ORF307 – Optimization

6. Constrained least squares

Bartolomeo Stellato — Spring 2024

Ed Forum

CONTROL $y = Ax + b$ GOAL $\min \|y - y^{ref}\|^2 + \lambda \|x\|^2$

ESTIMATION $y = Ax + v$ GOAL $\min_x \|Ax - y\|^2 + \lambda \|x\|^2$

- What is the difference between control and estimation?
- What is the definition of pareto frontier? Why can't you do better?

Recap

Multi-objective least squares

Goal choose n-vector x such that
k norm squared objectives are small

$$J_1 = \|A_1 x - b_1\|^2$$

$$\vdots$$

$$J_k = \|A_k x - b_k\|^2$$

A_i are $m_i \times n$ matrices and b_i are m_i-vectors for $i = 1, \ldots, k$

J_i are the objectives in a *multi-objective (-criterion) optimization problem*

Could choose x to minimize
any one J_i, but we want
tie make them all small

Weighted sum objective

Choose positive weights $\lambda_1, \ldots, \lambda_k$ and form *weighted sum objective*

$$\begin{aligned} J &= \lambda_1 J_1 + \cdots + \lambda_k J_k \\ &= \lambda_1 \|A_1 x - b_1\|^2 + \cdots + \lambda_k \|A_k x - b_k\|^2 \end{aligned}$$

Choose x to minimize J

Primary objective

- Often $\lambda_1 = 1$ and J_1 is the **primary objective**
- **Interpretation** λ_i is how much we care about J_i being small, relative to J_1

Bi-criterion optimization

$$J_1 + \lambda J_2 = \|A_1 x - b_1\|^2 + \lambda \|A_2 x - b_2\|^2$$

Optimal trade-off curve

Bi-criterion problem

$$\text{minimize} \quad J_1(x) + \lambda J_2(x) \quad \longrightarrow \quad x^\star(\lambda)$$

Pareto optimal $x^\star(\lambda)$

There is no point z that satisfies

$$J_1(z) \leq J_1(x^\star(\lambda)) \quad \text{and} \quad J_2(z) \leq J_2(x^\star(\lambda))$$

with one of the inequalities holding strictly
(no other point beats $x^\star$ on both objectives)

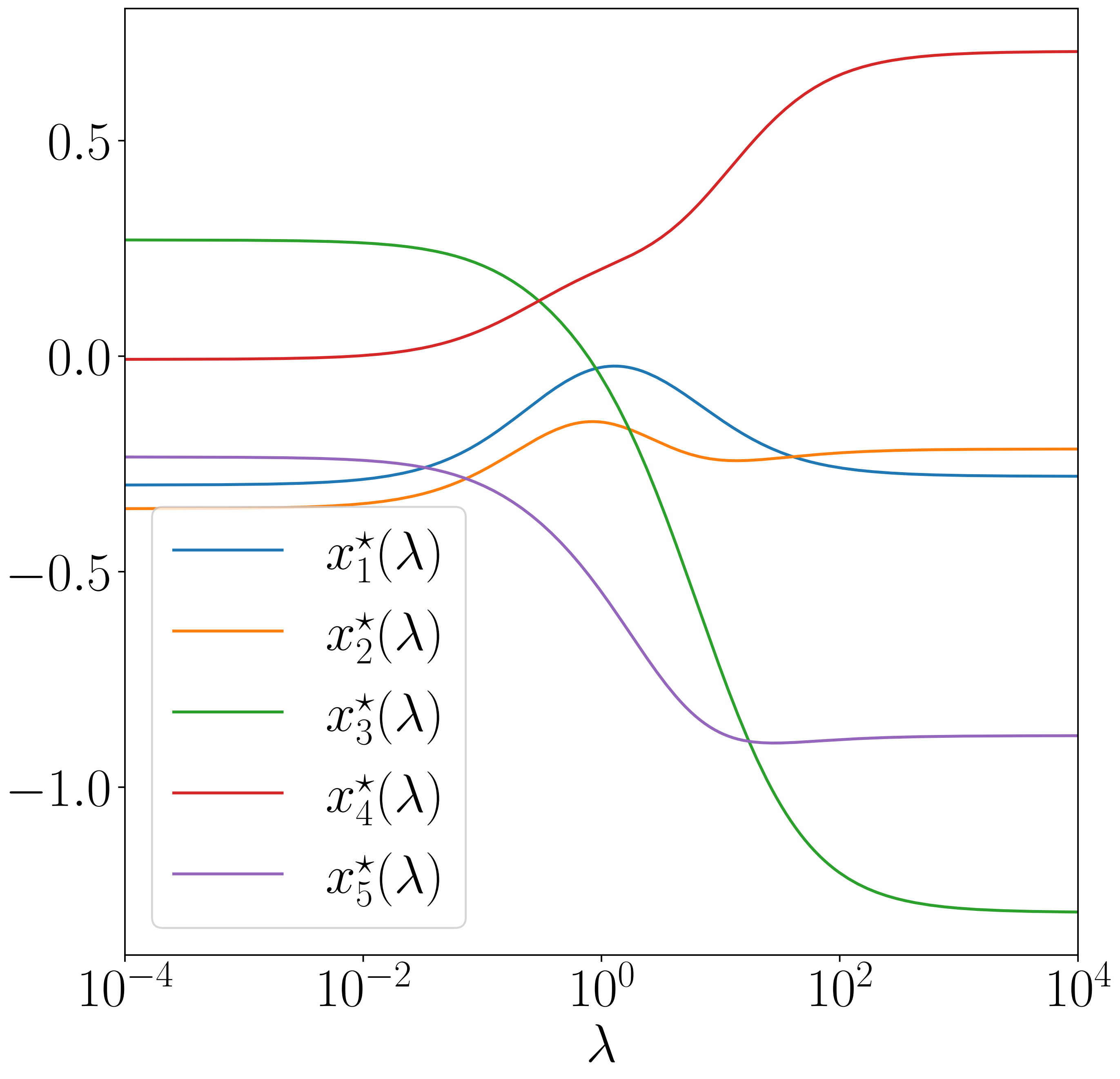
Optimal trade-off curve

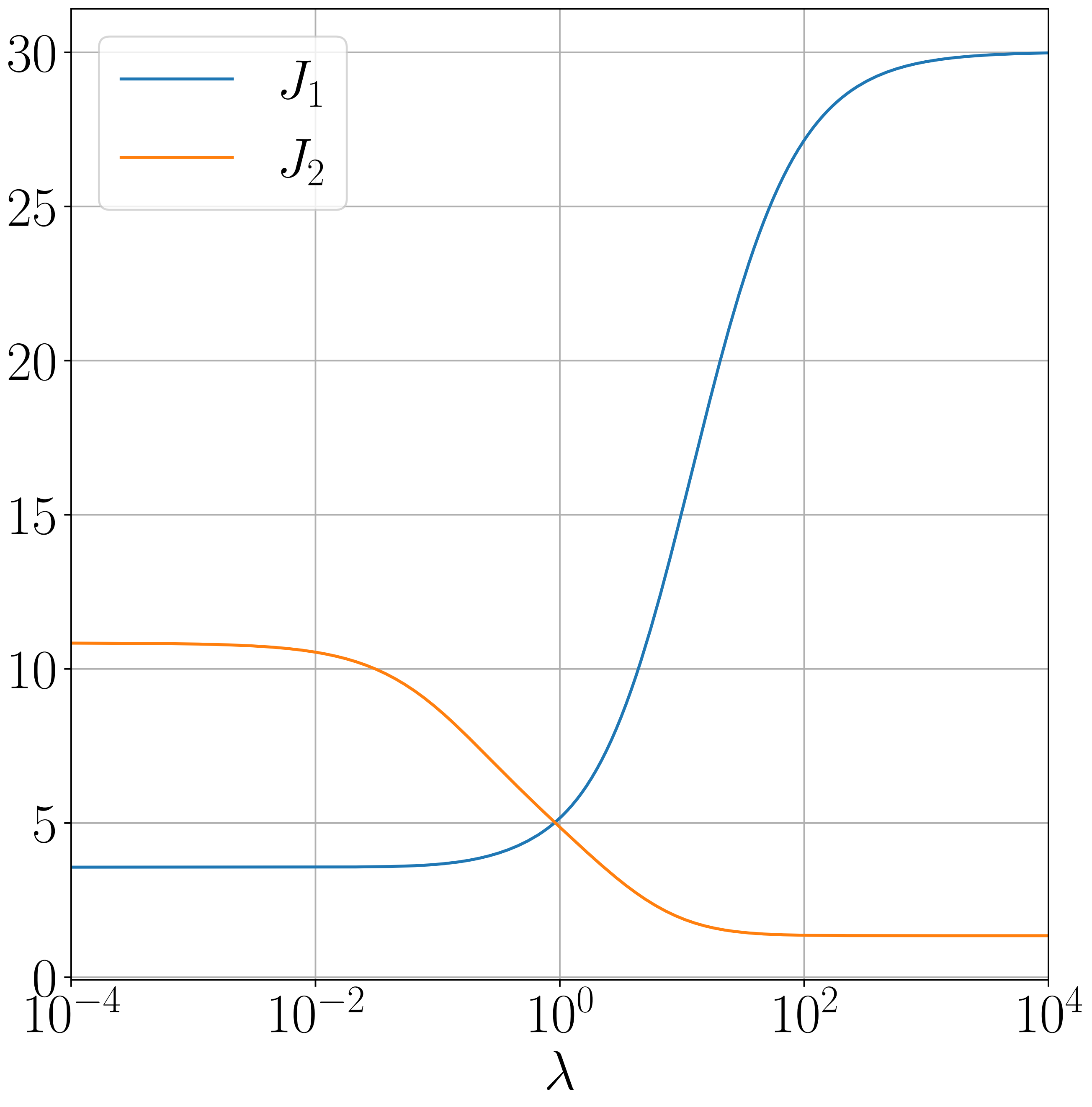
$$(J_1(x^\star(\lambda)), J_2(x^\star(\lambda)), \qquad \lambda > 0$$

Optimal trade-off curve

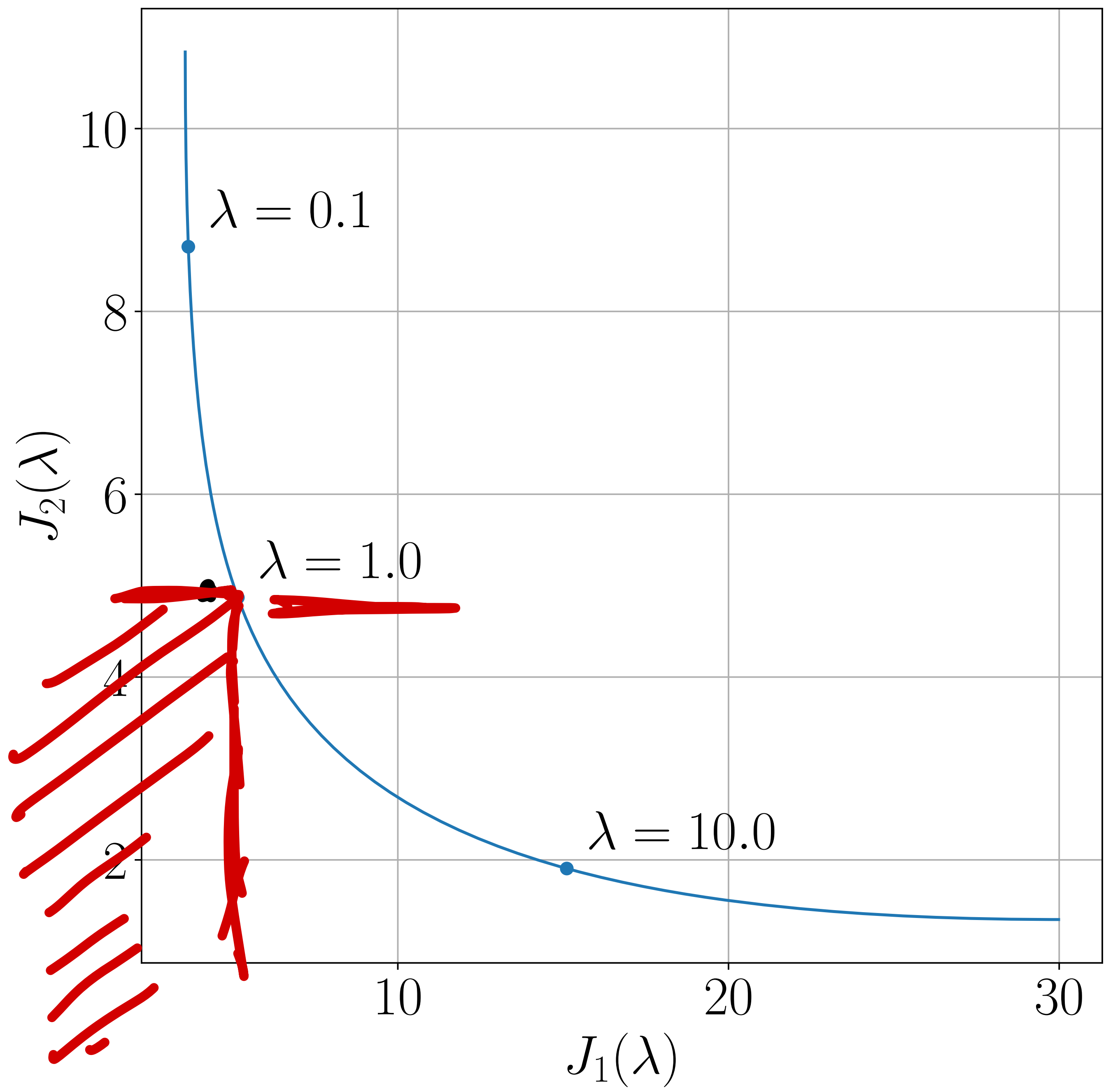
Example

minimize $\quad J_1(x) + \lambda J_2(x)$ $\qquad$ $(A_1, A_2$ are both $10 \times 5)$





