ORF522 – Linear and Nonlinear Optimization

11. Interior-point methods implementation

Bartolomeo Stellato — Fall 2022

Recap

Optimality conditions

			Primal		**Dual**
minimize	$c^T x$	$\longrightarrow$	minimize $\quad c^T x$		maximize $\quad -b^T y$
subject to	$Ax \leq b$		subject to $\quad Ax + s = b$		subject to $\quad A^T y + c = 0$
			$s \geq 0$		$y \geq 0$

Optimality conditions

$$\begin{aligned} Ax + s - b &= 0 \\ A^T y + c &= 0 \\ s_i y_i &= 0 \\ s, y &\geq 0 \end{aligned}$$

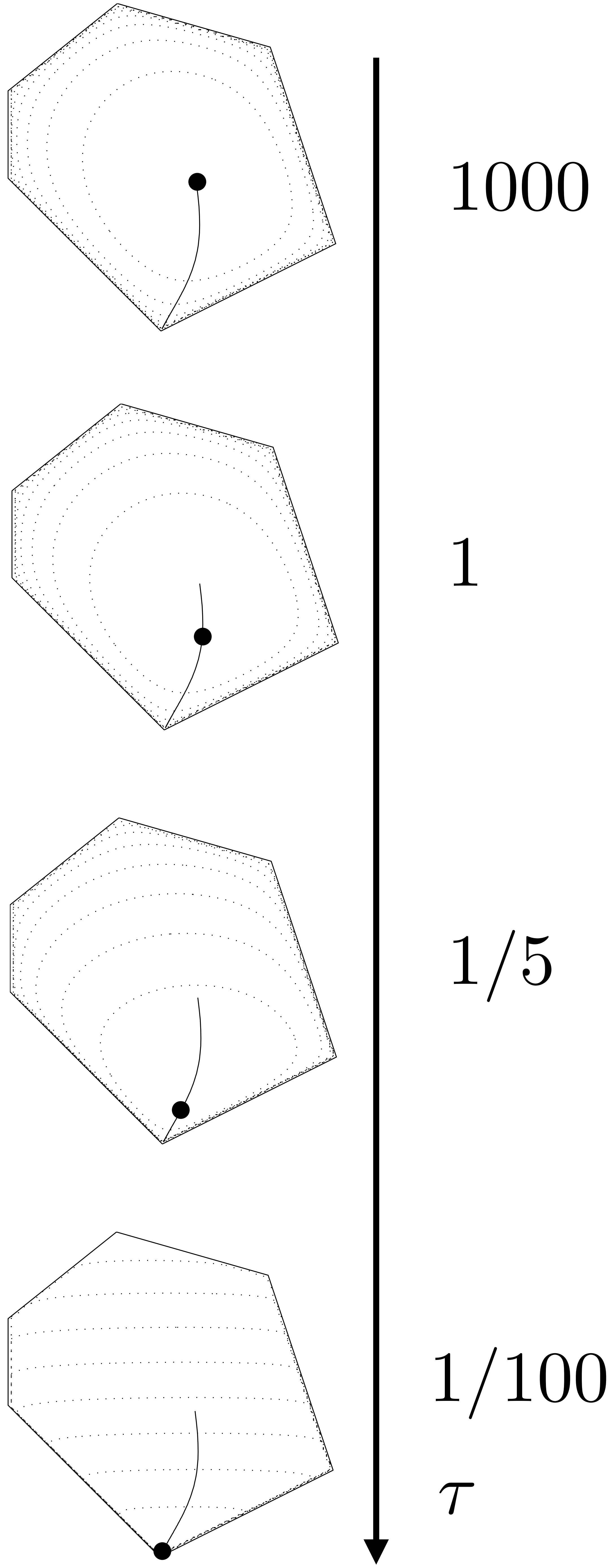
Central path

Analytic Center
$\tau \to \infty$

minimize $\quad c^T x - \tau \sum_{i=1}^m \log(s_i)$
subject to $\quad Ax + s = b$

