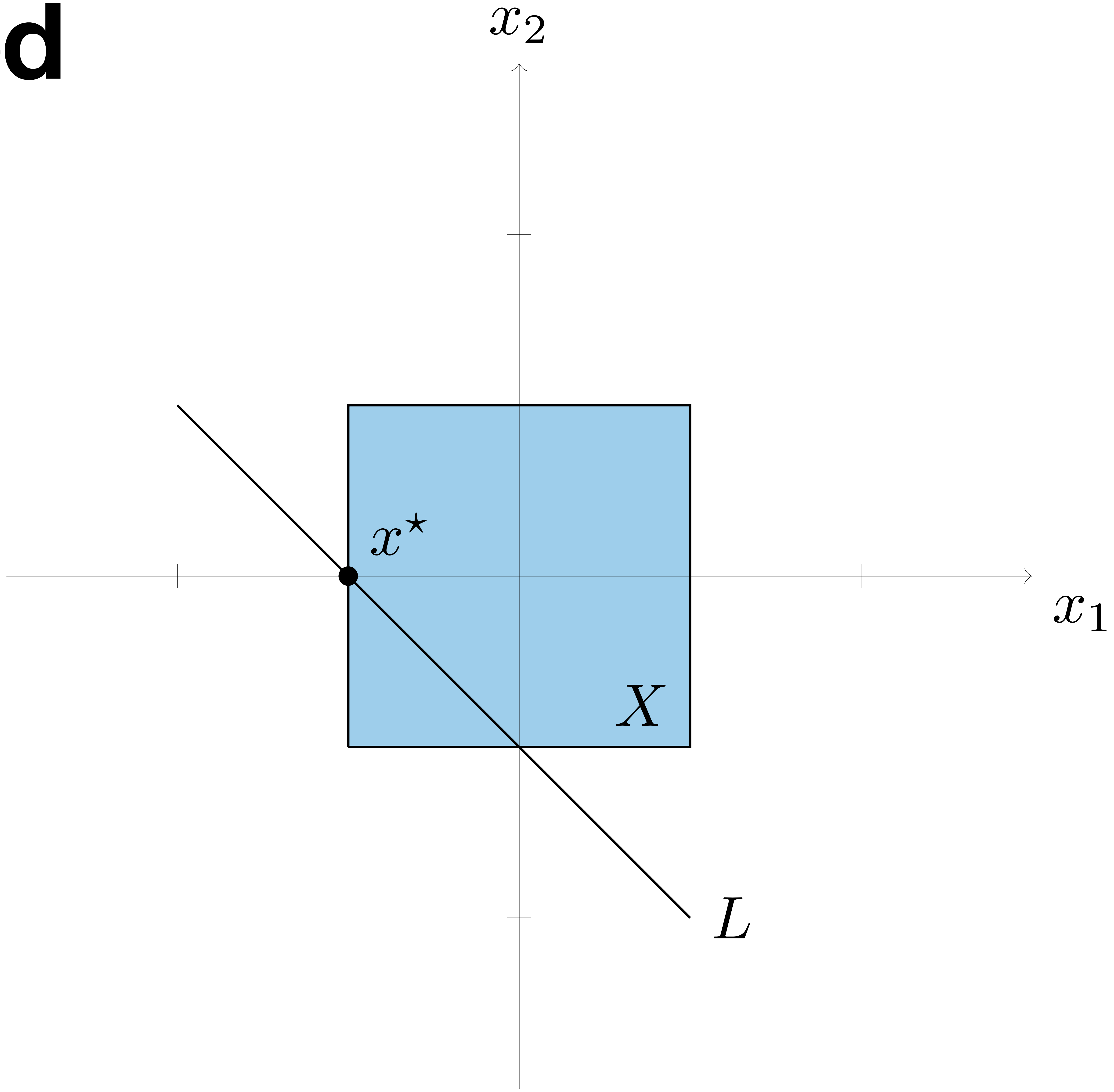
Goal find point as far left as possible,
in the unit box X,
and restricted to the line L

$$
\begin{array}{ll}
\text{minimize} & x_1 \\
\text{subject to} & \|x\|_\infty \leq 1 \\
 & x_1 + x_2 = -1
\end{array}
$$

$-1 \leq x_1 \leq 1$
$-1 \leq x_2 \leq 1$

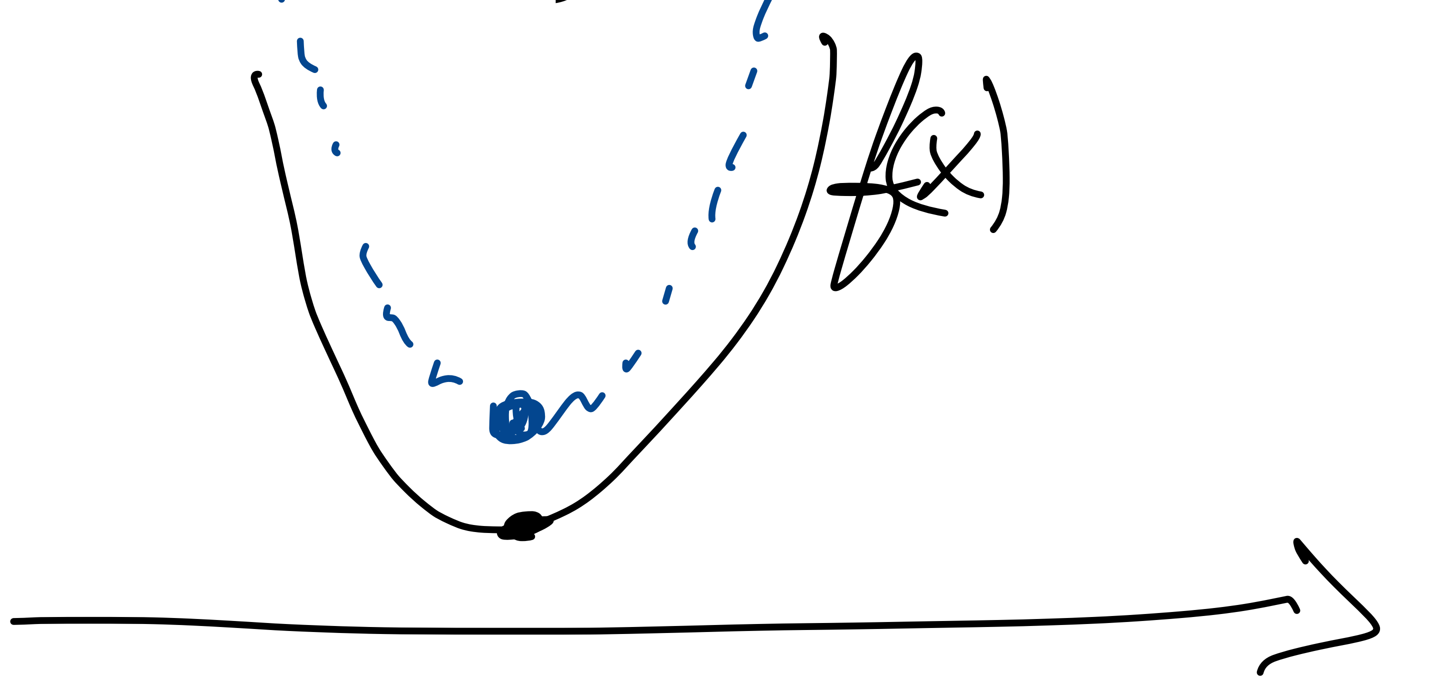
The (nonlinear) norm function appears in the constraints

Why is it a linear progam?