Trade-off curve



Today's lecture

Constrained least squares

- Linearly constrained least squares
- Solving the constrained least squares problem
- Portfolio optimization

Linearly constrained least squares

Least squares with equality constraints

The (linearly) constrained least squares problem is

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

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Least squares with equality constraints

The (linearly) constrained least squares problem is

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

objective function

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Least squares with equality constraints

The (linearly) constrained least squares problem is

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

equality constraints

objective function

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Least squares with equality constraints

The (linearly) constrained least squares problem is

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

equality constraints

objective function

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Definitions

x is *feasible* if $Cx = d$

Least squares with equality constraints

The (linearly) constrained least squares problem is

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

objective function

equality constraints

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Definitions

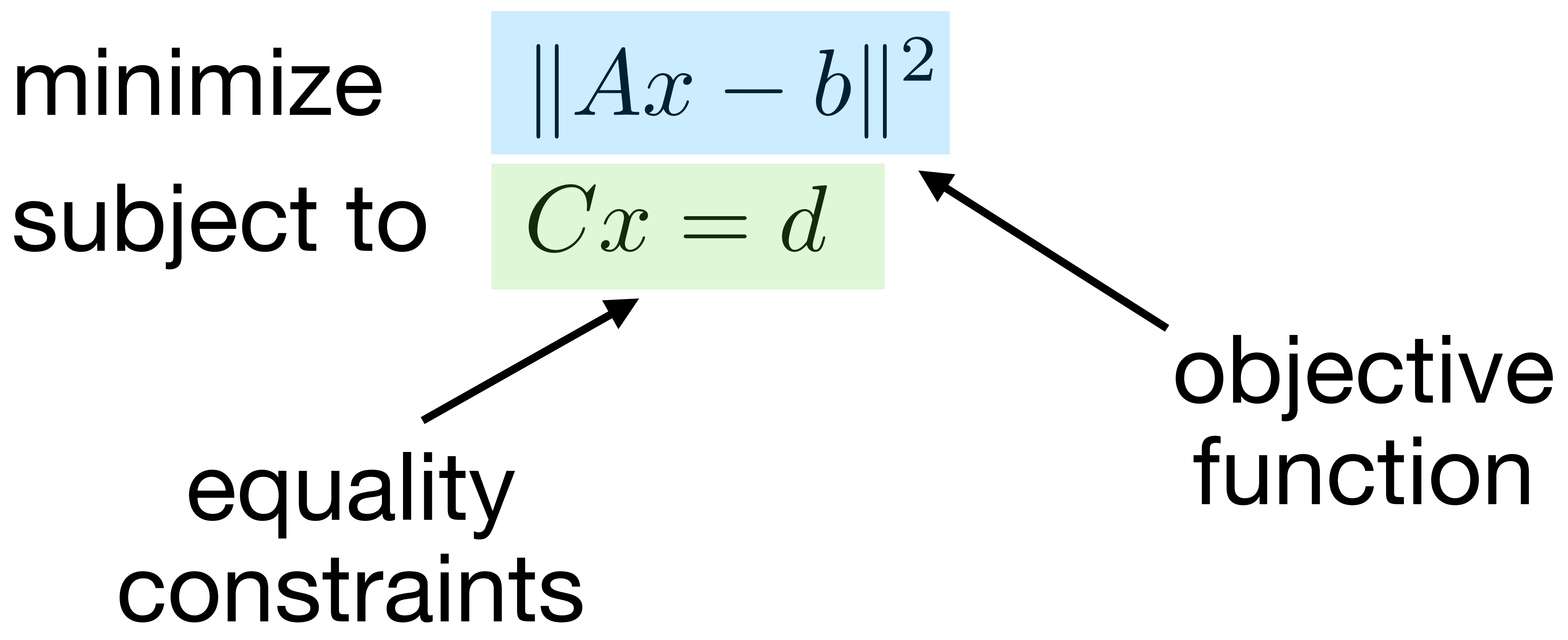
x is *feasible* if $Cx = d$

$x^\star$ is a *solution* if

- $Cx^\star = d$
- $\|Ax^\star - b\|^2 \leq \|Ax - b\|^2$ for any x satisfying $Cx = d$

Least squares with equality constraints

The (linearly) constrained least squares problem is



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

equality constraints

objective function

Problem data

- $m \times n$ matrix A, m-vector b
- $p \times n$ matrix C, p-vector d

Definitions

x is *feasible* if $Cx = d$

$x^\star$ is a *solution* if

- $Cx^\star = d$
- $\|Ax^\star - b\|^2 \leq \|Ax - b\|^2$ for any x satisfying $Cx = d$

Interpretations

- Combine solving linear equations with least squares.
- Like a bi-objective least squares with ∞ weight on second objective, $\|Cx - d\|^2$.

Optimal advertising with budget

m demographic groups
we want to advertise to

→

v^{des} is the m-vector
of desired views/impressions

n advertising channels
(web publishers, radio, print, etc.)

→

s is the n-vector
of purchases

$m \times n$ matrix A gives
demographic reach of channels

→

A_{ij} is the number of views
for group i and dollar spent
on channel j $(1000/\$)$

Optimal advertising with budget

m demographic groups
we want to advertise to

→

v^{des} is the m-vector
of desired views/impressions

n advertising channels
(web publishers, radio, print, etc.)

→

s is the n-vector
of purchases

$m \times n$ matrix A gives
demographic reach of channels

→

A_{ij} is the number of views
for group i and dollar spent
on channel j $(1000/\$)$

Views across
demographic groups

$$v = As$$

Optimal advertising with budget

m demographic groups we want to advertise to → v^{des} is the m-vector of desired views/impressions

n advertising channels (web publishers, radio, print, etc.) → s is the n-vector of purchases

$m \times n$ matrix A gives demographic reach of channels → A_{ij} is the number of views for group i and dollar spent on channel j $(1000/\$)$

Views across demographic groups

$$v = As$$

Goal

$$\begin{array}{ll} \text{minimize} & \|As - v^{\text{des}}\|^2 \\ \text{subject to} & \mathbf{1}^T s = B \end{array}$$

Optimal advertising with budget

m demographic groups we want to advertise to → v^{des} is the m-vector of desired views/impressions

n advertising channels (web publishers, radio, print, etc.) → s is the n-vector of purchases

$m \times n$ matrix A gives demographic reach of channels → A_{ij} is the number of views for group i and dollar spent on channel j $(1000/\$)$

Views across demographic groups

$$v = As$$

Goal

$$\begin{aligned} \text{minimize} \quad & \|As - v^{\text{des}}\|^2 \\ \text{subject to} \quad & \mathbf{1}^T s = B \end{aligned}$$

← allocated budget

Optimal advertising with budget

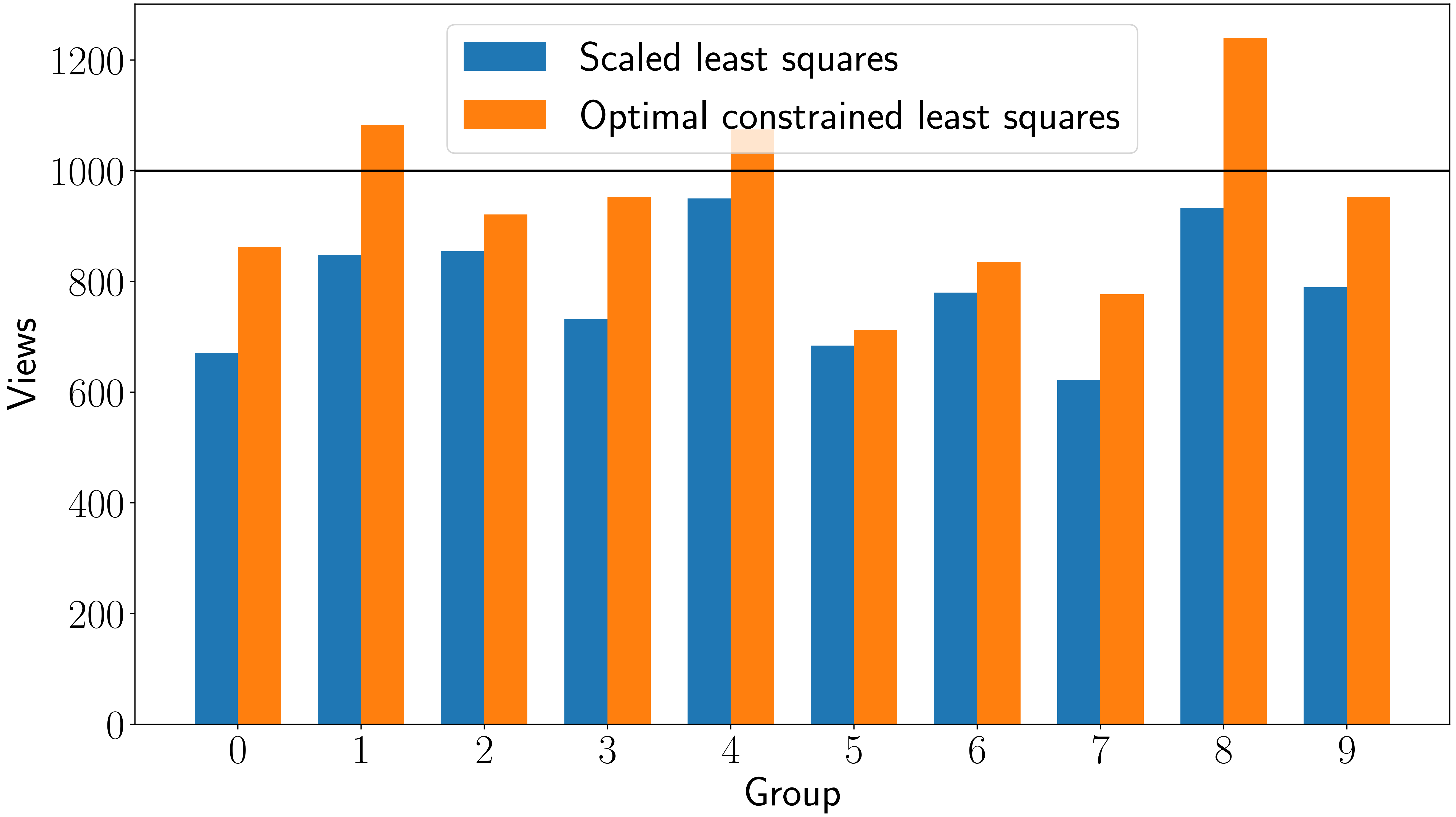
Results

$m = 10$ groups, $n = 3$ channels

budget $B = 1284$

desired views vector $v^{\text{des}} = (10^3)\mathbf{1}$

$$\begin{array}{ll} \text{minimize} & \|As - v^{\text{des}}\|^2 \\ \text{subject to} & \mathbf{1}^T s = B \end{array}$$