Set of points $(x^\star(\tau), s^\star(\tau), y^\star(\tau))$
with $\tau > 0$ such that

$$Ax + s - b = 0$$
$$A^T y + c = 0$$
$$s_i y_i = \tau$$
$$s, y \geq 0$$



Main idea

Follow central path as $\tau \to 0$

Strict complementarity

Primal

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax + s = b \\ & s \geq 0 \end{array}$$

Dual

$$\begin{array}{ll} \text{maximize} & -b^T y \\ \text{subject to} & A^T y + c = 0 \\ & y \geq 0 \end{array}$$

Theorem
If the two problems have feasible solutions, then there exist feasible s and y with a **strict complementary sparsity** pattern:

$$y_i > 0, s_i = 0 \quad \text{or} \quad y_i = 0, s_i > 0$$

In other words, $s_i + y_i > 0$

Proof (left as exercise)
Details in [Theorem 10.6, LP]

Main idea

Optimality conditions

$$h(x, s, y) = \begin{bmatrix} Ax + s - b \\ A^T y + c \\ SY\mathbf{1} \end{bmatrix} = 0 \qquad \begin{aligned} S &= \mathbf{diag}(s) \\ Y &= \mathbf{diag}(y) \end{aligned}$$

$$s, y \geq 0$$

- Apply variants of Newton's method to solve $h(x, s, y) = 0$
- Enforce $s, y > 0$ (strictly) at every iteration
- **Motivation** avoid getting stuck in "corners"

Algorithm step

Linear system

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} + \sigma\mu\mathbf{1} \end{bmatrix}$$

Duality measure

$$\mu = \frac{s^T y}{m}$$

Centering parameter

$\sigma \in [0, 1]$

$\sigma = 0 \quad \Rightarrow \quad$ Newton step

$\sigma = 1 \quad \Rightarrow \quad$ Centering step towards $(x^\star(\mu), s^\star(\mu), y^\star(\mu))$

Line search to enforce $s, y > 0$

$$(x, s, y) \leftarrow (x, s, y) + \alpha(\Delta x, \Delta s, \Delta y)$$

Primal-dual path-following algorithm

Initialization

1. Given (x_0, s_0, y_0) such that $s_0, y_0 > 0$

Iterations

1. Choose $\sigma \in [0, 1]$
2. Solve $\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY + \sigma\mu\mathbf{1} \end{bmatrix}$ where $\mu = s^T y/m$
3. Find maximum α such that $y + \alpha\Delta y > 0$ and $s + \alpha\Delta s > 0$
4. Update $(x, s, y) \leftarrow (x, s, y) + \alpha(\Delta x, \Delta s, \Delta y)$