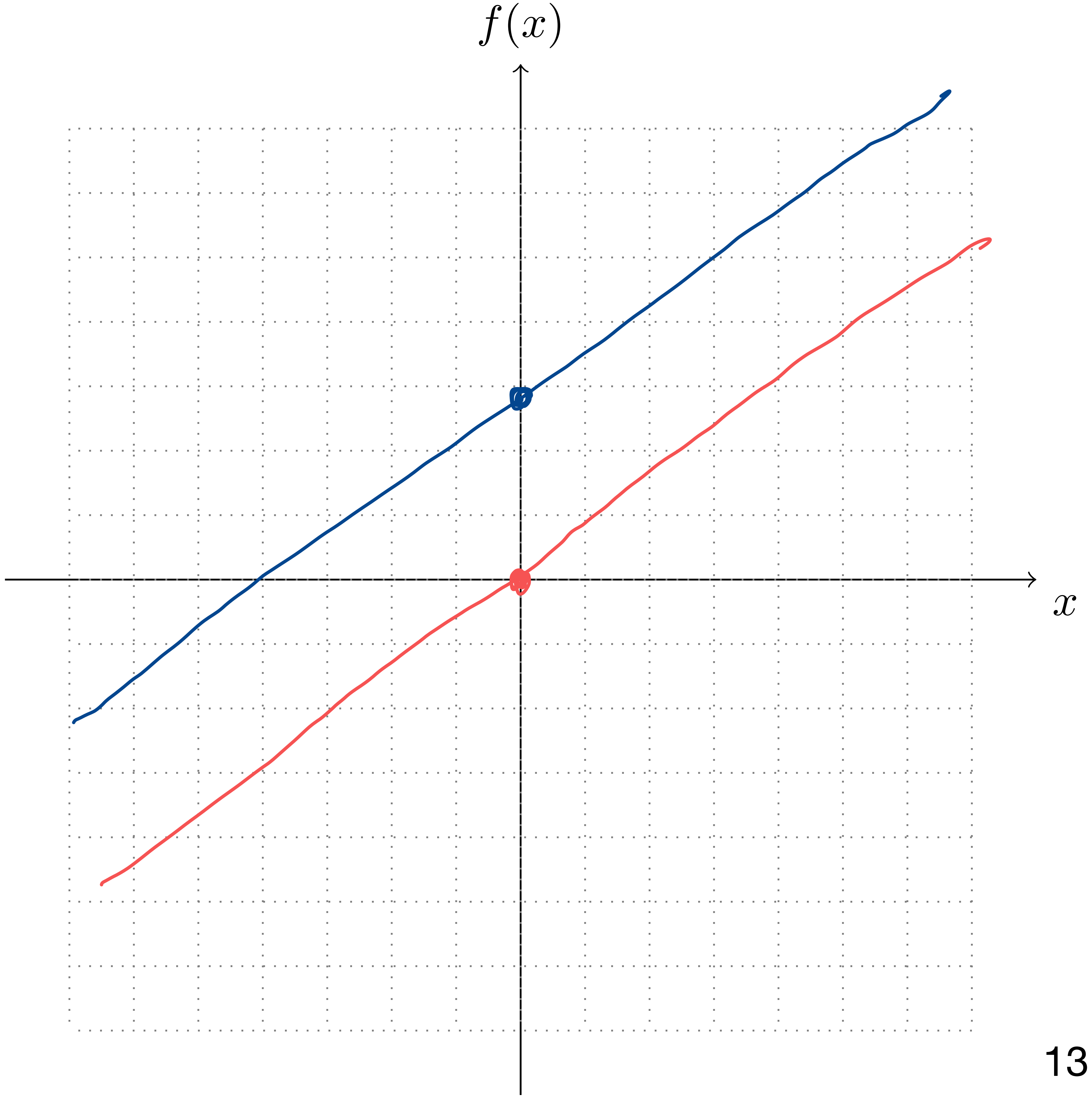
Piecewise linear optimization

Linear, affine and convex functions



Linear function: $f(x) = a^T x$

Affine function: $f(x) = a^T x + b$

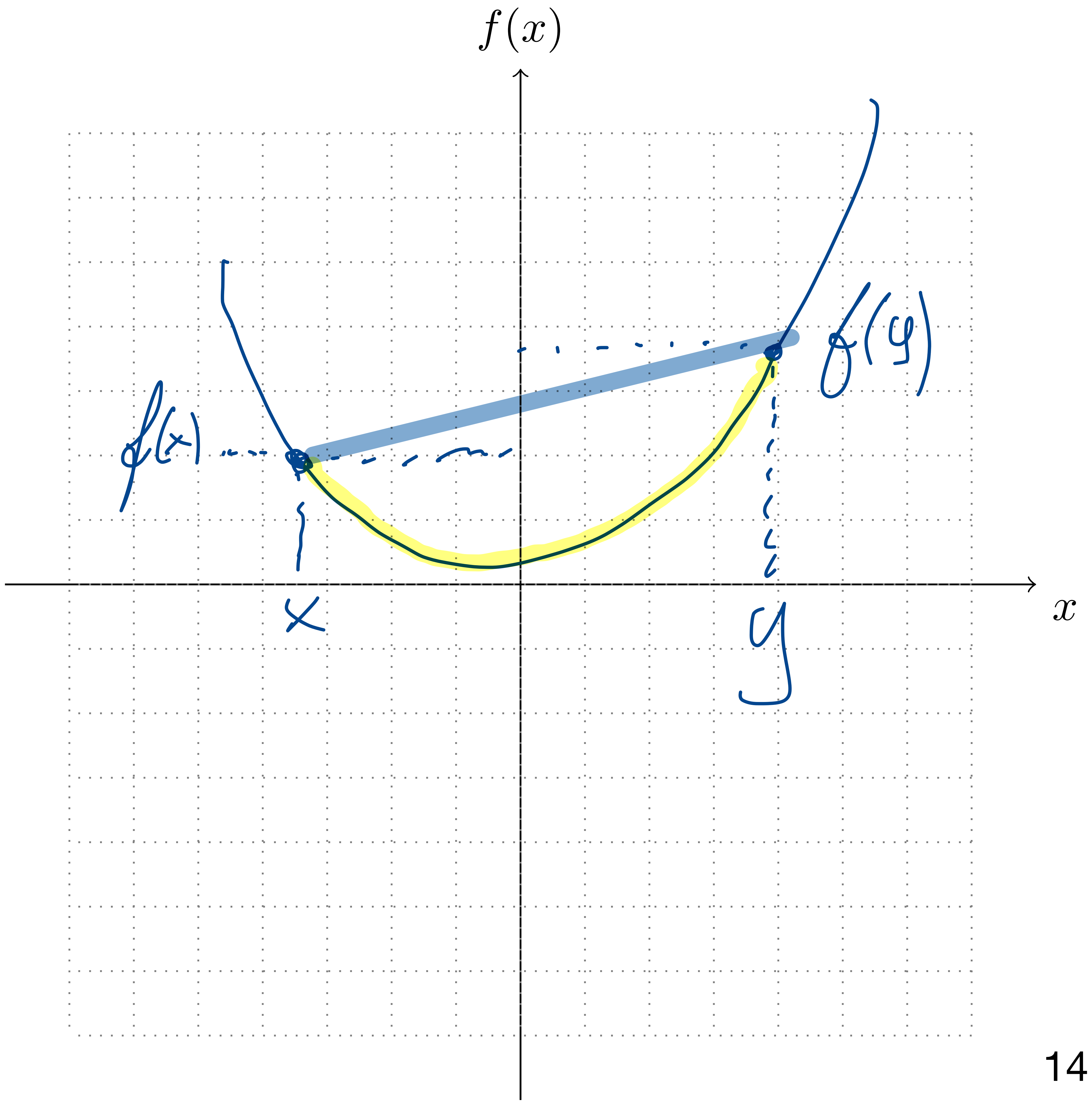


Linear, affine and convex functions

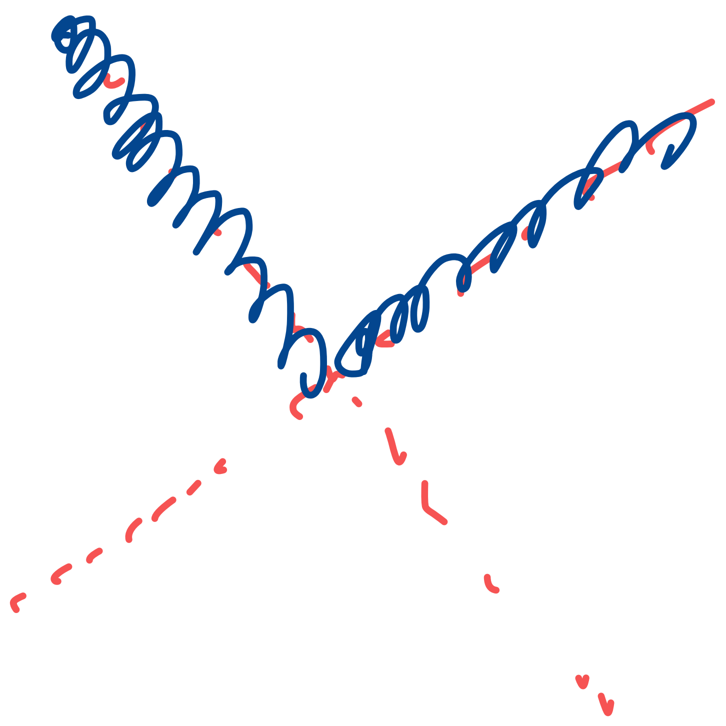
Convex function:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y),$$

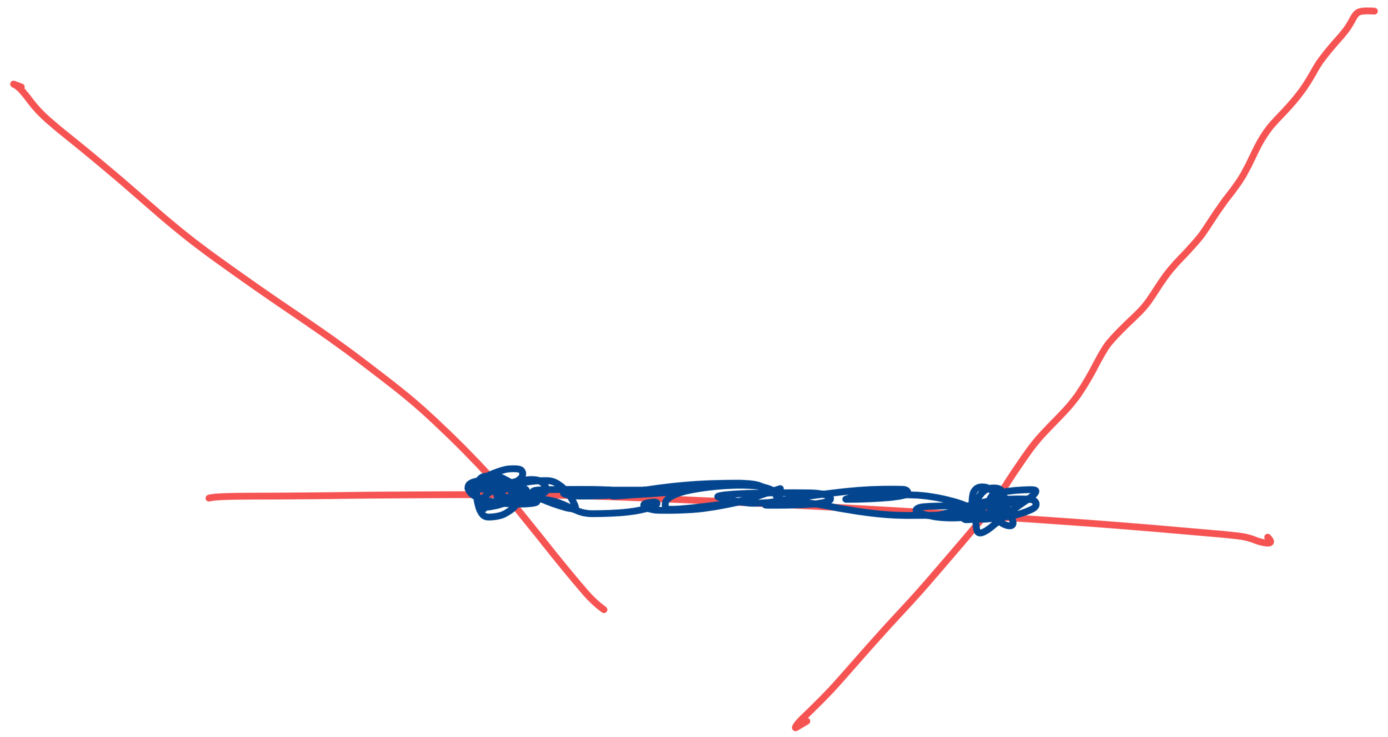
$\forall x, y \in \mathbf{R}^n, \; \alpha \in [0, 1]$