Optimal advertising with budget

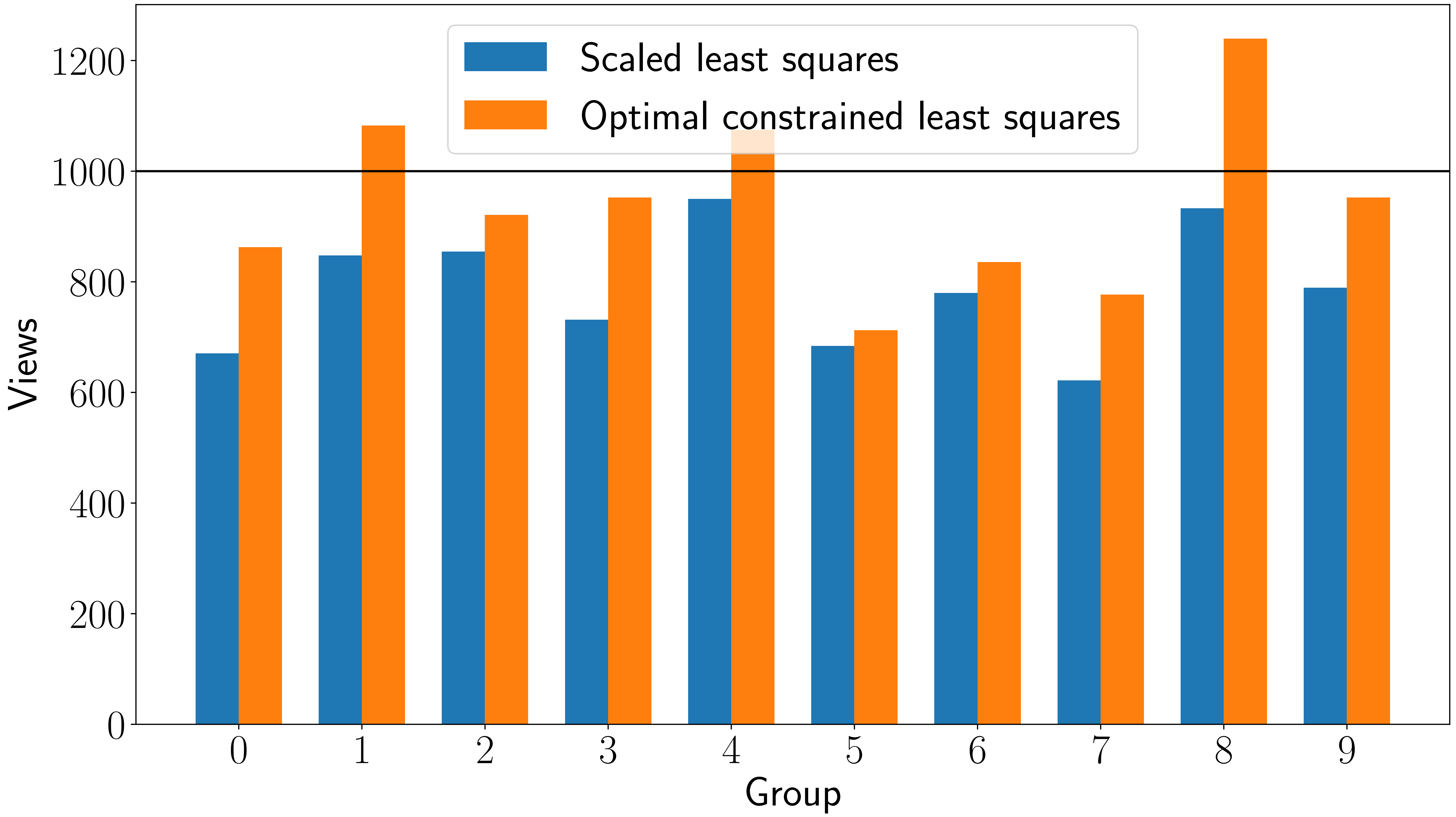
Results

$m = 10$ groups, $n = 3$ channels

budget $B = 1284$

desired views vector $v^{\text{des}} = (10^3)\mathbf{1}$

$$\begin{array}{ll} \text{minimize} & \|As - v^{\text{des}}\|^2 \\ \text{subject to} & \mathbf{1}^T s = B \end{array}$$



optimal spending $s^\star = (315, 110, 859)$ ⟶ RMS 16.10%

Optimal advertising with budget

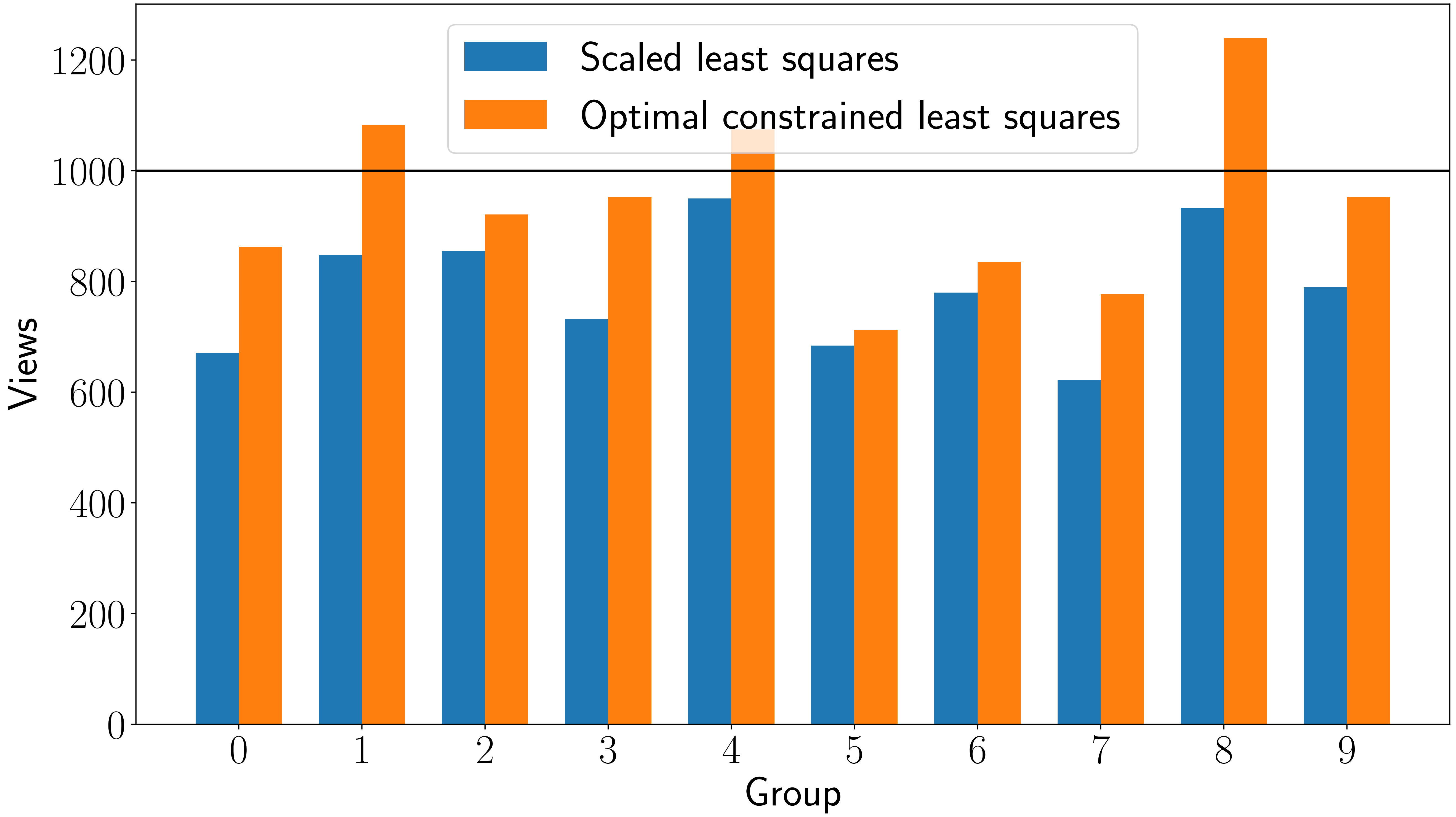
Results

$m = 10$ groups, $n = 3$ channels

budget $B = 1284$

desired views vector $v^{\text{des}} = (10^3)\mathbf{1}$

$$\begin{array}{ll} \text{minimize} & \|As - v^{\text{des}}\|^2 \\ \text{subject to} & \mathbf{1}^T s = B \end{array}$$



optimal spending $s^\star = (315, 110, 859)$ $\longrightarrow$ RMS 16.10%

rescaled least squares spending $s^\star = (50, 80, 1154)$ $\longrightarrow$ RMS 23.85%

Least norm problem

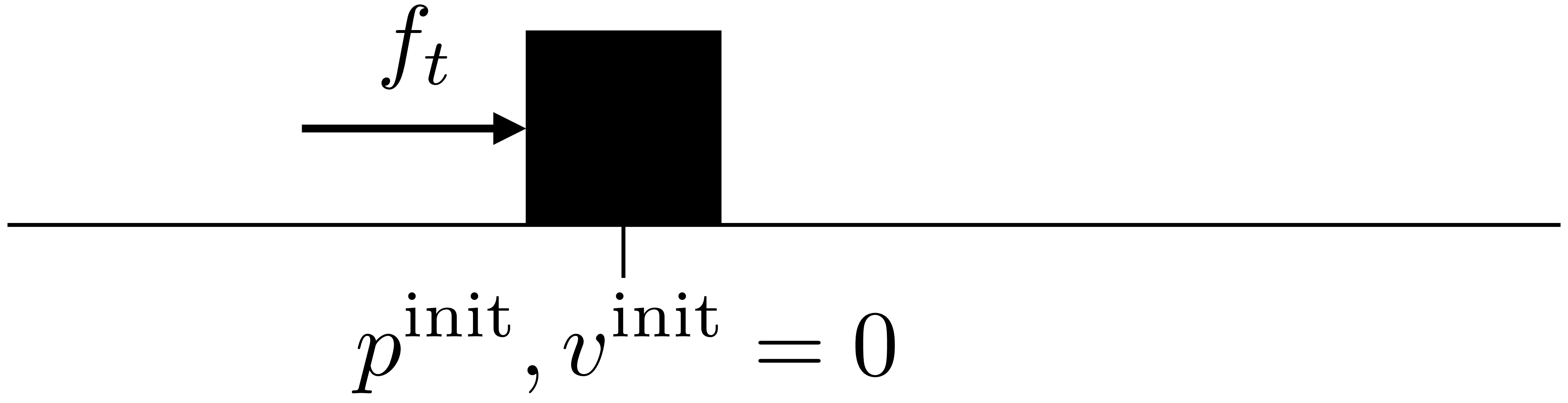
Special case of constrained least squares problem with $A = I$ and $b = 0$