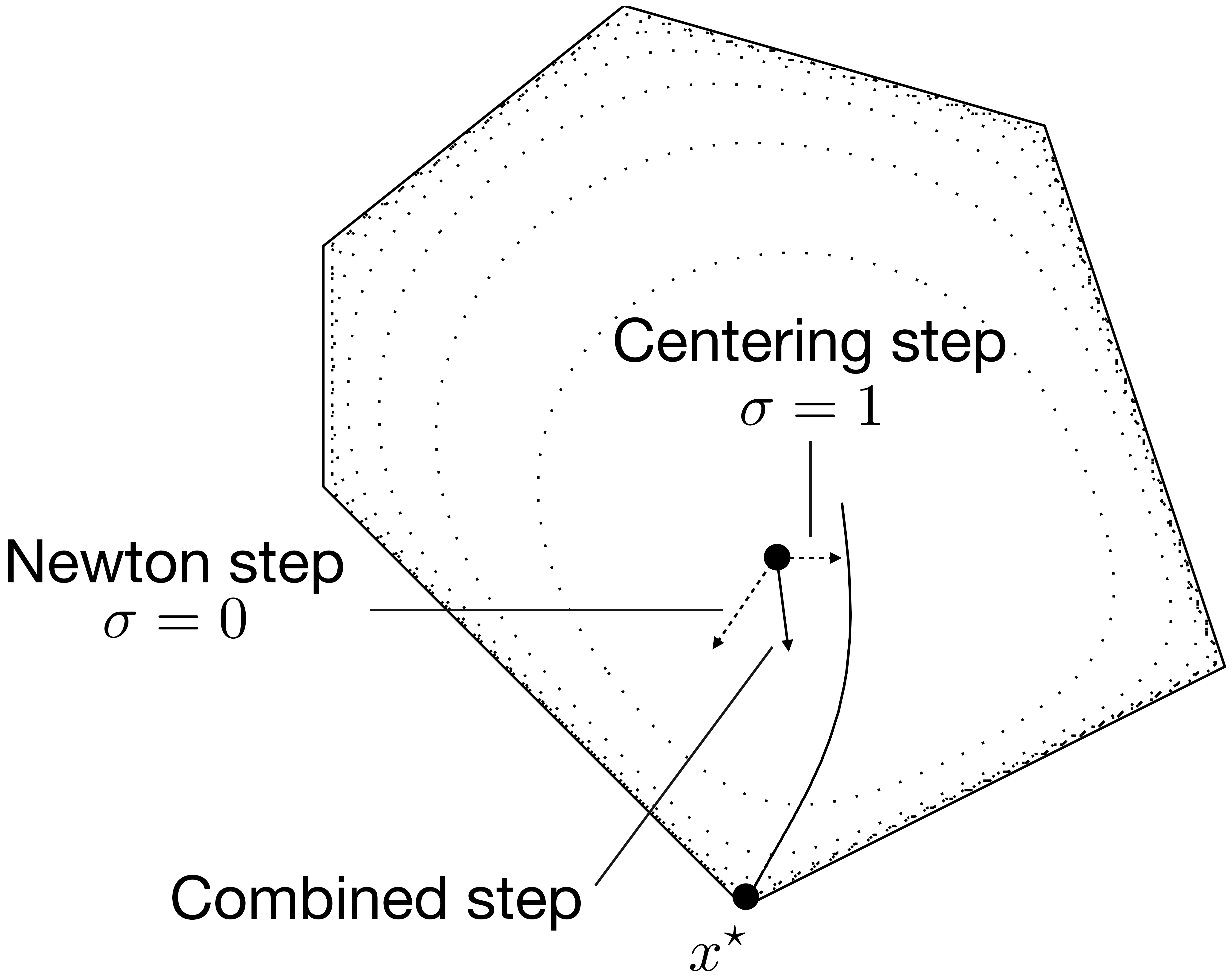
Today's lecture
[Chapter 14, NO][Chapter 22, LP]

- Mehrotra predictor-corrector algorithm
- Implementation details
- Homogeneous self-dual embedding
- Interior-point vs simplex

Predictor-corrector algorithm

Main idea:

Predict and select centering parameter



Predict

Compute Newton direction

Estimate

How good is the Newton step?
(how much can μ decrease?)

Select centering parameter

Very roughly:
Pick $\sigma \approx 0$ if Newton step is good
Pick $\sigma \approx 1$ if Newton step is bad

Select centering parameter

Newton step

$$(\Delta x_a, \Delta s_a, \Delta y_a)$$

Maximum step-size

$$\alpha_p = \max\{\alpha \in [0,1] \mid s + \alpha \Delta s_a \geq 0\}$$
$$\alpha_d = \max\{\alpha \in [0,1] \mid y + \alpha \Delta y_a \geq 0\}$$

Duality measure candidate
(after Newton step)

$$\mu_a = \frac{(s + \alpha_p \Delta s_a)^T (y + \alpha_d \Delta y_a)}{m}$$

⟶

Centering parameter heuristic σ

$$\sigma = \left(\frac{\mu_a}{\mu}\right)^3$$

Mehrotra correction

Newton step

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y_a \\ \Delta x_a \\ \Delta s_a \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} \end{bmatrix} \longrightarrow s_i(\Delta y_a)_i + y_i(\Delta s_a)_i + s_i y_i = 0$$