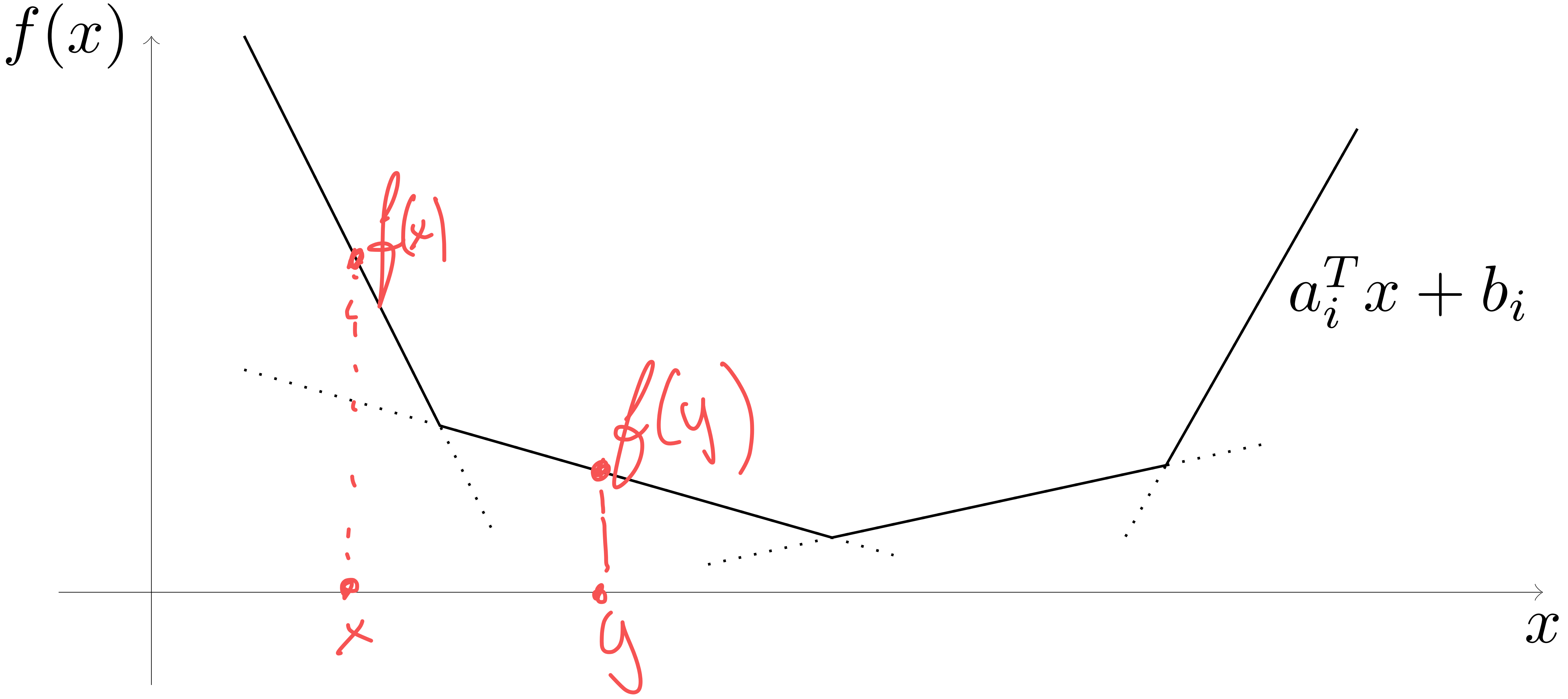


Convex piecewise-linear functions



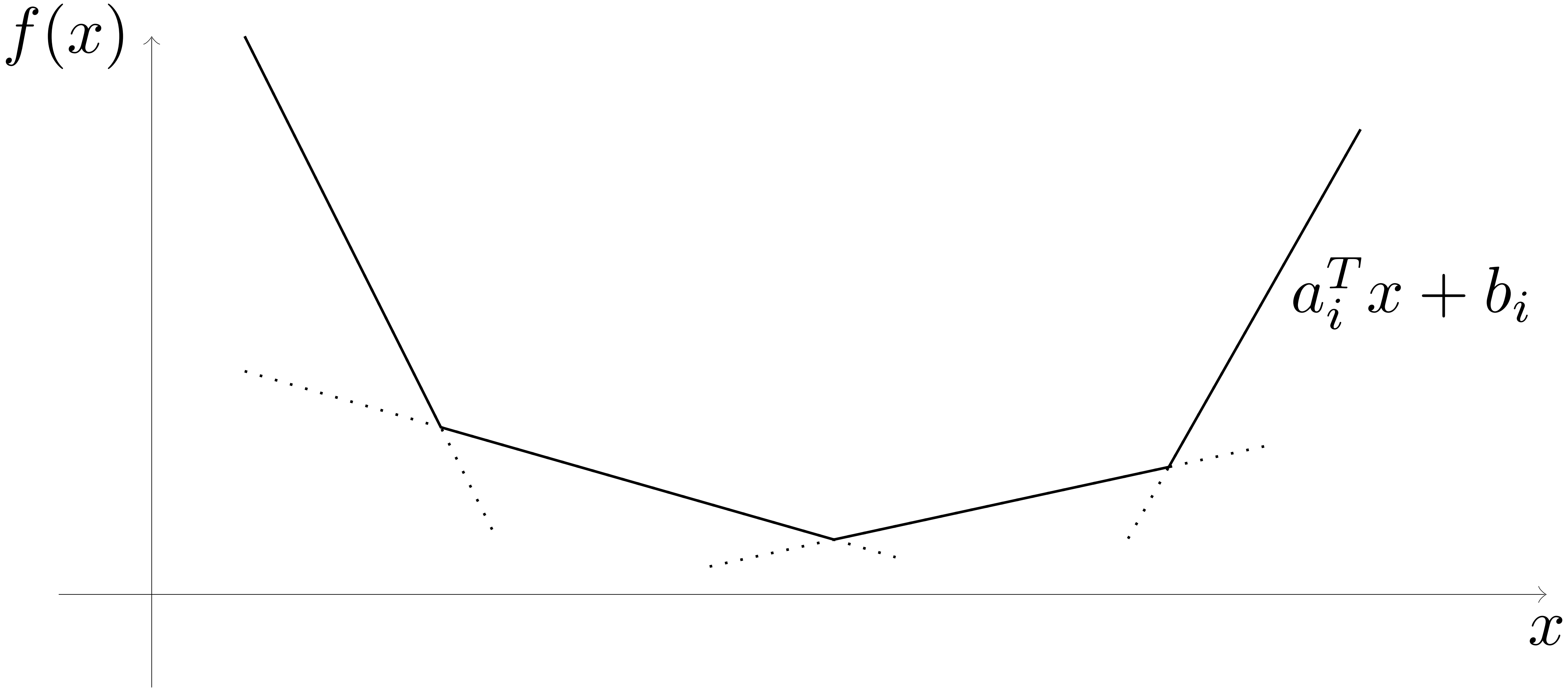
$$f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i)$$