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

Find the smallest vector that satisfies a set of linear equations

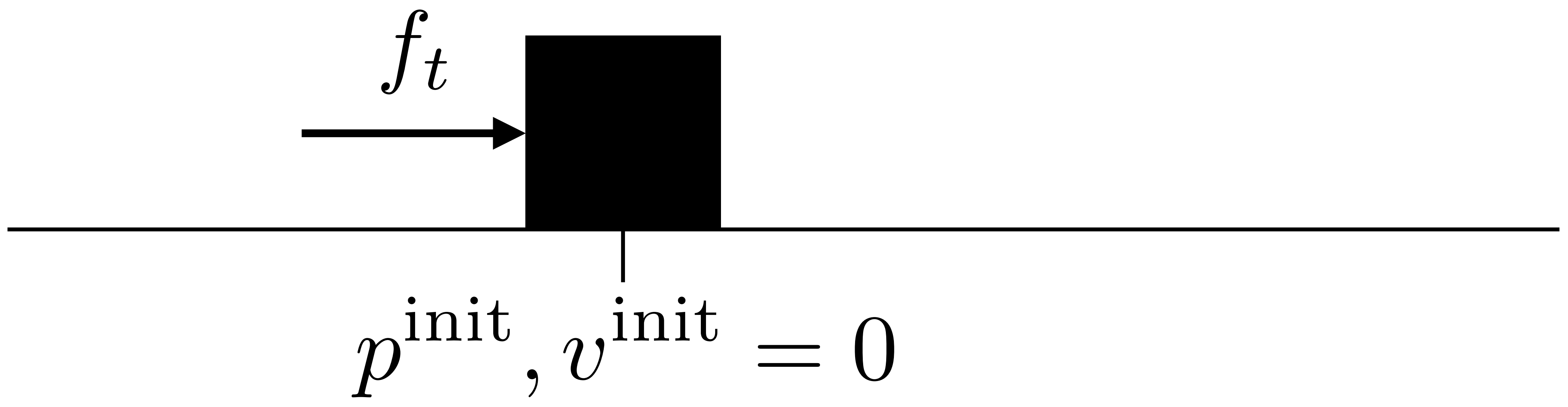
Force sequence

Unit mass on frictionless surface, initially at rest



Force sequence

Unit mass on frictionless surface, initially at rest



$v^{\text{init}} = 0$

$p^{\text{init}} = 0$

$m = 1$

$a = f$

$\Delta t = 1$

EXERCISE

DERIVE EQUALITIES

10-vector f gives the forces applied for one second each

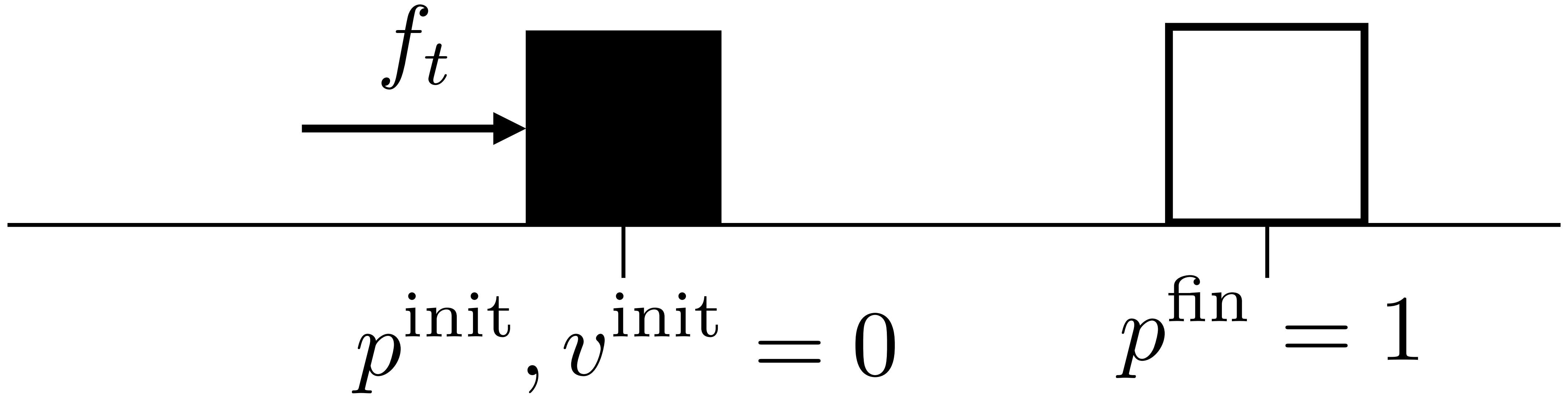
Final velocity and position (Newton's laws)

$$v^{\text{fin}} = f_1 + f_2 + \cdots + f_{10}$$

$$p^{\text{fin}} = (19/2)f_1 + (17/2)f_2 + \cdots + (1/2)f_{10}$$

Force sequence

Unit mass on frictionless surface, initially at rest



10-vector f gives the forces applied for one second each

Final velocity and position (Newton's laws)

$$v^{\text{fin}} = f_1 + f_2 + \cdots + f_{10}$$

$$p^{\text{fin}} = (19/2)f_1 + (17/2)f_2 + \cdots + (1/2)f_{10}$$

Goal

Let's find f such that $v^{\text{fin}} = 0$ and $p^{\text{fin}} = 1$

Least norm force sequence

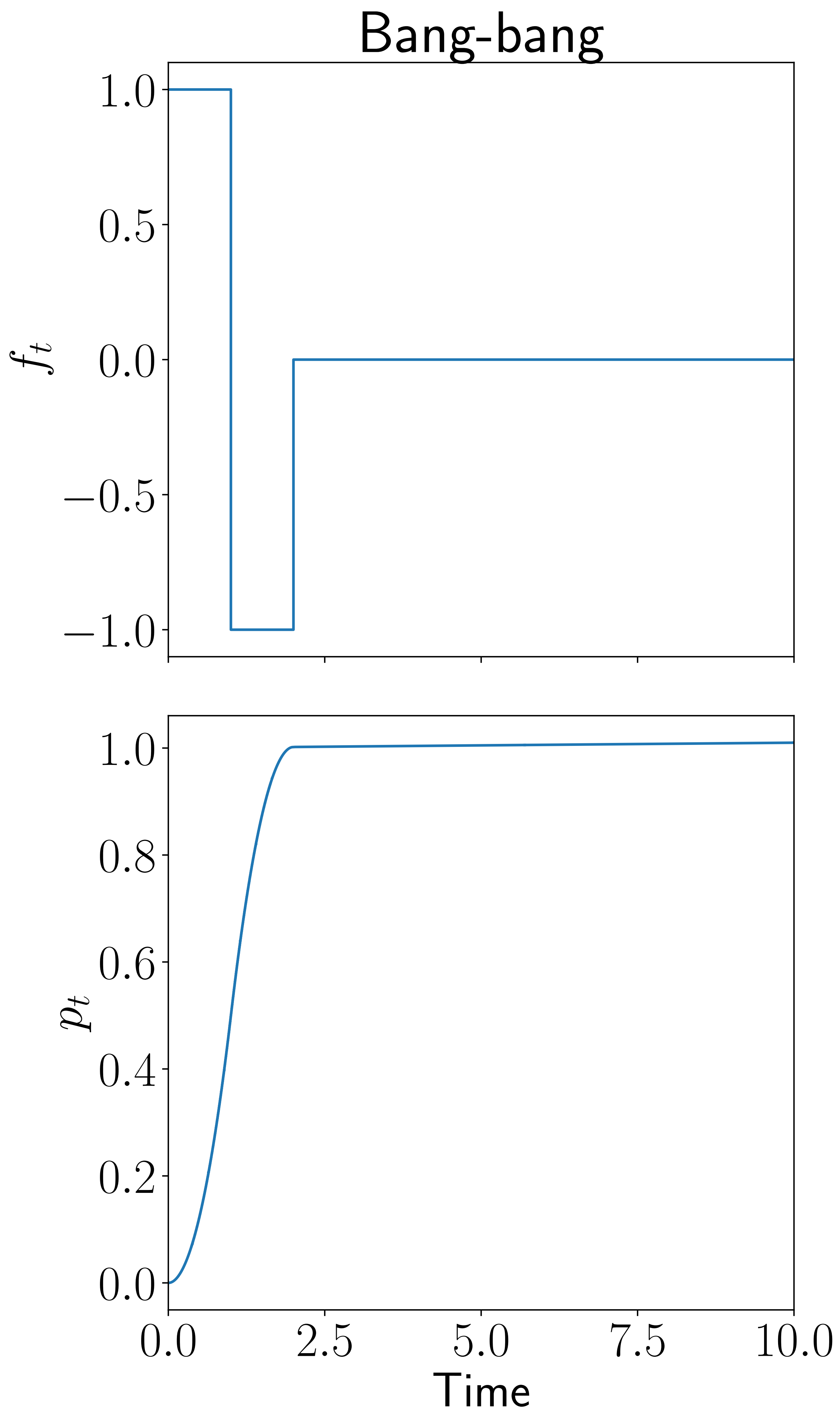
Find f that brings to $p^{\text{fin}} = 1$, $v^{\text{fin}} = 0$

Least norm force sequence

Find f that brings to $p^{\text{fin}} = 1$, $v^{\text{fin}} = 0$

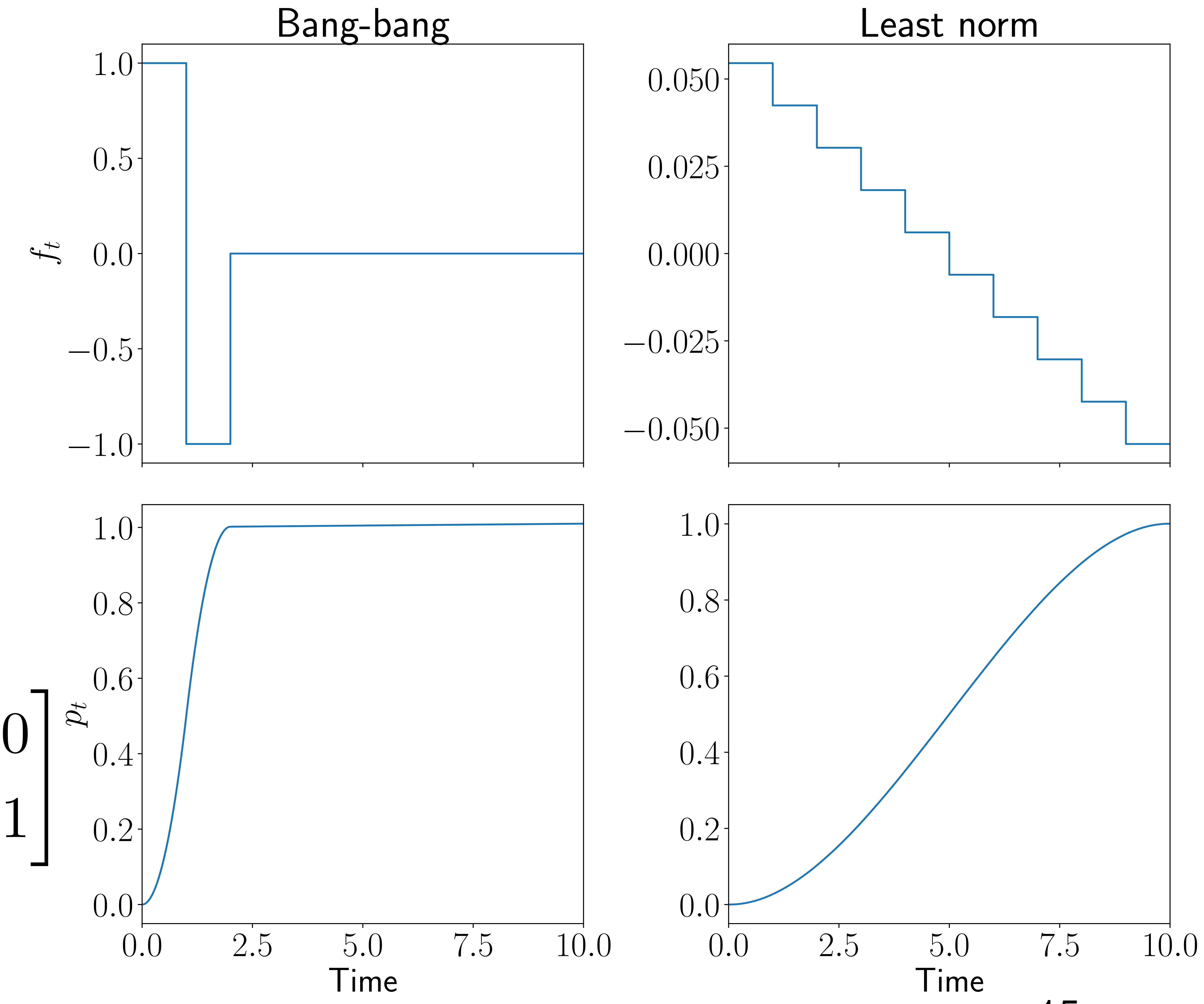
Bang-bang solution

$$f^{\text{bb}} = (1, -1, 0, \ldots, 0) \qquad \|f^{\text{bb}}\|^2 = 2$$



Least norm force sequence

Find f that brings to $p^{\text{fin}} = 1$, $v^{\text{fin}} = 0$



Bang-bang solution

$$f^{\text{bb}} = (1, -1, 0, \ldots, 0) \qquad \|f^{\text{bb}}\|^2 = 2$$

Least norm solution

minimize $\quad \|f\|^2$

subject to $\quad \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 19/2 & 17/2 & \ldots & 1/2 \end{bmatrix} f = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\|f^{\text{ln}}\|^2 = 0.012$ **Much cheaper effort!**