Full step

$$(s_i + (\Delta s_a)_i)(y_i + (\Delta y_a)_i) = (\Delta s_a)_i(\Delta y_a)_i \neq 0$$

Complementarity violation depends on step length

Corrected direction

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} - \Delta S_a \Delta Y_a \mathbf{1} + \sigma\mu\mathbf{1} \end{bmatrix}$$

$$\Delta S_a = \mathbf{diag}(\Delta s_a)$$
$$\Delta Y_a = \mathbf{diag}(\Delta y_a)$$

Mehrotra predictor-corrector algorithm

Initialization

Given (x, s, y) such that $s, y > 0$

1. **Termination conditions**

   $$r_p = Ax + s - b, \quad r_d = A^T y + c, \quad \mu = (s^T y)/m$$

   If $\|r_p\|, \|r_d\|, \mu$ are small, **break** Optimal solution $(x^\star, s^\star, y^\star)$

2. **Newton step** (affine scaling)

   $$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y_a \\ \Delta x_a \\ \Delta s_a \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} \end{bmatrix}$$

Mehrotra predictor-corrector algorithm

3. **Barrier parameter**

$$\alpha_p = \max\{\alpha \in [0,1] \mid s + \alpha \Delta s_a \geq 0\}$$

$$\alpha_d = \max\{\alpha \in [0,1] \mid y + \alpha \Delta y_a \geq 0\}$$

$$\mu_a = \frac{(s + \alpha_p \Delta s_a)^T (y + \alpha_d \Delta y_a)}{m}$$

$$\sigma = \left(\frac{\mu_a}{\mu}\right)^3$$

4. **Corrected direction**

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} - \Delta S_a \Delta Y_a \mathbf{1} + \sigma \mu \mathbf{1} \end{bmatrix}$$

Mehrotra predictor-corrector algorithm

5. **Update iterates**

$$\alpha_p = \max\{\alpha \geq 0 \mid s + \alpha \Delta s \geq 0\}$$
$$\alpha_d = \max\{\alpha \geq 0 \mid y + \alpha \Delta y \geq 0\}$$

$$(x, s) = (x, s) + \min\{1, \eta\alpha_p\}(\Delta x, \Delta s)$$
$$y = y + \min\{1, \eta\alpha_d\}\Delta y$$

Avoid corners

$$\eta = 1 - \epsilon \approx 0.99$$

Implementation details

Search equations

Step 2 (**Newton**) and 4 (**Corrected direction**) solve equations of the form

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} b_y \\ b_x \\ b_s \end{bmatrix}$$

The **Newton** step right hand side: $$\begin{bmatrix} b_y \\ b_x \\ b_s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} \end{bmatrix}$$

The **corrector** step right hand side: $$\begin{bmatrix} b_y \\ b_x \\ b_s \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -SY\mathbf{1} - \Delta S_a \Delta Y_a \mathbf{1} + \sigma\mu\mathbf{1} \end{bmatrix}$$

Solving the search equations

Our linear system is not symmetric

$$\begin{bmatrix} 0 & A & I \\ A^T & 0 & 0 \\ S & 0 & Y \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta s \end{bmatrix} = \begin{bmatrix} b_y \\ b_x \\ b_s \end{bmatrix}$$

Substitute last equation, $\Delta s = Y^{-1}(b_s - S\Delta y)$, into first

$$\begin{bmatrix} -Y^{-1}S & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \end{bmatrix} = \begin{bmatrix} b_y - Y^{-1}b_s \\ b_x \end{bmatrix}$$