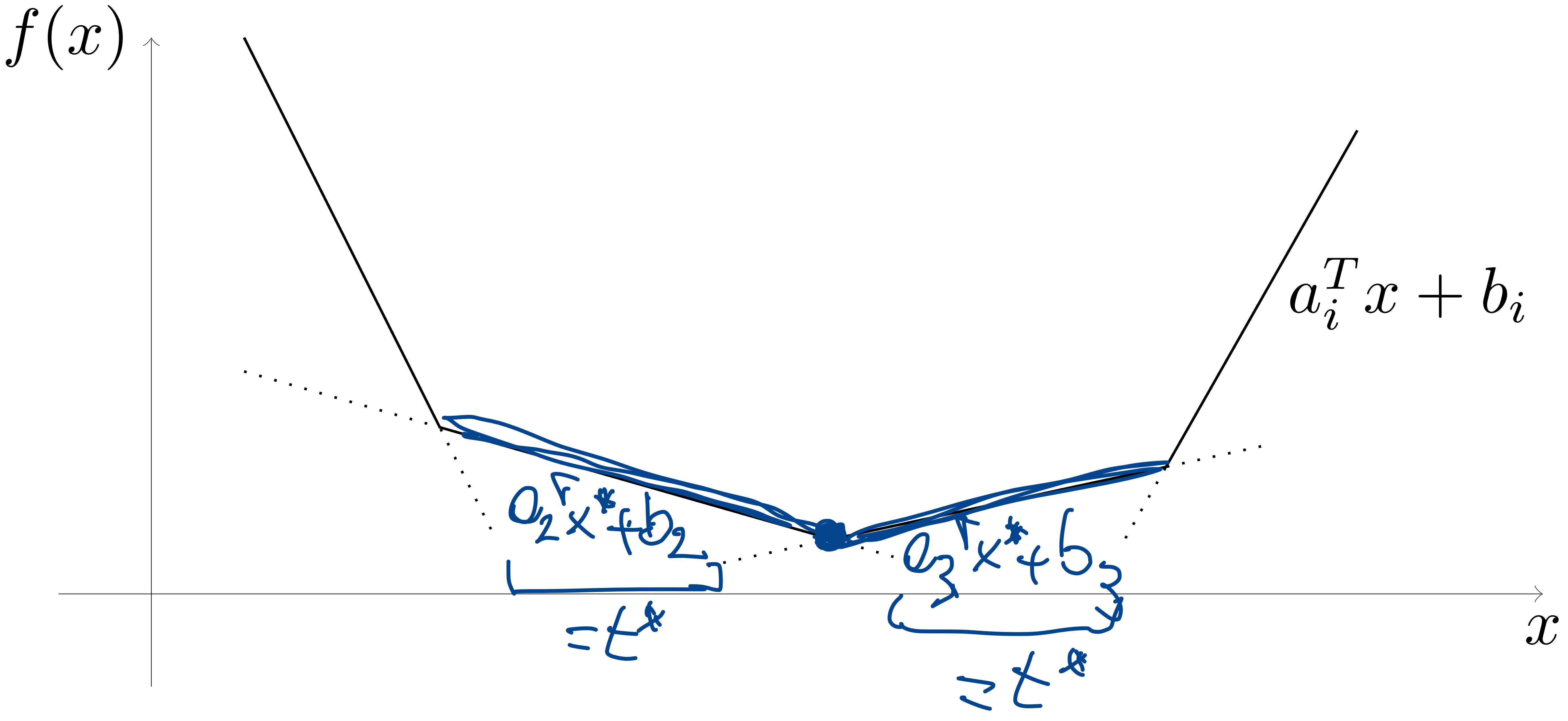
Convex piecewise-linear minimization

minimize $\max_{i=1,\ldots,m} (a_i^T x + b_i)$



Convex piecewise-linear minimization

minimize $\max_{i=1,\ldots,m} (a_i^T x + b_i)$



Equivalent linear optimization

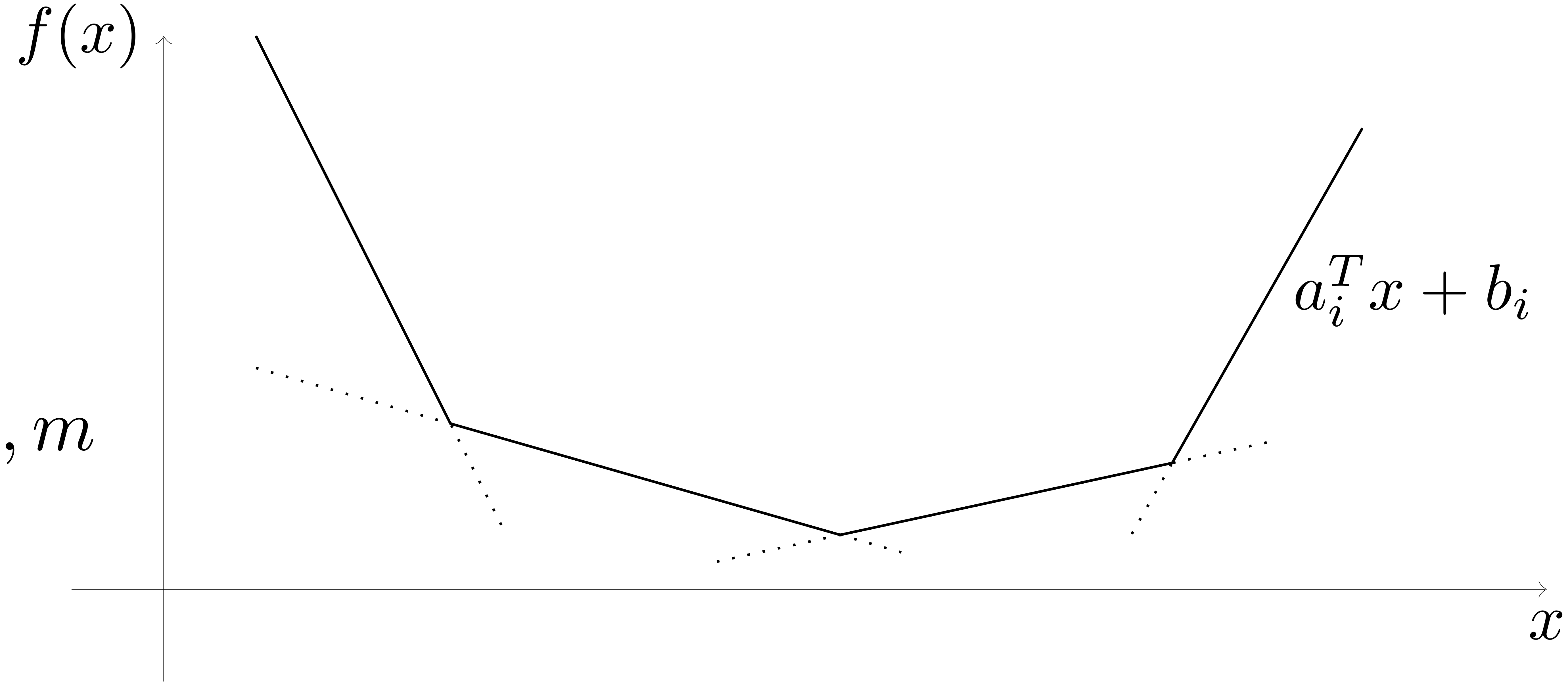
minimize t

subject to $a_i^T x + b_i \leq t, \quad i = 1, \ldots, m$

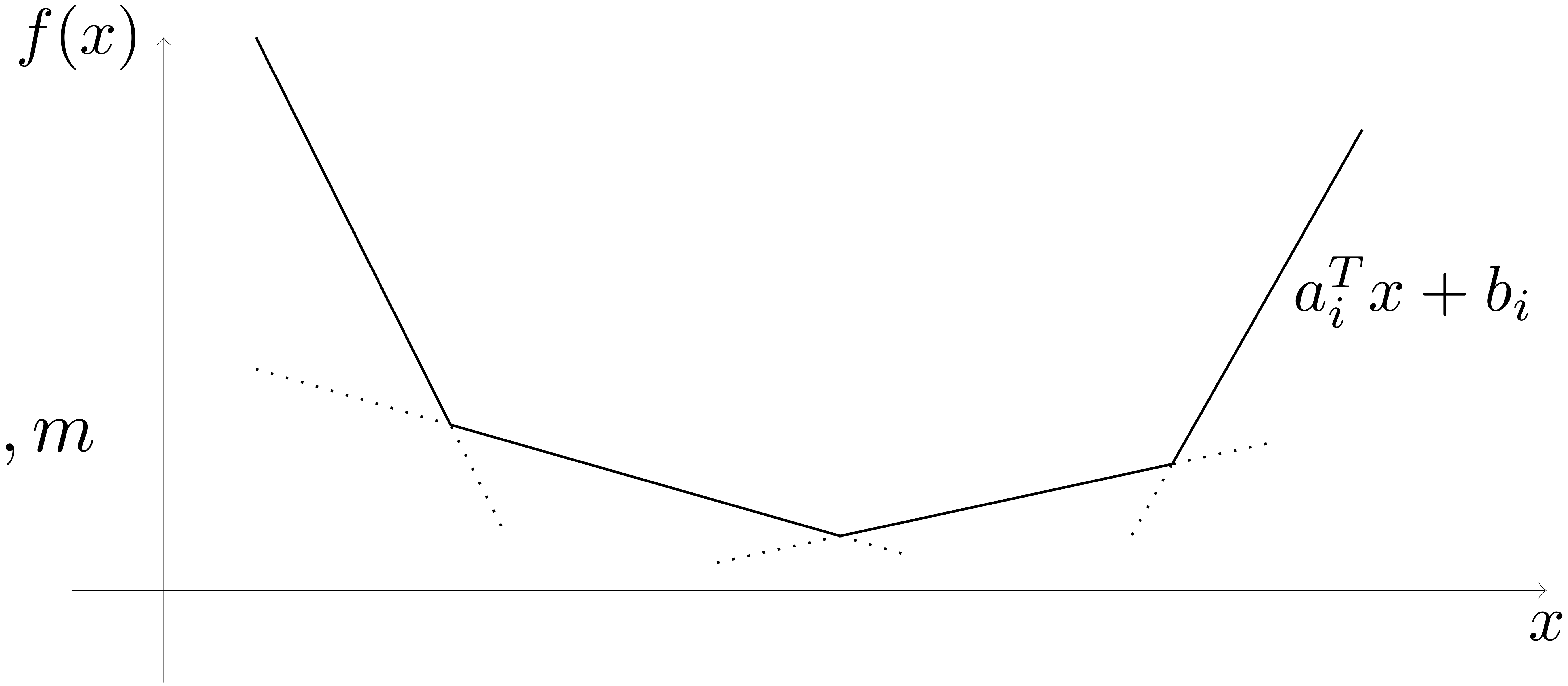
Convex piecewise-linear minimization

Equivalent linear optimization

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \dots, m \end{array}$$



Convex piecewise-linear minimization



Equivalent linear optimization

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & a_i^T x + b_i \le t, \quad i = 1, \dots, m
\end{array}
$$

$$a_i^T x - t \le -b_i$$

$$\begin{bmatrix} a_i^T & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \le -b_i$$

Matrix notation

$$
\begin{array}{ll}
\text{minimize} & \tilde{c}^T \tilde{x} \\
\text{subject to} & \tilde{A}\tilde{x} \le \tilde{b}
\end{array}
$$

$$\tilde{x} = \begin{bmatrix} x \\ t \end{bmatrix}, \quad \tilde{c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} a_1^T & -1 \\ \vdots & \vdots \\ a_m^T & -1 \end{bmatrix}, \quad \tilde{b} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \end{bmatrix}$$

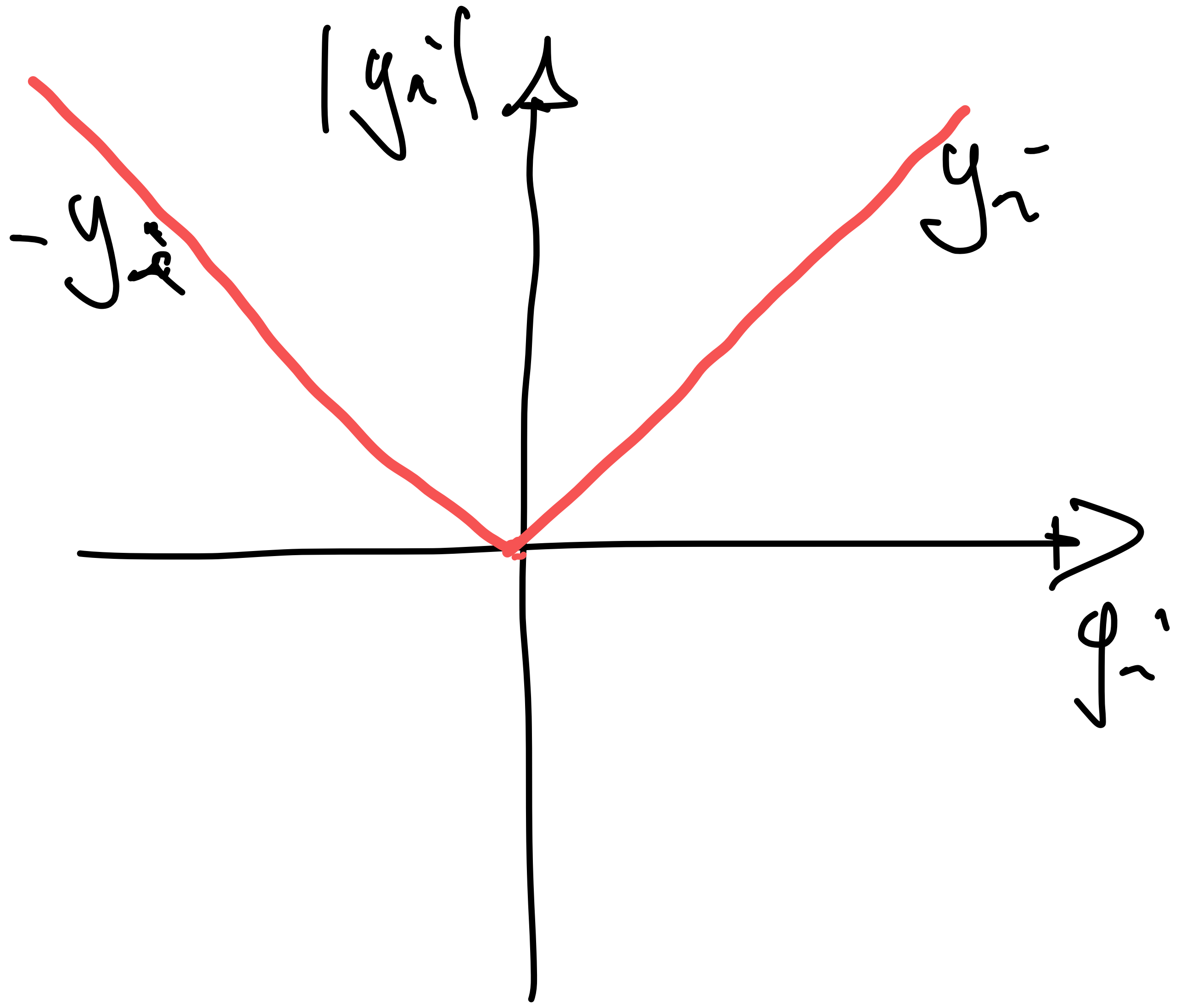
Vector norm problems as linear optimization

∞-norm regression

$$\text{minimize} \quad \|Ax - b\|_\infty$$

The ∞**-norm** of m-vector y is

$$\|y\|_\infty = \max_{i=1,\ldots,m} |y_i| = \max_{i=1,\ldots,m} \max\{y_i, -y_i\}$$