Solving the constrained least squares problem

Optimality conditions via calculus

minimize $\quad f(x) = \|Ax - b\|^2$

subject to $\quad Cx = d$

Optimality conditions via calculus

$$\begin{array}{ll} \text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & Cx = d \end{array} \quad \longrightarrow \quad \begin{array}{ll} \text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & c_i^T x = d_i, \quad i = 1, \ldots, p \end{array}$$

Optimality conditions via calculus

$$\begin{array}{ll}\text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & Cx = d\end{array} \longrightarrow \begin{array}{ll}\text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & c_i^T x = d_i, \quad i = 1, \ldots, p\end{array}$$

Lagrangian function

$$L(x, z) = f(x) + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions via calculus

$$\begin{array}{ll}\text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & Cx = d\end{array} \quad \longrightarrow \quad \begin{array}{ll}\text{minimize} & f(x) = \|Ax - b\|^2 \\ \text{subject to} & c_i^T x = d_i, \quad i = 1, \ldots, p\end{array}$$

Lagrangian function

$$L(x, z) = f(x) + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions

$$\left[\begin{aligned} &\frac{\partial L}{\partial x_i}(x^\star, z) = 0, \quad i = 1, \ldots, n, \\ &\frac{\partial L}{\partial z_i}(x^\star, z) = 0, \quad i = 1, \ldots, p \end{aligned}\right.$$

Optimality conditions via calculus

$$L(x, z) = x^T A^T A x - 2(A^T b)^T x + b^T b + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions via calculus

$$L(x, z) = x^T A^T A x - 2(A^T b)^T x + b^T b + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions

$$\frac{\partial L}{\partial z_i}(x^\star, z) = c_i^T x^{*} - d_i = 0$$ (we already knew)

Optimality conditions via calculus

$$L(x, z) = x^T A^T A x - 2(A^T b)^T x + b^T b + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions

$$\frac{\partial L}{\partial z_i}(x^\star, z) = c_i^T x - d_i = 0$$

(we already knew)

$$\frac{\partial L}{\partial x_i}(x^\star, z) = \underbrace{2\sum_{j=1}^{n}(A^T A)_{ij} x_j^\star - 2(A^T b)_i}_{2A^T(Ax-b)_i} + \sum_{j=1}^{p} z_j (c_j)_i = 0$$

Optimality conditions via calculus

$$C = \begin{bmatrix} \text{—}\; c_1^T \;\text{—} \\ \text{—}\; c_2^T \;\text{—} \\ \vdots \\ \text{—}\; c_p^T \;\text{—} \end{bmatrix}$$

$$L(x,z) = x^TA^TAx - 2(A^Tb)^Tx + b^Tb + z_1(c_1^Tx - d_1) + \cdots + z_p(c_p^Tx - d_p)$$

Optimality conditions

$$\frac{\partial L}{\partial z_i}(x^\star, z) = c_i^Tx - d_i = 0$$ (we already knew)

$$\frac{\partial L}{\partial x_i}(x^\star, z) = 2\sum_{j=1}^{n}(A^TA)_{ij}x_j^\star - 2(A^Tb)_i + \sum_{j=1}^{p} z_j(c_j)_i = 0$$

$$\longrightarrow$$

Vector form

$$Cx = d$$

$$2A^TAx^\star - 2A^Tb + C^Tz = 0$$

$$\begin{bmatrix} | & | & | & & | \\ c_1 & c_2 & c_3 & \cdots & c_p \\ | & | & | & & | \end{bmatrix}$$

Optimality conditions via calculus

$$L(x, z) = x^T A^T A x - 2(A^T b)^T x + b^T b + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions

$$\frac{\partial L}{\partial z_i}(x^\star, z) = c_i^T x - d_i = 0$$ (we already knew)

$$\frac{\partial L}{\partial x_i}(x^\star, z) = 2\sum_{j=1}^{n}(A^T A)_{ij} x_j^\star - 2(A^T b)_i + \sum_{j=1}^{p} z_j (c_j)_i = 0$$

→

Vector form

$$Cx = d$$

$$2A^T A x^\star - 2A^T b + C^T z = 0$$

Karush-Kuhn-Tucker (KKT) conditions

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

(square set of $n + p$ linear equations)

Optimality conditions via calculus

$$L(x, z) = x^T A^T A x - 2(A^T b)^T x + b^T b + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

Optimality conditions

$$\frac{\partial L}{\partial z_i}(x^\star, z) = c_i^T x - d_i = 0$$

(we already knew)

$$\frac{\partial L}{\partial x_i}(x^\star, z) = 2\sum_{j=1}^{n}(A^T A)_{ij} x_j^\star - 2(A^T b)_i + \sum_{j=1}^{p} z_j (c_j)_i = 0$$

→

Vector form

$$Cx = d$$

$$2A^T A x^\star - 2A^T b + C^T z = 0$$

Karush-Kuhn-Tucker (KKT) conditions

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

(square set of $n + p$ linear equations)

Note KKT equations are extension of normal equations to constrained least squares

Invertibility of KKT matrix

no longer positive definite in general ⟶

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ z \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

Invertibility of KKT matrix

no longer positive definite in general ⟶

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ z \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

The KKT matrix is invertible if and only if

- C has linearly independent rows
- $\begin{bmatrix} A \\ C \end{bmatrix}$ has linearly independent columns (true when A has linearly indep. cols)

Invertibility of KKT matrix

no longer positive definite in general ⟶ $$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ z \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

The KKT matrix is invertible if and only if

- C has linearly independent rows ⟶ $p \leq n$ (C is wide)
- $\begin{bmatrix} A \\ C \end{bmatrix}$ has linearly independent columns ⟶ $m + p \geq n$ $\left(\begin{bmatrix} A \\ C \end{bmatrix} \text{ is tall} \right)$
  (true when A has linearly indep. cols)

Invertibility of KKT matrix

no longer positive definite in general ⟶

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ z \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

The KKT matrix is invertible if and only if

- C has linearly independent rows ⟶ $p \leq n$ (C is wide)
- $\begin{bmatrix} A \\ C \end{bmatrix}$ has linearly independent columns (true when A has linearly indep. cols) ⟶ $m + p \geq n$ $\left(\begin{bmatrix} A \\ C \end{bmatrix} \text{ is tall}\right)$

Complexity (with $p \leq n \leq m$)

- Factor + solve: $2mn^2 + (2/3)(n+p)^3 + 2(n+p)^2 \approx 2mn^2$
- Solve given a new b (prefactored): $2mn + 2(n+p)^2 \approx 2mn$

same as unconstrained

Optimality from KKT solution

For $x^\star$ and $z^\star$ such that

$$2A^TAx^\star + C^Tz^\star = 2A^Tb, \quad Cx^\star = d$$

Optimality from KKT solution

For $x^\star$ and $z^\star$ such that

$$2A^TAx^\star + C^Tz^\star = 2A^Tb, \quad Cx^\star = d$$

Given a feasible x and z, we can write the objective (just as least squares)

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

Optimality from KKT solution

For $x^\star$ and $z^\star$ such that

$$2A^TAx^\star + C^Tz^\star = 2A^Tb, \quad Cx^\star = d$$

Given a feasible x and z, we can write the objective (just as least squares)

$$\|Ax - b\|^2 = \|(Ax - Ax^\star) + (Ax^\star - b)\|^2$$

$$= \|A(x - x^\star)\|^2 + \|Ax^\star - b\|^2 + \overset{=0}{\xcancel{2(x - x^\star)^TA^T(Ax^\star - b)}}$$

We can expand last term, using $2A^T(Ax^\star - b) = -C^Tz^\star$ and $Cx = Cx^\star = d$

$$2(x - x^\star)^TA^T(Ax^\star - b) = -(x - x^\star)^TC^Tz^\star = -(C(x - x^\star))^Tz^\star \boxed{= 0}$$

Optimality from KKT solution

For $x^\star$ and $z^\star$ such that

$$2A^TAx^\star + C^Tz^\star = 2A^Tb, \quad Cx^\star = d$$

Given a feasible x and z, we can write the objective (just as least squares)

$$\begin{aligned}\|Ax - b\|^2 &= \|(Ax - Ax^\star) + (Ax^\star - b)\|^2 \\ &= \|A(x - x^\star)\|^2 + \|Ax^\star - b\|^2 + \underbrace{2(x - x^\star)^TA^T(Ax^\star - b)}_{=0}\end{aligned}$$

We can expand last term, using $2A^T(Ax^\star - b) = -C^Tz^\star$ and $Cx = Cx^\star = d$

$$2(x - x^\star)^TA^T(Ax^\star - b) = -(x - x^\star)^TC^Tz^\star = -(C(x - x^\star))^Tz^\star = 0$$

$$\|Ax - b\|^2 = \overbrace{\|A(x - x^\star)\|^2}^{\geq 0} + \|Ax^\star - b\|^2 \boxed{\geq} \|Ax^\star - b\|^2$$

$x^\star$ is **optimal**

Portfolio optimization

Portfolio allocation weights

We want to invest V dollars in n different *assets* (stocks, bonds, ...) over periods $t = 1, \ldots, T$

Portfolio allocation weights

n-vector w gives the fraction of our total portfolio held in each asset

Portfolio allocation weights

We want to invest V dollars in n different *assets* (stocks, bonds, ...)
over periods $t = 1, \ldots, T$

Portfolio allocation weights

n-vector w gives the fraction of our total portfolio held in each asset

Properties

- Vw_j dollar value hold in asset j
- $\mathbf{1}^T w = 1$ (normalized)
- $w_j < 0$ means short positions (you borrow) (must be returned at time T)
- Example: $w = (-0.2, 0.0, 1.2)$

Portfolio allocation weights

We want to invest V dollars in n different *assets* (stocks, bonds, ...)
over periods $t = 1, \ldots, T$

Portfolio allocation weights

n-vector w gives the fraction of our total portfolio held in each asset

Properties

- Vw_j dollar value hold in asset j
- $\mathbf{1}^T w = 1$ (normalized)
- $w_j < 0$ means short positions (you borrow) (must be returned at time T)
- Example: $w = (-0.2, 0.0, 1.2)$

Short position
of $0.2V$ on asset 1

Portfolio allocation weights

We want to invest V dollars in n different *assets* (stocks, bonds, ...) over periods $t = 1, \ldots, T$

Portfolio allocation weights

n-vector w gives the fraction of our total portfolio held in each asset

Properties

- $V w_j$ dollar value hold in asset j
- $\mathbf{1}^T w = 1$ (normalized)
- $w_j < 0$ means short positions (you borrow) (must be returned at time T)
- Example: $w = (-0.2, 0.0, 1.2)$

Short position of $0.2V$ on asset 1

Don't hold any of asset 2

Portfolio allocation weights

We want to invest V dollars in n different *assets* (stocks, bonds, ...)
over periods $t = 1, \ldots, T$

Portfolio allocation weights

n-vector w gives the fraction of our total portfolio held in each asset

Properties

- Vw_j dollar value hold in asset j
- $\mathbf{1}^T w = 1$ (normalized)
- $w_j < 0$ means short positions (you borrow) (must be returned at time T)
- Example: $w = (-0.2, 0.0, 1.2)$

Short position of $0.2V$ on asset 1

Don't hold any of asset 2

Hold $1.2V$ in asset 3

Leverage, long-only portfolios, and cash

Leverage

$$L = |w_1| + \cdots + |w_n| = \|w\|_1$$

$L = 1$ when all weights are nonnegative (“long only portfolio”)

Leverage, long-only portfolios, and cash

Leverage

$$L = |w_1| + \cdots + |w_n| = \|w\|_1$$

$L = 1$ when all weights are nonnegative (“long only portfolio”)

Uniform portfolio

$$w = \mathbf{1}/n$$

Leverage, long-only portfolios, and cash

Leverage

$$L = |w_1| + \cdots + |w_n| = \|w\|_1$$

$L = 1$ when all weights are nonnegative (“long only portfolio”)

Uniform portfolio

$$w = \mathbf{1}/n$$

Risk free asset

We often assume asset n is ”risk-free” (e.g., cash)

if $w = e_n$, it means the portfolio is all cash

Return over a period

Asset returns

$\tilde{r}_t$ is the (fractional) return of each asset over period t

example: $\tilde{r}_t = (0.01, -0.023, 0.02)$
(often expressed as percentage)

Return over a period

Asset returns

$\tilde{r}_t$ is the (fractional) return of each asset over period t

example: $\tilde{r}_t = (0.01, -0.023, 0.02)$
(often expressed as percentage)