Substitute first equation, $\Delta y = S^{-1}Y(A\Delta x - b_y + Y^{-1}b_s)$, into second

$$A^T S^{-1} Y A \Delta x = b_x + A^T S^{-1} Y b_y - A^T S^{-1} b_s$$

Simplified linear system

Coefficient matrix

$$B = A^T S^{-1} Y A$$

Characteristics

- A is **large** and **sparse**
- $S^{-1}Y$ is **positive** and **diagonal**, different at each iteration
- B is **positive definite** if $\mathrm{rank}(A) = n$
- Sparsity pattern of B is the **pattern** of $A^T A$ (independent of $S^{-1}Y$)

Cholesky factorizations

$$B = PLL^T P^T$$

- Reordering only once to get P
- One numerical factorizaton per interior-point iteration $O(n^3)$ ⟶ **Per-iteration complexity** $O(n^3)$
- Forward/backward substitution twice per iteration $O(n^2)$

Convergence

Mehrotra's algorithm

No convergence theory → Examples where it **diverges** (rare!)

Fantastic convergence **in practice** → Less than 30 iterations

Floating point operations

Theoretical iteration complexity

Alternative versions (slower than Mehrotra) converge in $O(\sqrt{n})$ iterations → $O(n^{3.5})$

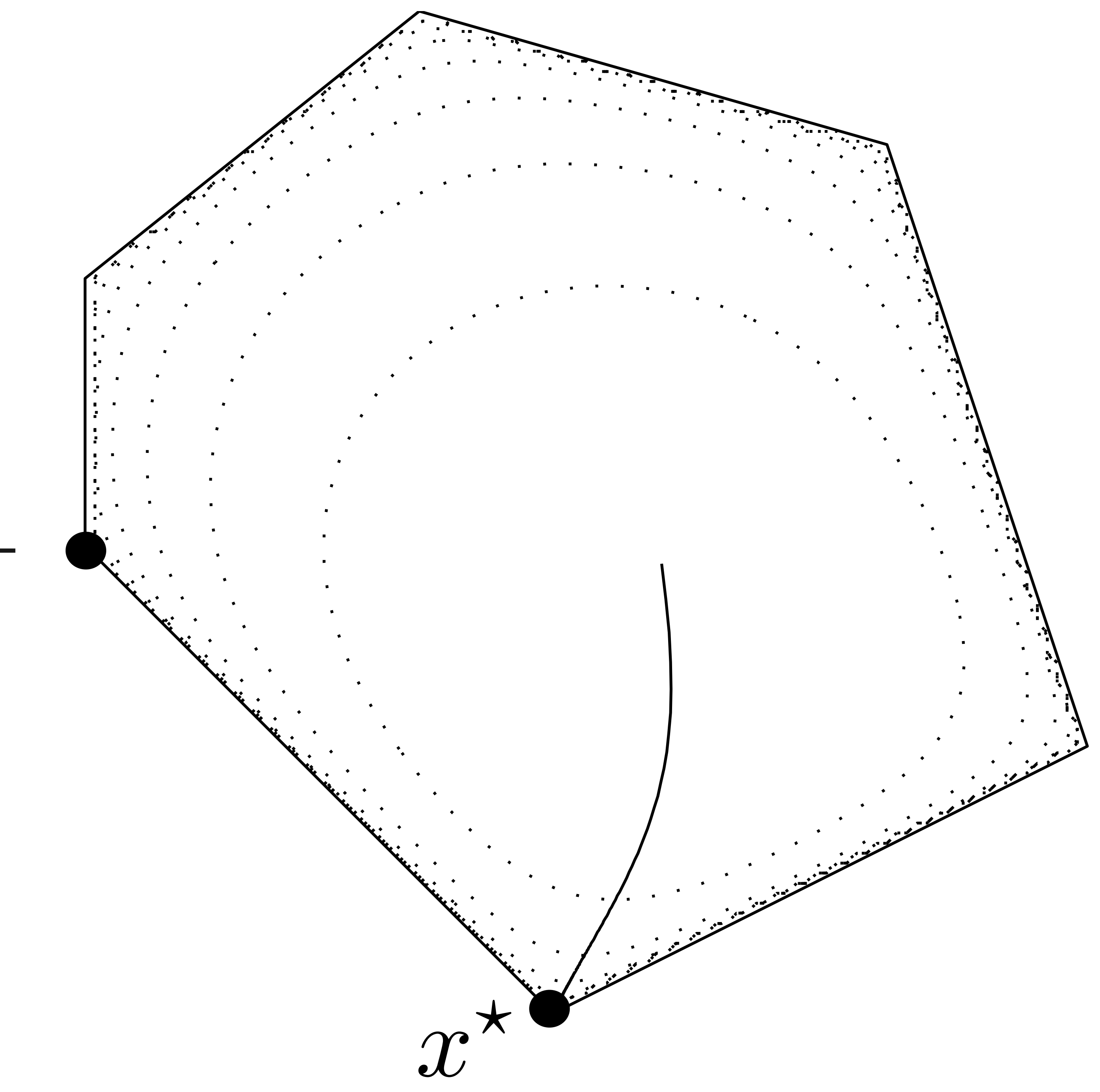
Average iteration complexity

Average iterations complexity is $O(\log n)$ → $O(n^3 \log n)$

Warm-starting

Interior-point methods are **difficult to warm-start**

Previous solution



Badly centered
initial point

↓

Hard to make progress
with long steps

Homogeneous self-dual embedding

Optimality conditions

Primal

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax + s = b \\ & s \geq 0 \end{array}$$

Dual

$$\begin{array}{ll} \text{maximize} & -b^T y \\ \text{subject to} & A^T y + c = 0 \\ & y \geq 0 \end{array}$$

Optimality conditions

$$\begin{bmatrix} 0 \\ s \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & A^T \\ -A & 0 \\ c^T & b^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ b \\ 0 \end{bmatrix}$$

$$s, y \geq 0$$