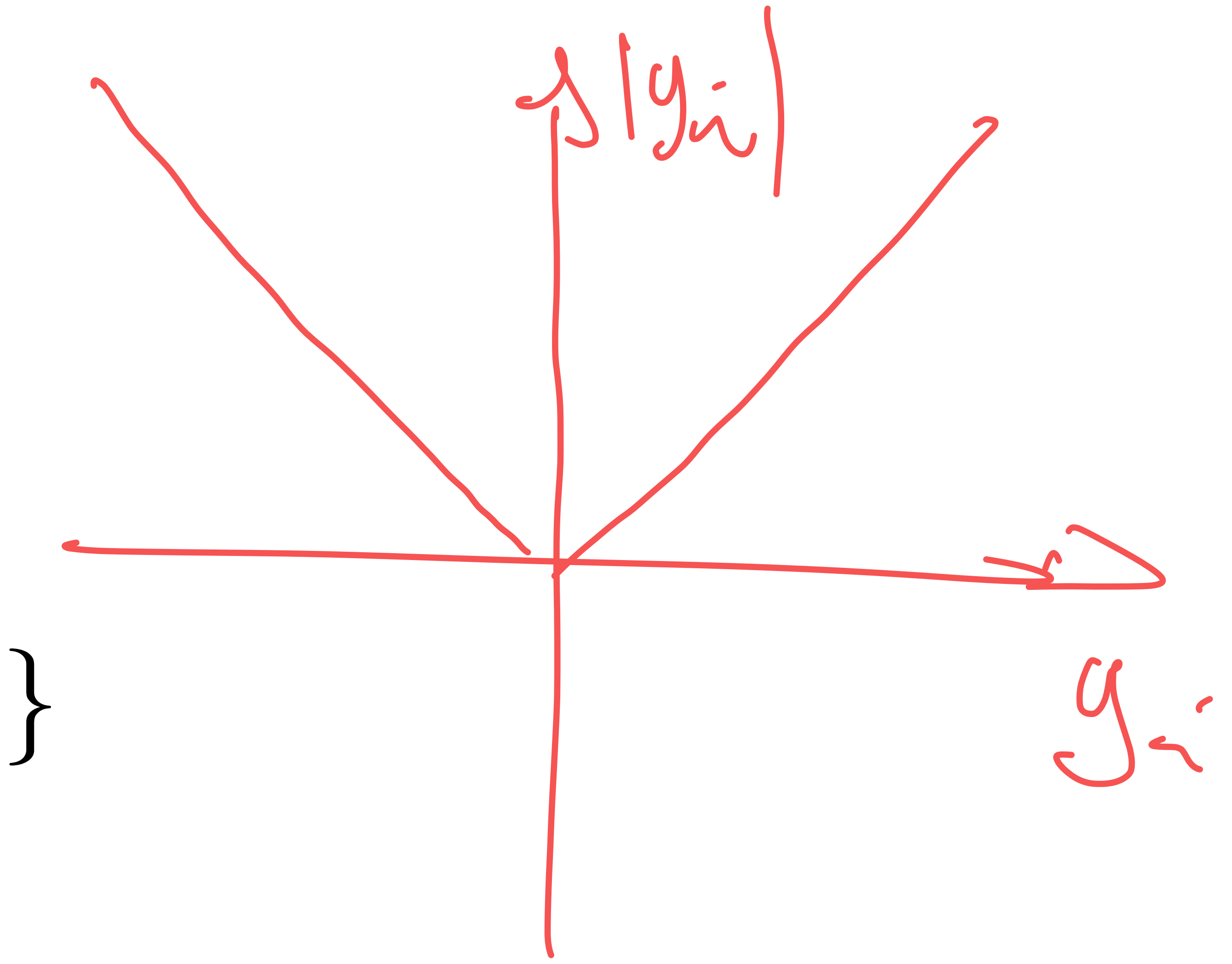
∞-norm regression

$y_i = (Ax - b)_i$

$$\text{minimize} \quad \|Ax - b\|_\infty$$

The ∞**-norm** of m-vector y is

$$\|y\|_\infty = \max_{i=1,\ldots,m} |y_i| = \max_{i=1,\ldots,m} \max\{y_i, -y_i\}$$



min t
s.t. $|y_i| \leq t \quad i = 1, \ldots, m$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & (Ax - b)_i \leq t, \quad i = 1, \ldots, m \\ & -(Ax - b)_i \leq t, \quad i = 1, \ldots, m \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax - b \leq t\mathbf{1} \\ & -(Ax - b) \leq t\mathbf{1} \end{array}$$

∞-norm regression

$$\text{minimize} \quad \|Ax - b\|_\infty$$

The ∞**-norm** of m-vector y is

$$\|y\|_\infty = \max_{i=1,\ldots,m} |y_i| = \max_{i=1,\ldots,m} \max\{y_i, -y_i\}$$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax - b \leq t\mathbf{1} \\ & -(Ax - b) \leq t\mathbf{1} \end{array}$$

∞-norm regression

$$\text{minimize} \quad \|Ax - b\|_\infty$$

The ∞**-norm** of m-vector y is

$$\|y\|_\infty = \max_{i=1,\ldots,m} |y_i| = \max_{i=1,\ldots,m} \max\{y_i, -y_i\}$$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax - b \leq t\mathbf{1} \\ & -(Ax - b) \leq t\mathbf{1} \end{array}$$

Matrix notation

$$\begin{array}{ll} \text{minimize} & \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} x \\ t \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} A & -\mathbf{1} \\ -A & -\mathbf{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{array}$$

Sum of piecewise-linear functions

minimize $\quad f(x) + g(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i) + \max_{i=1,\ldots,p} (c_i^T x + d_i)$

Sum of piecewise-linear functions

minimize $\quad f(x) + g(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i) + \max_{i=1,\ldots,p} (c_i^T x + d_i)$

Equivalent linear optimization

$$
\begin{array}{ll}
\text{minimize} & t_1 + t_2 \\
\text{subject to} & a_i^T x + b_i \leq t_1, \quad i = 1, \ldots, m \\
 & c_i^T x + d_i \leq t_2, \quad i = 1, \ldots, p
\end{array}
$$

1-norm regression

$$\text{minimize} \quad \|Ax - b\|_1$$

The 1-**norm** of m-vector y is

$$\|y\|_1 = \sum_{i=1}^{m} |y_i| = \sum_{i=1}^{m} \max\{y_i, -y_i\}$$

1-norm regression

$y_i = (Ax - b)_i$

$$\text{minimize} \quad \|Ax - b\|_1$$

The 1**-norm** of m-vector y is

$$\|y\|_1 = \sum_{i=1}^{m} |y_i| = \sum_{i=1}^{m} \max\{y_i, -y_i\}$$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} u_i \\ \text{subject to} & (Ax - b)_i \leq u_i, \quad i = 1, \dots, m \\ & -(Ax - b)_i \leq u_i, \quad i = 1, \dots, m \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & Ax - b \leq u \\ & -(Ax - b) \leq u \end{array}$$

1-norm regression

$$\text{minimize} \quad \|Ax - b\|_1$$

The 1**-norm** of m-vector y is

$$\|y\|_1 = \sum_{i=1}^{m} |y_i| = \sum_{i=1}^{m} \max\{y_i, -y_i\}$$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & Ax - b \leq u \\ & -(Ax - b) \leq u \end{array}$$

1-norm regression

$$\text{minimize} \quad \|Ax - b\|_1$$

The 1-**norm** of m-vector y is

$$\|y\|_1 = \sum_{i=1}^{m} |y_i| = \sum_{i=1}^{m} \max\{y_i, -y_i\}$$

Equivalent problem

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & Ax - b \leq u \\ & -(Ax - b) \leq u \end{array}$$

Matrix notation

$$\begin{array}{ll} \text{minimize} & \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}^T \begin{bmatrix} x \\ u \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{array}$$

Summary: 1 and ∞-norm regression

∞-norm

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

Equivalent to

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & Ax - b \leq t\mathbf{1} \\
& -(Ax - b) \leq t\mathbf{1}
\end{array}
$$

Absolute value of every element $(Ax - b)_i$ is bounded by the same **scalar** t

Summary: 1 and ∞-norm regression

∞-norm

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

Equivalent to

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & Ax - b \le t\mathbf{1} \\
& -(Ax - b) \le t\mathbf{1}
\end{array}
$$

Absolute value of every element $(Ax - b)_i$ is bounded by the same **scalar** t

1-norm

minimize $\quad \|Ax - b\|_1$

Equivalent to

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T u \\
\text{subject to} & Ax - b \le u \\
& -(Ax - b) \le u
\end{array}
$$

Absolute value of every element $(Ax - b)_i$ is bounded by a component of the **vector** u

Example : converting to an LP

minimize $\|Ax - b\|_\infty$

subject to $\|x\|_1 \leq k$

$$\begin{aligned} \min \quad & t \\ \text{st} \quad & \|x\|_1 \leq k \\ & Ax - b \leq t\mathbf{1} \\ & -(Ax-b) \leq t\mathbf{1} \end{aligned}$$

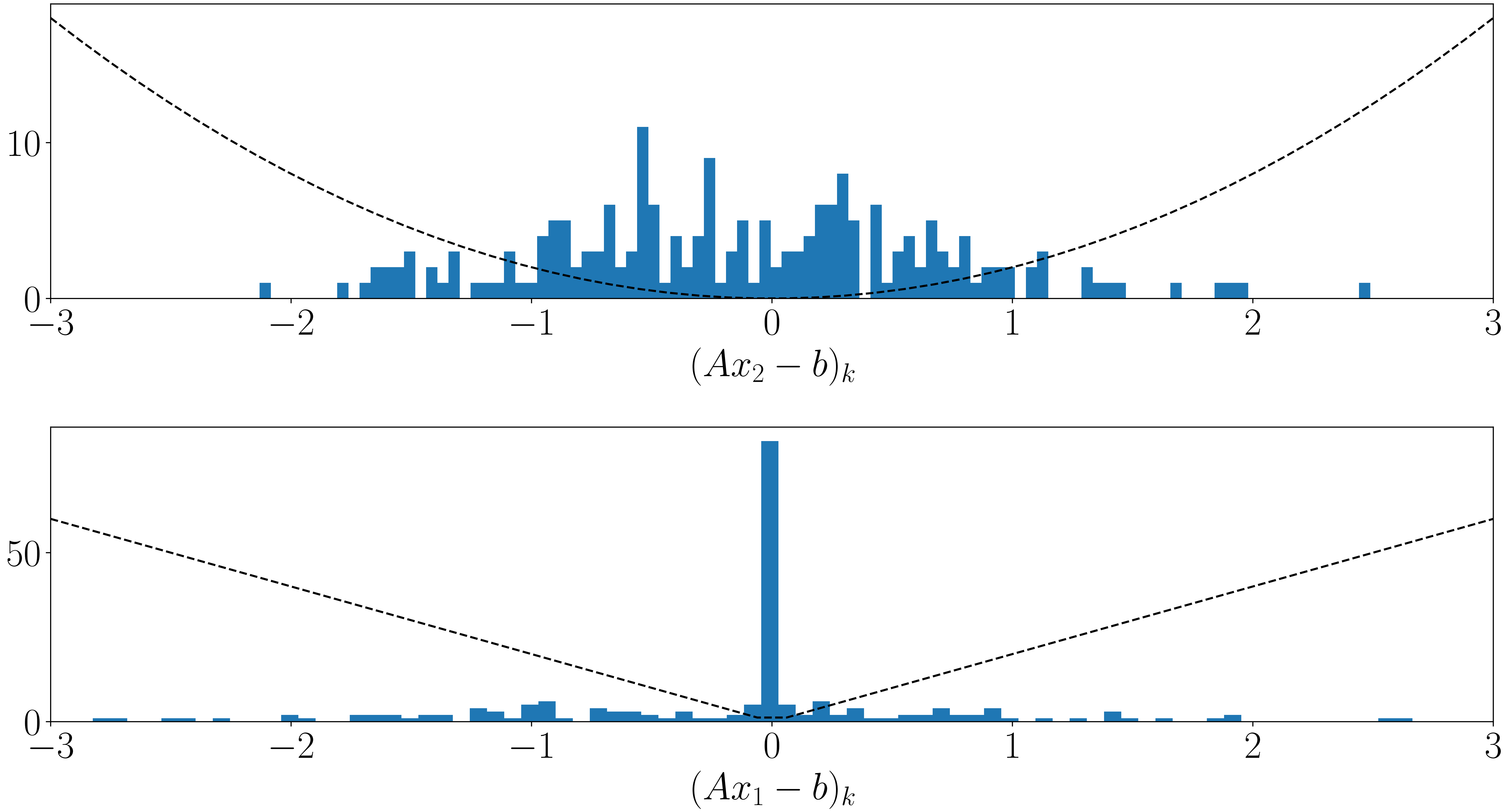
$$\begin{aligned} \min \quad & t \\ \text{st.} \quad & \mathbf{1}^T u \leq k \\ & Ax - b \leq t\mathbf{1} \\ & -(Ax-b) \leq t\mathbf{1} \\ & x \leq u \\ & -x \leq u \end{aligned}$$