Portfolio return

$$r_t = \tilde{r}_t^T w$$

It is the (fractional) return for the entire portfolio over period t

Return over a period

Asset returns

$\tilde{r}_t$ is the (fractional) return
of each asset over period t

example: $\tilde{r}_t = (0.01, -0.023, 0.02)$
(often expressed as percentage)

Portfolio return

$$r_t = \tilde{r}_t^T w$$

It is the (fractional) return
for the entire portfolio over period t

Total portfolio value after a period

$$V_{t+1} = V_t + V_t \tilde{r}_t^T w = V_t(1 + r_t)$$

Return matrix

Hold constant portfolio
with weights w over T periods

R is the $T \times n$ matrix of **asset returns**

R_{tj} is the return of asset j in period t

$$R = \begin{array}{c|cccc|l} & \text{AAPL} & \text{GOOG} & \text{MMM} & \text{US \$} & \\ \hline & 0.00219 & 0.0006 & -0.00113 & 0.00005 & \text{Mar 1, 2016} \\ & 0.00744 & -0.00894 & -0.00019 & 0.00005 & \text{Mar 2, 2016} \\ & 0.01488 & -0.00215 & 0.00433 & 0.00005 & \text{Mar 3, 2016} \end{array}$$

Return matrix

Hold constant portfolio
with weights w over T periods

Columns interpretation

Column j is time series
of asset j returns

R is the $T \times n$ matrix of **asset returns**

R_{tj} is the return of asset j in period t

$$\begin{array}{ccc} & \begin{array}{cccc} \text{AAPL} & \text{GOOG} & \text{MMM} & \text{US \$} \end{array} & \\ R = & \begin{bmatrix} 0.00219 & 0.0006 & -0.00113 & 0.00005 \\ 0.00744 & -0.00894 & -0.00019 & 0.00005 \\ 0.01488 & -0.00215 & 0.00433 & 0.00005 \end{bmatrix} & \begin{array}{l} \text{Mar 1, 2016} \\ \text{Mar 2, 2016} \\ \text{Mar 3, 2016} \end{array} \end{array}$$

Note. If nth asset risk-free,
the last column of R is $\mu^{\mathrm{rf}}\mathbf{1}$,
where μ^{rf} is the risk-free
per-period interest reate

Return matrix

Hold constant portfolio
with weights w over T periods

R is the $T \times n$ matrix of **asset returns**

R_{tj} is the return of asset j in period t

$$R = \begin{array}{cc} \begin{array}{cccc} \text{AAPL} & \text{GOOG} & \text{MMM} & \text{US \$} \end{array} & \\ \begin{bmatrix} 0.00219 & 0.0006 & -0.00113 & 0.00005 \\ 0.00744 & -0.00894 & -0.00019 & 0.00005 \\ 0.01488 & -0.00215 & 0.00433 & 0.00005 \end{bmatrix} & \begin{array}{l} \text{Mar 1, 2016} \\ \text{Mar 2, 2016} \\ \text{Mar 3, 2016} \end{array} \end{array}$$

Columns interpretation

Column j is time series
of asset j returns

Rows interpretation

Row t is $\tilde{r}_t$ is the asset
return vector over period t

Note. If nth asset risk-free,
the last column of R is $\mu^{\mathrm{rf}}\mathbf{1}$,
where μ^{rf} is the risk-free
per-period interest reate

Return matrix

R is the $T \times n$ matrix of **asset returns**

R_{tj} is the return of asset j in period t

$$R = \begin{array}{c} \begin{array}{cccc} \text{AAPL} & \text{GOOG} & \text{MMM} & \text{US \$} \end{array} \\ \begin{bmatrix} 0.00219 & 0.0006 & -0.00113 & 0.00005 \\ 0.00744 & -0.00894 & -0.00019 & 0.00005 \\ 0.01488 & -0.00215 & 0.00433 & 0.00005 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Mar 1, 2016} \\ \text{Mar 2, 2016} \\ \text{Mar 3, 2016} \end{array}$$

Hold constant portfolio
with weights w over T periods

Columns interpretation

Column j is time series of asset j returns

Note. If nth asset risk-free, the last column of R is $\mu^{\mathrm{rf}}\mathbf{1}$, where μ^{rf} is the risk-free per-period interest reate

Rows interpretation

Row t is $\tilde{r}_t$ is the asset return vector over period t

Portfolio returns (time series)

$$r = Rw \quad (T\text{-vector})$$

Returns over multiple periods

r is time series T-vector of portfolio returns

Returns over multiple periods

r is time series T-vector of portfolio returns

average return
(or just return)

$$\mathbf{avg}(r) = \mathbf{1}^T r/T$$

Returns over multiple periods

r is time series T-vector of portfolio returns

average return
(or just return)

$$\mathbf{avg}(r) = \mathbf{1}^T r / T$$

risk
(standard deviation)

$$\mathbf{std}(r) = \|r - \mathbf{avg}(r)\mathbf{1}\| / \sqrt{T}$$

Returns over multiple periods

r is time series T-vector of portfolio returns

average return
(or just return)

$$\mathbf{avg}(r) = \mathbf{1}^T r / T$$

risk
(standard deviation)

$$\mathbf{std}(r) = \|r - \mathbf{avg}(r)\mathbf{1}\| / \sqrt{T}$$

Total portfolio value

$$\begin{aligned} V_{T+1} &= V_1(1 + r_1) \cdots (1 + r_T) \\ &\approx V_1 + V_1(r_1 + \cdots + r_T) \\ &= V_1 + T\mathbf{avg}(r)V_1 \end{aligned}$$

(for $|r_t|$ small, e.g., ≤ 0.01
ignore higher order terms)

Returns over multiple periods

r is time series T-vector of portfolio returns

average return
(or just return)

$$\mathbf{avg}(r) = \mathbf{1}^T r / T$$

risk
(standard deviation)

$$\mathbf{std}(r) = \|r - \mathbf{avg}(r)\mathbf{1}\| / \sqrt{T}$$

Total portfolio value

$$\begin{aligned} V_{T+1} &= V_1(1 + r_1) \cdots (1 + r_T) \\ &\approx V_1 + V_1(r_1 + \cdots + r_T) \\ &= V_1 + T\mathbf{avg}(r)V_1 \end{aligned}$$

(for $|r_t|$ small, e.g., ≤ 0.01
ignore higher order terms)

For high portfolio value we need large $\mathbf{avg}(r)$

Annualized return and risk

Mean return and and risk are often expressed in **annualized form** (per year)

Annualized return and risk

Mean return and and risk are often expressed in **annualized form** (per year)

Given P trading periods per year (i.e., 250 days)

annualized return $= P\mathbf{avg}(r)$, annualized risk $= \sqrt{P}\mathbf{std}(r)$

Portfolio optimization

How shall we choose the portfolio weight vector w?

Portfolio optimization

How shall we choose the portfolio weight vector w?

Goals

High (mean) return
$\mathbf{avg}(r)$

Low risk
$\mathbf{std}(r)$

Portfolio optimization

How shall we choose the portfolio weight vector w?

Goals

High (mean) return	Low risk
$\mathbf{avg}(r)$	$\mathbf{std}(r)$

Data

- We know **realized asset returns** but not future ones
- **Optimization.** We choose w that would have worked well in the past
- **True goal.** Hope it will work well in the future (just like data fitting)

$r = Rw$

Portfolio optimization

Minimize risk given a target return

Chose n-vector w to solve

$$\begin{array}{ll} \text{minimize} & \mathbf{std}(Rw)^2 = (1/T)\|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \mathbf{1}^T w = 1 \\ & \mathbf{avg}(Rw) = \rho \end{array}$$

Portfolio optimization

Minimize risk given a target return

Chose n-vector w to solve

$$\begin{array}{ll} \text{minimize} & \mathbf{std}(Rw)^2 = (1/T)\|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \mathbf{1}^T w = 1 \\ & \mathbf{avg}(Rw) = \rho \end{array}$$

(past) portfolio
returns time series

Portfolio optimization

Minimize risk given a target return

Chose n-vector w to solve

$$\begin{array}{ll} \text{minimize} & \mathbf{std}(Rw)^2 = (1/T)\|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \mathbf{1}^T w = 1 \\ & \mathbf{avg}(Rw) = \rho \end{array}$$

(past) target
mean return

(past) portfolio
returns time series

Portfolio optimization

Minimize risk given a target return

Chose n-vector w to solve

$$\begin{array}{ll} \text{minimize} & \mathbf{std}(Rw)^2 = (1/T)\|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \mathbf{1}^T w = 1 \\ & \mathbf{avg}(Rw) = \rho \end{array}$$

(past) target mean return

(past) portfolio returns time series

Solutions w are **Pareto optimal**

Portfolio optimization

Minimize risk given a target return

Chose n-vector w to solve

$$\begin{array}{ll} \text{minimize} & \mathbf{std}(Rw)^2 = (1/T)\|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \mathbf{1}^T w = 1 \\ & \mathbf{avg}(Rw) = \rho \end{array}$$

(past) target mean return

(past) portfolio returns time series

Solutions w are **Pareto optimal**

Our question

what would have been the best constant allocation w,
had we known future returns?

Example allocations

Annual return 1% (risk-free asset has 1% return)

$$w = (0.00, 0.00, 0.00, \ldots, 0.00, 0.00, 1.00)$$

Example allocations

Annual return 1% (risk-free asset has 1% return)

$$w = (0.00, 0.00, 0.00, \ldots, 0.00, 0.00, 1.00)$$

Annual return 13%

$$w = (0.02, -0.07, -0.05, \ldots, -0.03, 0.06, 0.56)$$

Example allocations

Annual return 1% (risk-free asset has 1% return)

$$w = (0.00, 0.00, 0.00, \ldots, 0.00, 0.00, 1.00)$$

Annual return 13%

$$w = (0.02, -0.07, -0.05, \ldots, -0.03, 0.06, 0.56)$$

Annual return 25%

$$w = (0.05, -0.143, -0.09, \ldots, -0.07, 0.12, 0.12)$$

Example allocations

Annual return 1% (risk-free asset has 1% return)

$$w = (0.00, 0.00, 0.00, \dots, 0.00, 0.00, 1.00)$$

Annual return 13%

$$w = (0.02, -0.07, -0.05, \dots, -0.03, 0.06, 0.56)$$

Annual return 25%

$$w = (0.05, -0.143, -0.09, \dots, -0.07, 0.12, 0.12)$$

Asking for higher annual returns yields

- More invested in risky, but high return assets
- Larger short positions (”leveraging”)

Portfolio optimization

As constrained least squares

$$\begin{array}{ll} \text{minimize} & \|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix} \end{array}$$

Portfolio optimization

As constrained least squares

$$\begin{array}{ll} \text{minimize} & \|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix} \end{array}$$

μ is the n-vector of average returns per asset

$$\begin{aligned} \mathbf{avg}(r) &= (1/T)\mathbf{1}^T(\overbrace{Rw}^{r}) \\ &= (1/T)(R^T\mathbf{1})^T w = \mu^T w \end{aligned}$$

Portfolio optimization

As constrained least squares

$$\begin{array}{ll} \text{minimize} & \|\overbrace{Rw}^{A} - \overbrace{\rho\mathbf{1}}^{b}\|^2 \\ \text{subject to} & \underbrace{\begin{bmatrix}\mathbf{1}^T \\ \mu^T\end{bmatrix}}_{C} w = \left.\begin{bmatrix}1 \\ \rho\end{bmatrix}\right\} d \end{array}$$

μ is the n-vector of average returns per asset

$$\begin{aligned} \mathbf{avg}(r) &= (1/T)\mathbf{1}^T(Rw) \\ &= (1/T)(R^T\mathbf{1})^T w = \mu^T w \end{aligned}$$

Solution via KKT linear system

$$C\begin{bmatrix} \overbrace{2R^TR}^{A^TA} & \overset{C^T}{\boxed{\mathbf{1} \quad \mu}} \\ \boxed{\begin{matrix}\mathbf{1}^T \\ \mu^T\end{matrix}} & \begin{matrix}0 & 0 \\ 0 & 0\end{matrix} \end{bmatrix} \begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix} \begin{matrix} \}\, 2A^Tb \\ \left.\begin{matrix}\\ \\\end{matrix}\right\} d \end{matrix}$$

Optimal portfolios

Rewrite right-hand side

$$\begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2T\mu \\ 0 \\ 0 \end{bmatrix}$$