Any $(x^\star, s^\star, y^\star)$ satisfying these conditions is **optimal**

What happens if the problem is infeasible?

How do you detect infeasibility/unboundedness?

Primal

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax + s = b \\
& s \geq 0
\end{array}
$$

Dual

$$
\begin{array}{ll}
\text{maximize} & -b^T y \\
\text{subject to} & A^T y + c = 0 \\
& y \geq 0
\end{array}
$$

Alternatives (Farkas lemma) Write feasibility problem and dualize…

- **primal feasible:** $Ax + s = b, \quad s \geq 0$
- **primal infeasible:** $A^T y = 0, \quad b^T y < 0, \quad y \geq 0$ (primal infeasibility certificate)

- **dual feasible:** $A^T y + c = 0, \quad y \geq 0$
- **dual infeasible:** $Ax \leq 0, \quad c^T x < 0$ (dual infeasibility certificate)

The homogeneous self-dual embedding

Derivation

Introduce two new variables $\kappa, \tau \geq 0$

Homogeneous self-dual embedding

$$\begin{bmatrix} 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \tau \end{bmatrix}$$

$$s, y, \kappa, \tau \geq 0$$

$\longrightarrow$

$$Qu = v$$
$$u, v \geq 0$$

$$Q = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix}$$

$$u = (x, y, \tau)$$
$$v = (0, s, \kappa)$$

The homogeneous self-dual embedding

Properties

$$Qu = v$$
$$u, v \geq 0$$

$$Q = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix}$$

$$u = (x, y, \tau)$$
$$v = (0, s, \kappa)$$

Matrix

- Q is **skew-symmetric**: $Q^T = -Q \quad \Rightarrow \quad u^T Q u = 0$
- $u \perp v$ **proof** $Qu - v = 0 \quad \Rightarrow \quad u^T Q u - u^T v = 0 \quad \Rightarrow \quad u^T v = 0 \quad \blacksquare$

Homogeneous

(u, v) satisfy $Qu = v,\ (v, u) \geq 0 \quad \Rightarrow \quad \alpha(u, v)$ with $\alpha \geq 0$ feasible

Always feasible

$\alpha = 0 \quad \Rightarrow \quad (0, 0)$ is feasible

Self-dual problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Qu = v \\ & u, v \geq 0 \end{array}$$