Comparison with least-squares

$r = Ax - b$

Histogram of residuals $Ax - b$ with randomly generated $A \in \mathbf{R}^{200\times 80}$

$$x_2 = \text{argmin } \|Ax - b\|_2^2, \qquad x_1 = \text{argmin } \|Ax - b\|_1$$



1-norm distribution is **wider** with a **high peak at zero**

Modeling software does most of this for you

∞-norm

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

```
import numpy as np
import cvxpy as cp

m = 200; n = 80

A = np.random.randn(200, 80)
b = np.random.randn(200)
x = cp.Variable(80)

objective = cp.norm(A @ x - b, np.inf)
problem = cp.Problem(cp.Minimize(objective))
problem.solve()
```

Modeling software does most of this for you

∞-norm

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

```
import numpy as np
import cvxpy as cp

m = 200; n = 80

A = np.random.randn(200, 80)
b = np.random.randn(200)
x = cp.Variable(80)

objective = cp.norm(A @ x - b, np.inf)
problem = cp.Problem(cp.Minimize(objective))
problem.solve()
```

1-norm

minimize $\quad ||Ax - b||_1$

```
import numpy as np
import cvxpy as cp

m = 200; n = 80

A = np.random.randn(200, 80)
b = np.random.randn(200)
x = cp.Variable(80)

objective = cp.norm(A @ x - b, 1)
problem = cp.Problem(cp.Minimize(objective))
problem.solve()
```

Sparse signal recovery

Sparse signal recovery via 1–norm minimization

$\hat{x} \in \mathbf{R}^n$ is unknown signal, known to be sparse

We make linear measurements $y = A\hat{x}$ with $A \in \mathbf{R}^{m \times n}, m < n$

Estimate signal with smallest ℓ_1-norm, consistent with measurements

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = y \end{array}$$

Sparse signal recovery via 1–norm minimization

$\hat{x} \in \mathbf{R}^n$ is unknown signal, known to be sparse

We make linear measurements $y = A\hat{x}$ with $A \in \mathbf{R}^{m \times n}, m < n$

Estimate signal with smallest ℓ_1-norm, consistent with measurements

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = y \end{array}$$

Equivalent linear optimization

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & -u \leq x \leq u \\ & Ax = y \end{array}$$

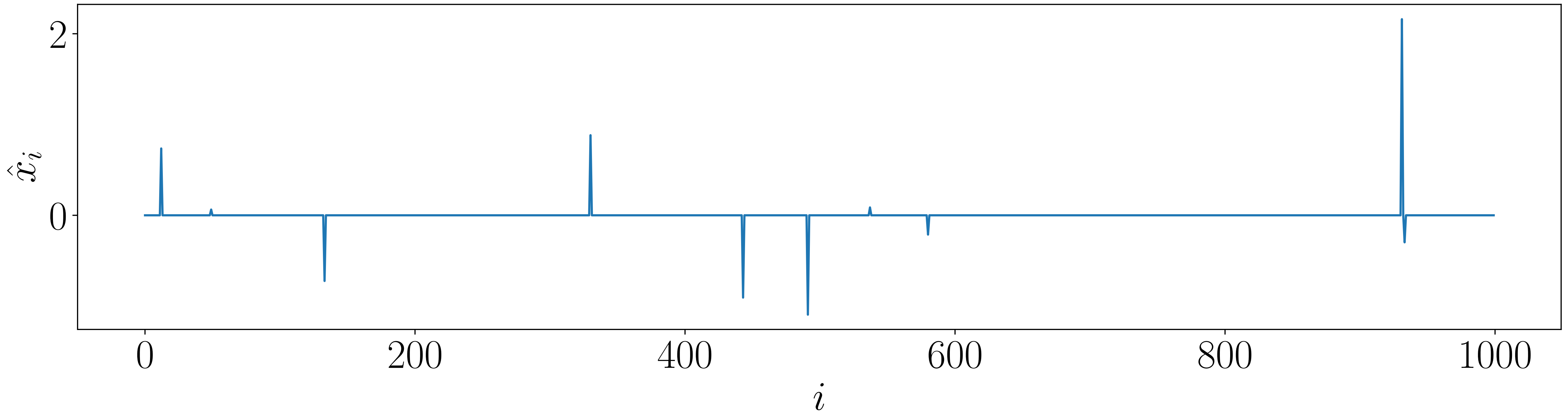
Sparse signal recovery via 1–norm minimization

Example

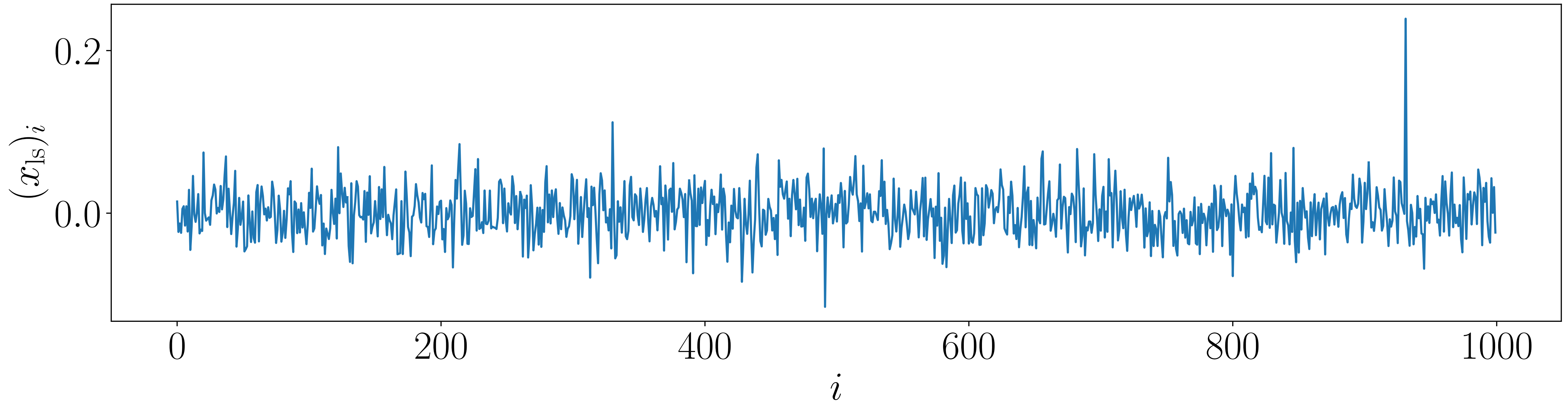
Exact signal $\hat{x} \in \mathbf{R}^{1000}$

10 nonzero components

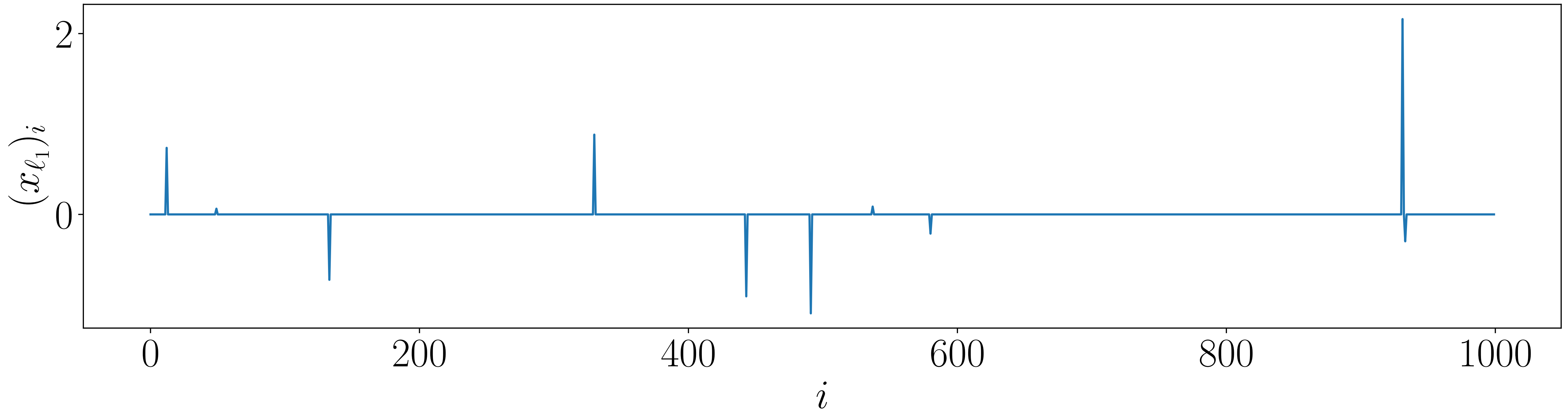
Random $A \in \mathbf{R}^{100 \times 1000}$



The least squares estimate cannot recover the sparse signal



The 1-norm estimate is **exact**



Support vector machines

Linear classification

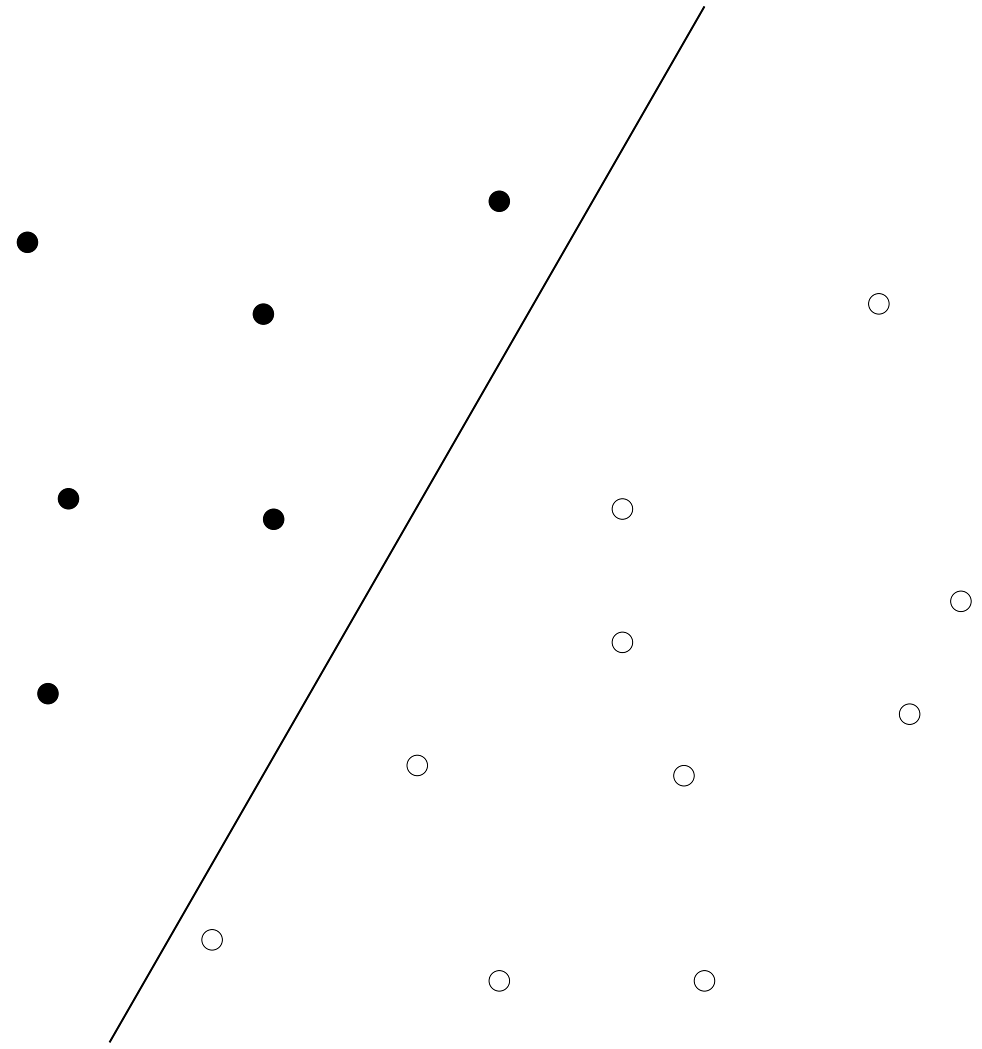
Support vector machine (linear separation)