Rewrite right-hand side

$$\begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2T\mu \\ 0 \\ 0 \end{bmatrix}$$

Optimal portfolios

Two fund theorem

Optimal portfolio w is an affine function of ρ

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2T\mu \\ 0 \\ 1 \end{bmatrix}$$

Optimal portfolios

Rewrite right-hand side

$$\begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2T\mu \\ 0 \\ 0 \end{bmatrix}$$

Two fund theorem

Optimal portfolio w is an affine function of ρ

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2T\mu \\ 0 \\ 1 \end{bmatrix}$$

We can rewrite the first n-components as
the combination of two portfolios (funds)

$$w = w_0 + \rho v$$

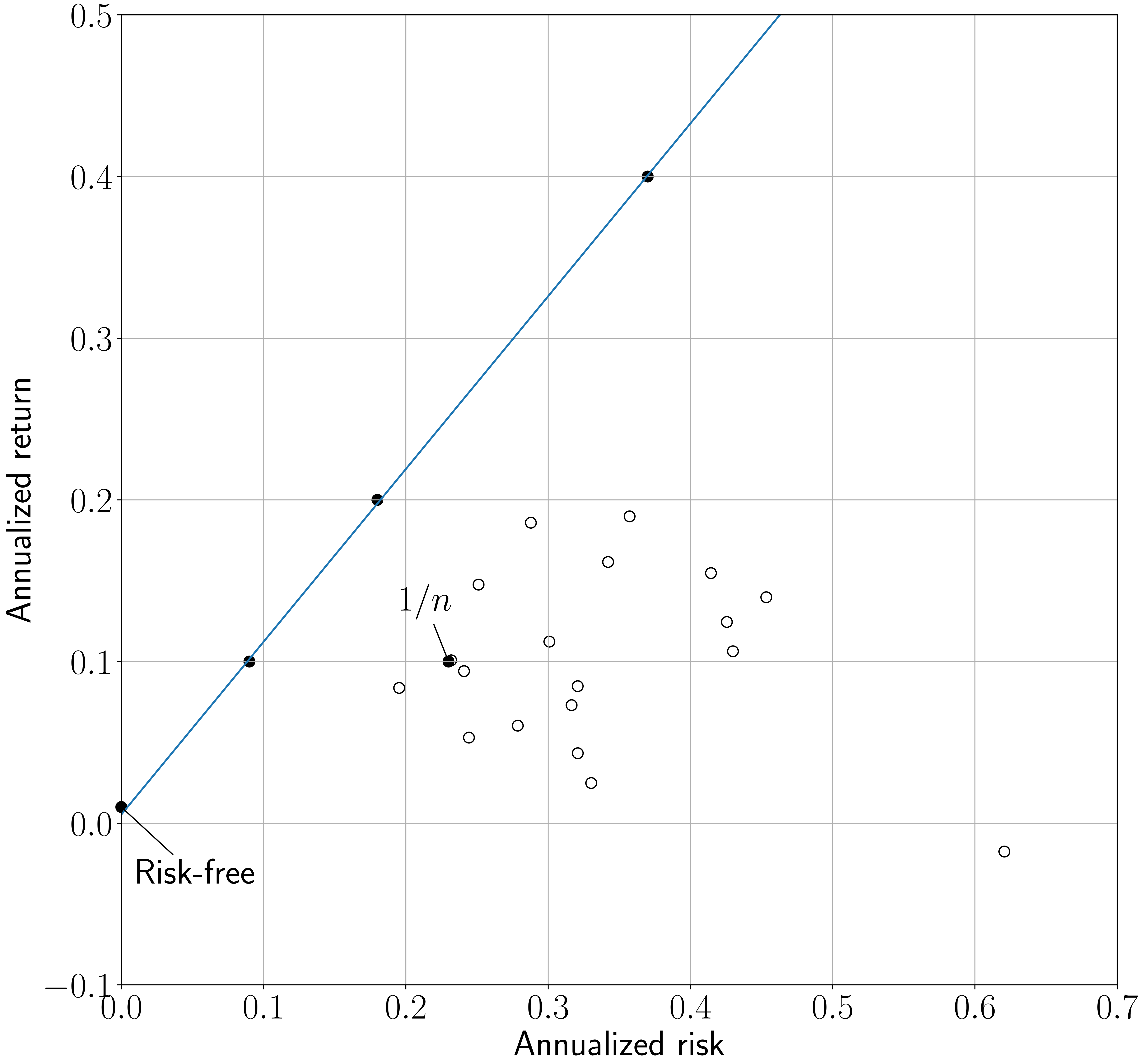
Risk-free
($\rho = 0$)

Other optimal portfolio

Example

20 assets over 2000 days (past)

- Optimal portfolios on a **straight line**
- Line starts at risk-free portfolio ($\rho = 0$)
- $1/n$ much better than single portfolios



The big assumption

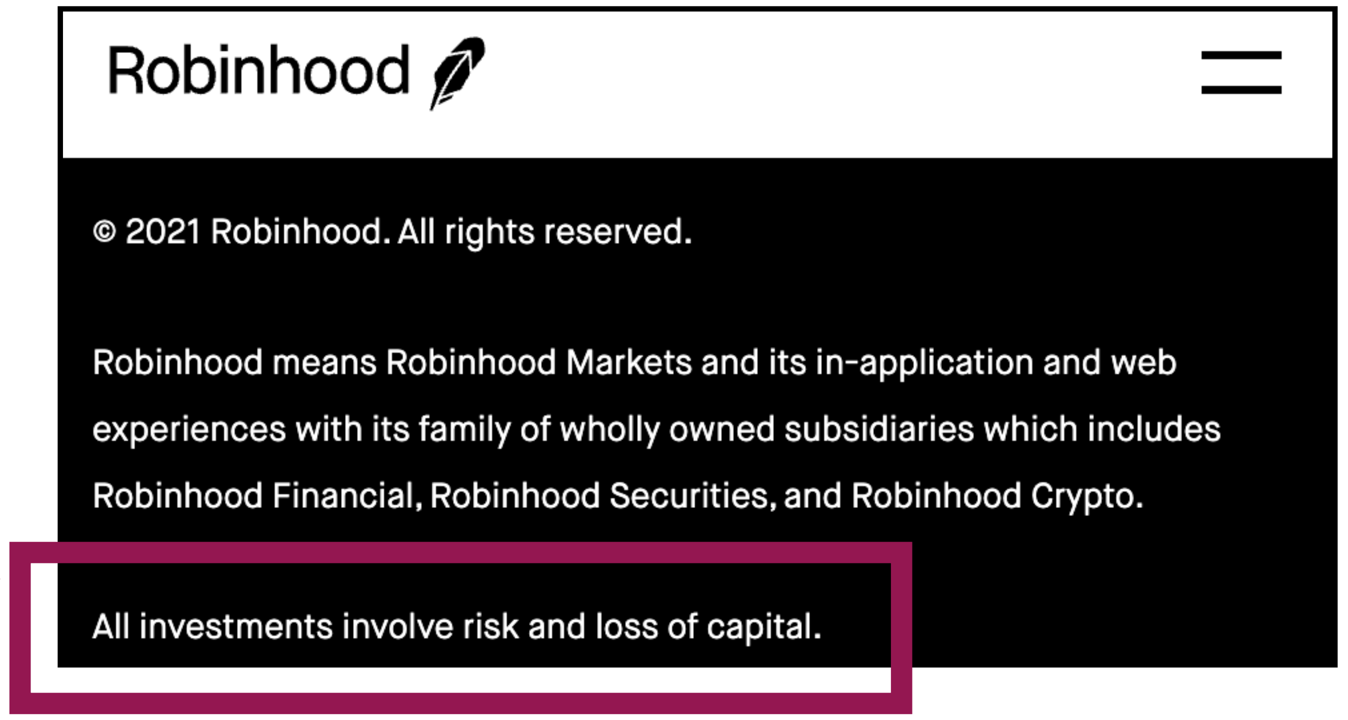
Future returns will look like past ones

- You are warned this is false, every time you invest
- It is often reasonable
- During crisis, market shifts, other big events not true

The big assumption

Future returns will look like past ones

- You are warned this is false, every time you invest
- It is often reasonable
- During crisis, market shifts, other big events not true



Robinhood

© 2021 Robinhood. All rights reserved.

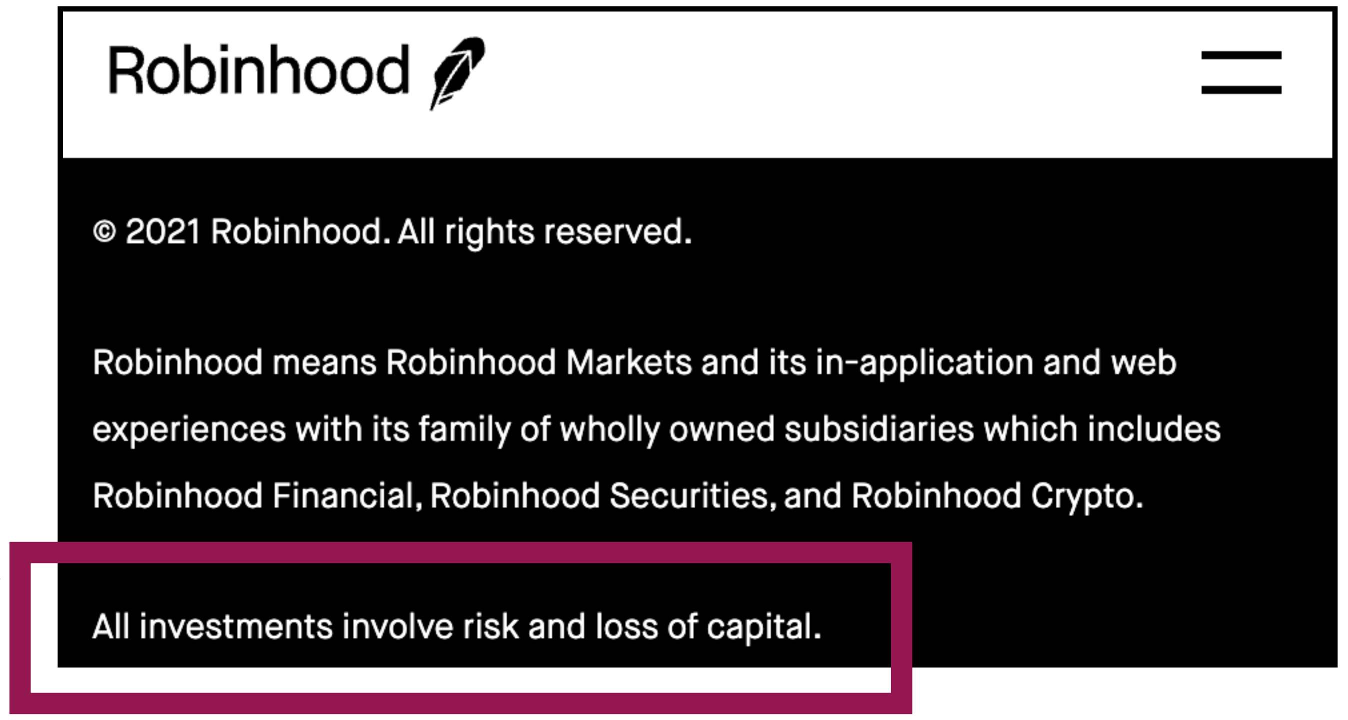
Robinhood means Robinhood Markets and its in-application and web experiences with its family of wholly owned subsidiaries which includes Robinhood Financial, Robinhood Securities, and Robinhood Crypto.

All investments involve risk and loss of capital.

The big assumption

Future returns will look like past ones

- You are warned this is false, every time you invest
- It is often reasonable
- During crisis, market shifts, other big events not true



© 2021 Robinhood. All rights reserved.

Robinhood means Robinhood Markets and its in-application and web experiences with its family of wholly owned subsidiaries which includes Robinhood Financial, Robinhood Securities, and Robinhood Crypto.