Given a set of points $\{v_1, \dots, v_N\}$ with binary labels $s_i \in \{-1, 1\}$

Find hyperplane that strictly separates the tho classes

×10, ×1000

$$a^T v_i + b > 0 \quad \text{if} \quad s_i = 1$$
$$a^T v_i + b < 0 \quad \text{if} \quad s_i = -1$$



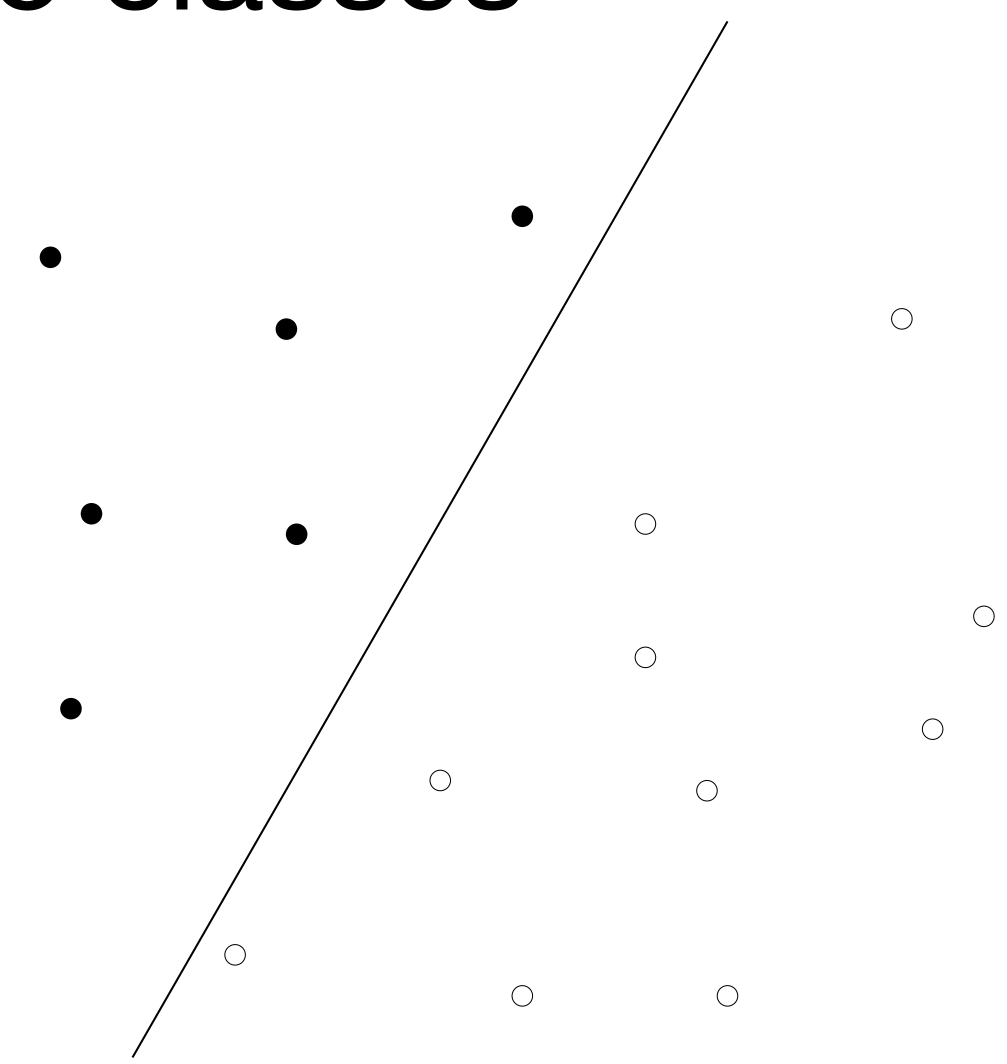
Linear classification

Support vector machine (linear separation)

Given a set of points $\{v_1, \ldots, v_N\}$ with binary labels $s_i \in \{-1, 1\}$

Find hyperplane that strictly separates the tho classes

$$a^T v_i + b > 0 \quad \text{if} \quad s_i = 1$$
$$a^T v_i + b < 0 \quad \text{if} \quad s_i = -1$$



Homogeneous in (a, b), hence equivalent to the linear inequalities (in a, b)

$$s_i(a^T v_i + b) \geq 1$$

Linear classification

Separable case

Feasibility problem

$$
\begin{array}{ll}
\text{find} & a, b \\
\text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \dots, N
\end{array}
$$

Linear classification

Separable case

Feasibility problem

$$\begin{array}{ll} \text{find} & a, b \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \ldots, N \end{array}$$

Which can be seen as a special case of LP with

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \ldots, N \end{array}$$

Linear classification

Separable case

Feasibility problem

$$\begin{array}{ll} \text{find} & a, b \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \dots, N \end{array}$$

Which can be seen as a special case of LP with

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \dots, N \end{array}$$

$p^\star = 0$ if problem feasible (points separable)

$p^\star = \infty$ if problem infeasible (points not separable)

Linear classification

Separable case

Feasibility problem

$$\begin{array}{ll} \text{find} & a, b \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \dots, N \end{array}$$

Which can be seen as a special case of LP with

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & s_i(a^T v_i + b) \geq 1, \quad i = 1, \dots, N \end{array}$$

$p^\star = 0$ if problem feasible (points separable)