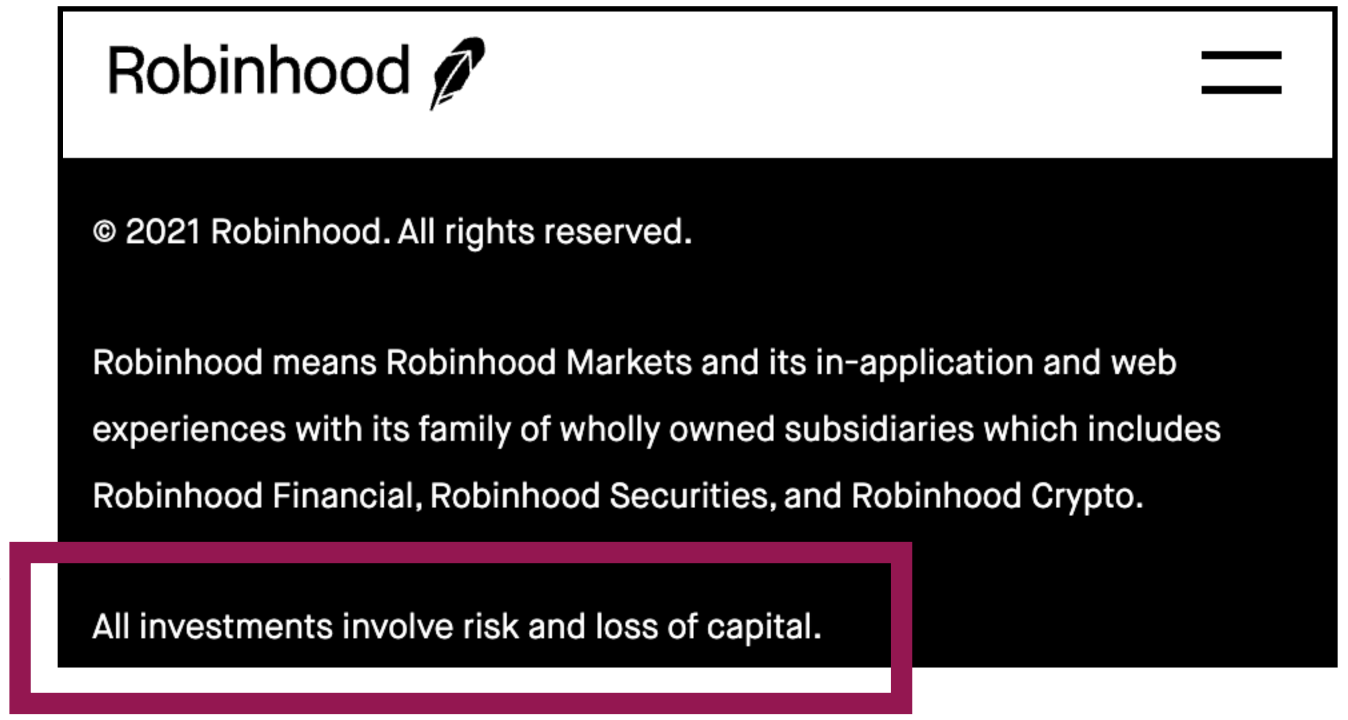
All investments involve risk and loss of capital.

If assumption holds (even approximately), a good w on past returns leads to good future (unknown) returns

The big assumption

Future returns will look like past ones

- You are warned this is false, every time you invest
- It is often reasonable
- During crisis, market shifts, other big events not true



Robinhood

© 2021 Robinhood. All rights reserved.

Robinhood means Robinhood Markets and its in-application and web experiences with its family of wholly owned subsidiaries which includes Robinhood Financial, Robinhood Securities, and Robinhood Crypto.

All investments involve risk and loss of capital.

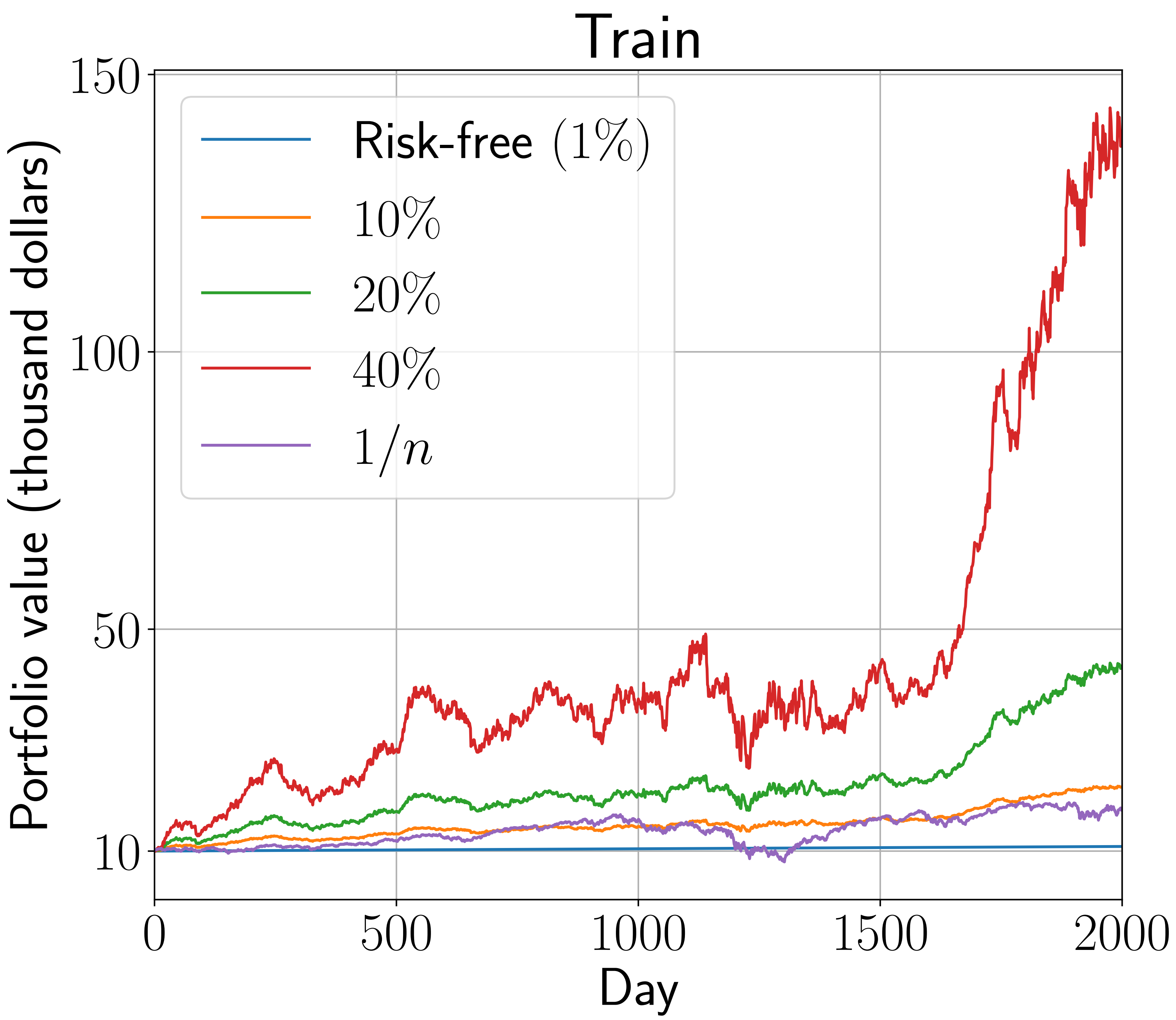
If assumption holds (even approximately), a good w on past returns leads to good future (unknown) returns

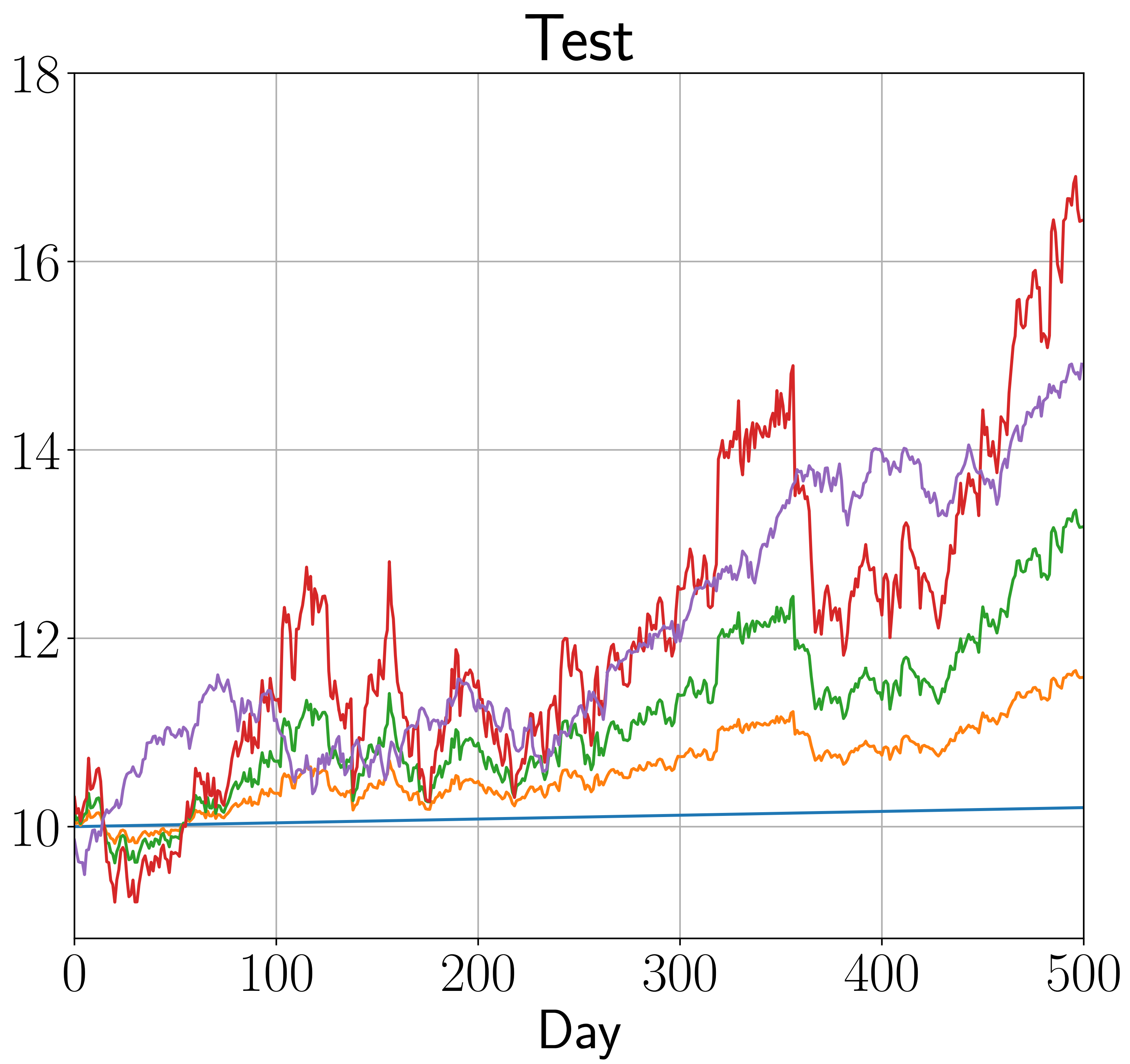
Example

- Pick w based on last 2 years of returns
- Use w during next 6 months

Total portfolio value

	Return		Risk		
	Train	Test	Train	Test	Leverage
Risk-free (1%)	0.01	0.01	0.00	0.00	1.00
10%	0.10	0.08	0.09	0.07	1.96
20%	0.20	0.15	0.18	0.15	3.03
40%	0.40	0.30	0.37	0.31	5.48
$1/n$	0.10	0.21	0.23	0.13	1.00





Build your quantitative hedge fund

Rolling portfolio optimization

For each period t, find weight w_t using L past returns

$$r_{t-1}, \dots, r_{t-L}$$

Build your quantitative hedge fund

Rolling portfolio optimization

For each period t, find weight w_t using L past returns

$$r_{t-1}, \ldots, r_{t-L}$$

Variations

- Update w every K periods (monthly, quarterly, ...)
- Add secondary objective $\lambda \|w_t - w_{t-1}\|^2$ to discourage turnover, reduce transaction cost
- Add logic to detect when the future is likely to not look like the past
- Add "signals" that predict future return of assets (Twitter sentiment analysis)

Constrained least squares

Today, we learned to:

- **Formulate** (linearly) and **solve** constrained least squares problems
- **Solve** portfolio allocations problems
- **Understand** the difference between past and future returns (be careful!)

References

- S. Boyd, L. Vandenberghe: Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares
  - Chapter 16 and 17: constrained least squares

Next lecture

- Linear optimization