$p^\star = \infty$ if problem infeasible (points not separable) ⟶ **What then?**

Linear classification

Approximate linear separation of non-separable points

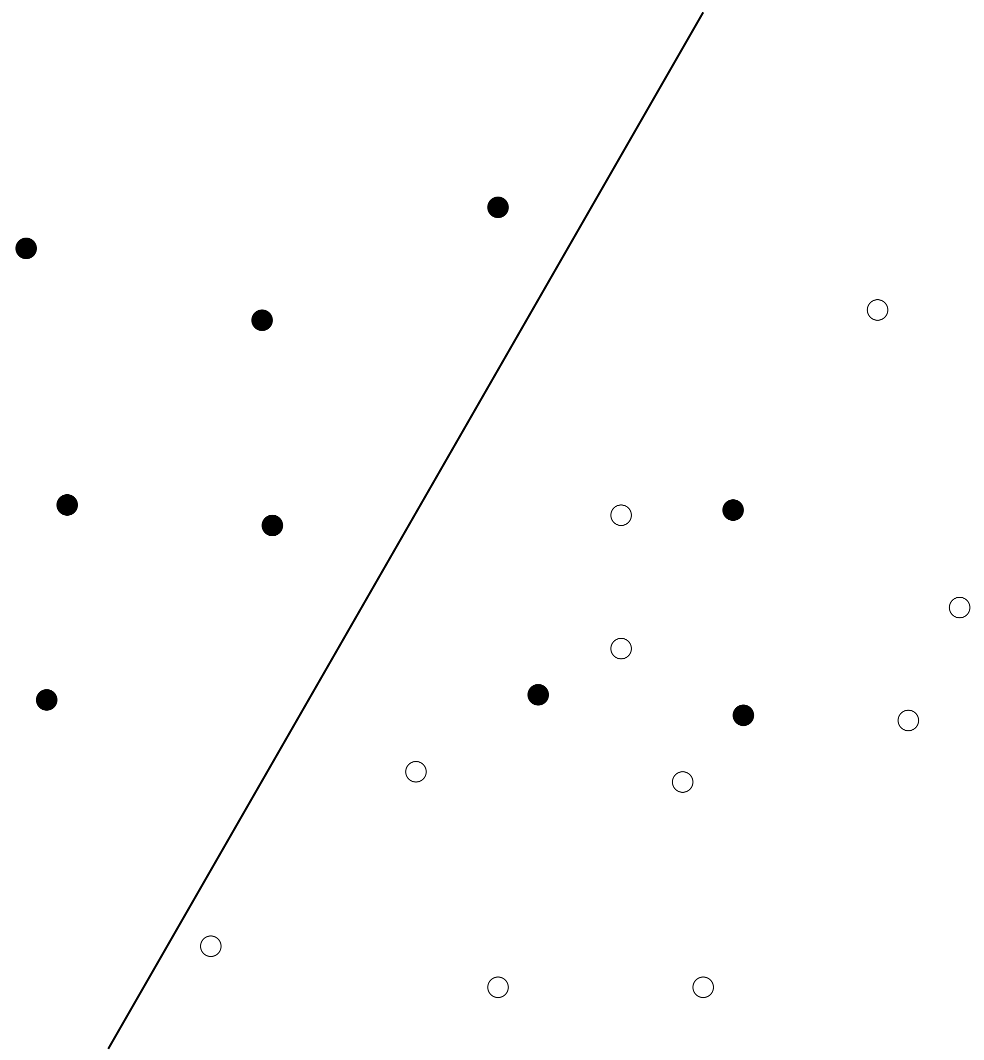
Each of our constraints is

$$s_i(a^T v_i + b) \geq 1$$

→

Violation

$$\max\{0, 1 - s_i(a^T v_i + b)\}$$



Linear classification

Approximate linear separation of non-separable points

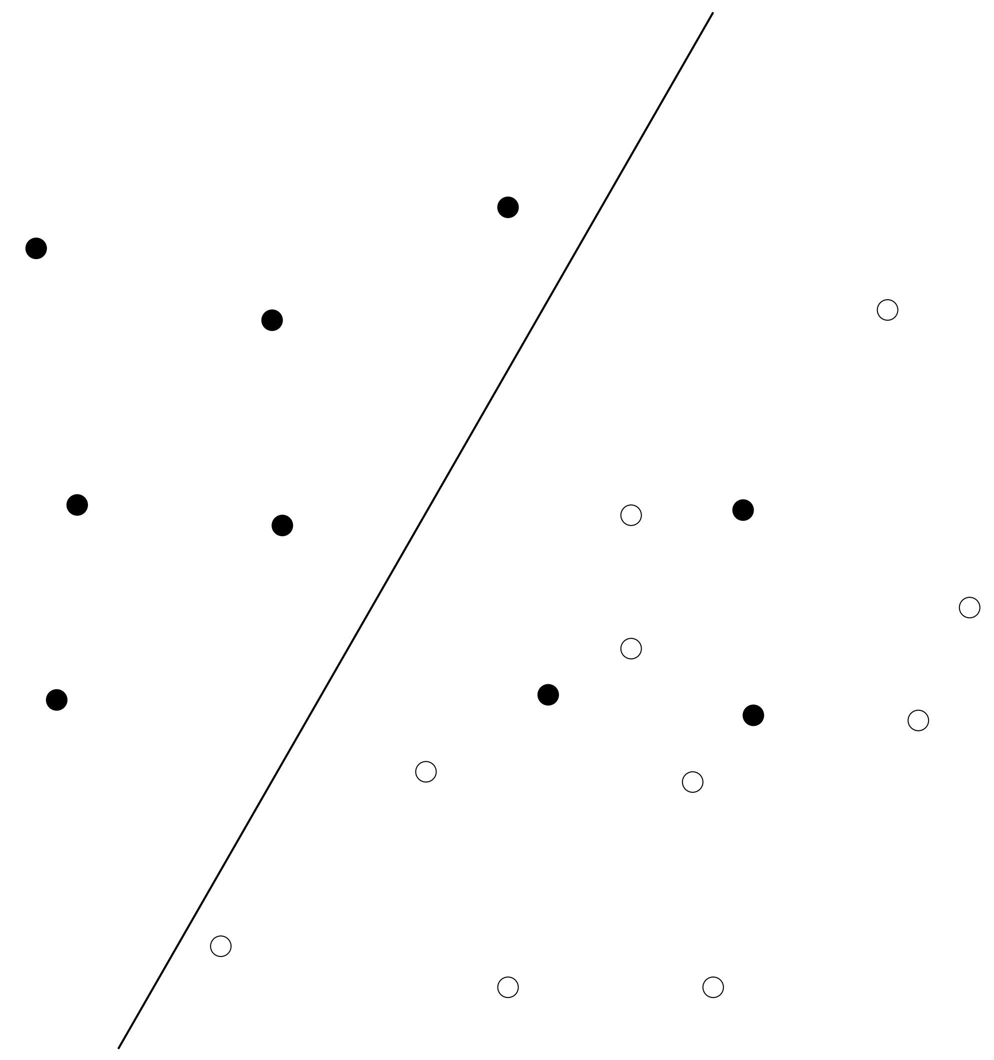
Each of our constraints is

$$s_i(a^T v_i + b) \geq 1$$

↳ $1 - s_i(a^T v_i + b) \leq 0$

Violation

$$\max\{0, 1 - s_i(a^T v_i + b)\}$$



Goal

Minimize sum of the violations

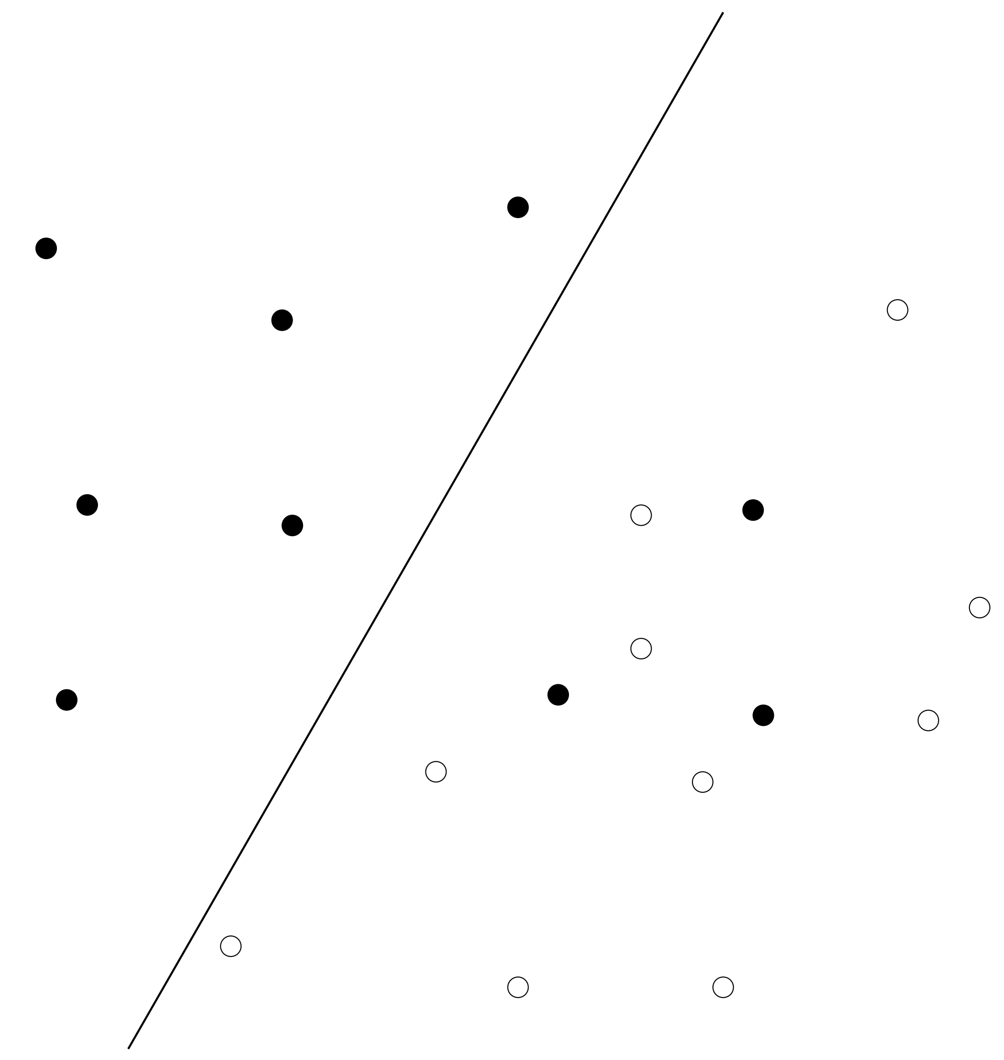
minimize $\quad \sum_{i=1}^{N} \max\{0, 1 - s_i(a^T v_i + b)\}$

Piecewise-linear minimization problem with variables a, b

Linear classification

Approximate linear separation of non-separable points

minimize $\sum_{i=1}^{N} \max\{0, 1 - s_i(a^T v_i + b)\}$

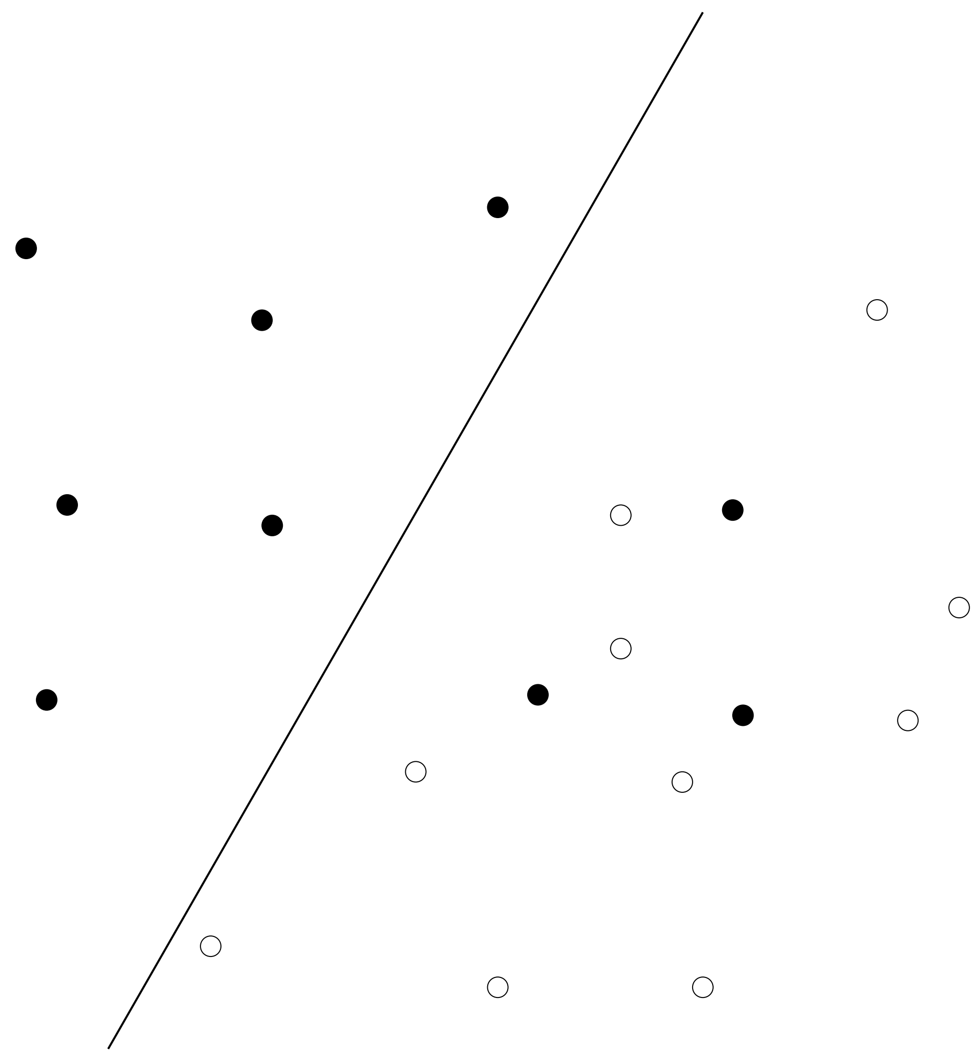


Linear classification

Approximate linear separation of non-separable points

minimize $\quad \sum_{i=1}^{N} \max\{0, 1 - s_i(a^T v_i + b)\}$

$u \in \mathbb{R}^N$



As a linear optimization problem

$$\min \quad \mathbf{1}^T u$$

$$1 - s_i(a^T v_i + b) \leq u_i \quad i = 1, \dots, N$$

$$0 \leq u_i \quad i = 1, \dots, N$$

$$\tilde{x} = (a, b, u), \quad a \in \mathbb{R}^n, \; b \in \mathbb{R}, \; u \in \mathbb{R}^N$$

AS EXERCISE

$$\min \quad \tilde{c}^T \tilde{x}$$

$$\text{s.t.} \quad \tilde{A} \tilde{x} \leq \tilde{b}$$

Piecewise-linear optimization

Today, we learned to:

- **Understand** the differences between vector norms
- **Reformulate** convex piecewise linear minimization as linear optimization
- **Apply** these techniques to sparse signal recovery and classification problems

References

- Bertsimas, Tsitsiklis: Introduction to Linear Optimization
  - Chapter 1.3: piecewise linear optimization

- R. Vanderbei: Linear Programming — Foundations and Extensions
  - Chapter 12.4,12.7: 1-norm regression and SVMs

Next time

- Linear optimization geometry
- Optimality conditions