Q skew-symmetric: $Q^T = -Q$

The dual is identical to the primal

Proof

$$g(\nu, \lambda, \mu) = \underset{u,v}{\text{minimize}}\ \mathcal{L}(u, v, \nu, \lambda, \mu) = \nu^T(Qu - v) - \lambda^T u - \mu^T v$$

$$\frac{\partial \mathcal{L}}{\partial u} = Q^T \nu - \lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial v} = -\nu^T - \mu = 0 \quad \Rightarrow \quad \nu = -\mu$$

$\longrightarrow$

Dual

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Q\mu = \lambda \\ & \mu, \lambda \geq 0 \end{array}$$

■

From self-duality to strict complementarity

Primal

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Qu = v \\ & u, v \geq 0 \end{array}$$

Dual

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Q\mu = \lambda \\ & \mu, \lambda \geq 0 \end{array}$$

$$u = (x, y, \tau)$$
$$v = (0, s, \kappa)$$

LP strict complementarity

$$u^T \lambda = 0, \quad u + \lambda > 0$$
$$v^T \mu = 0, \quad v + \mu > 0$$

$\longrightarrow$

Self-dual

$$u = \mu, \quad v = \lambda$$

$\longrightarrow$

Strict complementarity

$$u + v > 0 \quad \Rightarrow \quad \begin{array}{l} y + s > 0 \\ \tau + \kappa > 0 \end{array}$$

The homogeneous self-dual embedding

Outcomes

Find x, s, y, κ, τ such that

$$\begin{bmatrix} 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \tau \end{bmatrix}$$

$$s, y, \kappa, \tau \geq 0$$

Note. By strict complementarity, we can ensure $\kappa + \tau > 0$

Case 1: feasibility

$\tau > 0, \kappa = 0$ define $(\hat{x}, \hat{s}, \hat{y}) = (x^\star/\tau, s^\star/\tau, y^\star/\tau)$

$$\begin{aligned} 0 &= A^T \hat{y} + c \\ \hat{s} &= -A\hat{x} + b \end{aligned} \qquad \hat{s} \geq 0, \quad \hat{y} \geq 0, \quad \hat{s}^T \hat{y} = 0$$

⟶ $(\hat{x}, \hat{s}, \hat{y})$ is a **solution** to the original problem

The homogeneous self-dual embedding

Outcomes

Find x, s, y, κ, τ such that
$$\begin{bmatrix} 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \tau \end{bmatrix}$$
$$s, y, \kappa, \tau \geq 0$$

Case 2: infeasibility

$\tau = 0, \kappa > 0$ ⟶ $c^T x + b^T y < 0$ (**impossible**). Must have infeasibility

If $b^T y < 0$ then $\hat{y} = y/(-b^T y)$ is a **certificate of primal infeasibility**
$$A^T \hat{y} = 0, \quad b^T \hat{y} = -1 < 0, \quad \hat{y} \geq 0$$

If $c^T x < 0$ then $\hat{x} = x/(-c^T x)$ is a **certificate of dual infeasibility**
$$A\hat{x} \leq 0, \quad c^T \hat{x} = -1 < 0$$

Interior-point method for homogeneous self-dual embedding

Linear complementarity problem

$$
\begin{aligned}
Qu &= v \\
u^T v &= 0 \\
u, v &\geq 0
\end{aligned}
$$

Equations

$$
h(u, v) = \begin{bmatrix} Qu - v \\ UV\mathbf{1} \end{bmatrix} = 0
$$

$$
u, v \geq 0
$$

Directions

$$
\begin{bmatrix} Q & -I \\ V & U \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \begin{bmatrix} -r_e \\ -UV\mathbf{1} + \sigma\mu\mathbf{1} \end{bmatrix} \qquad \begin{aligned} r_e &= Qu - v \\ \mu &= (u^T v)/d \end{aligned}
$$

Line search to enforce $u, v > 0$

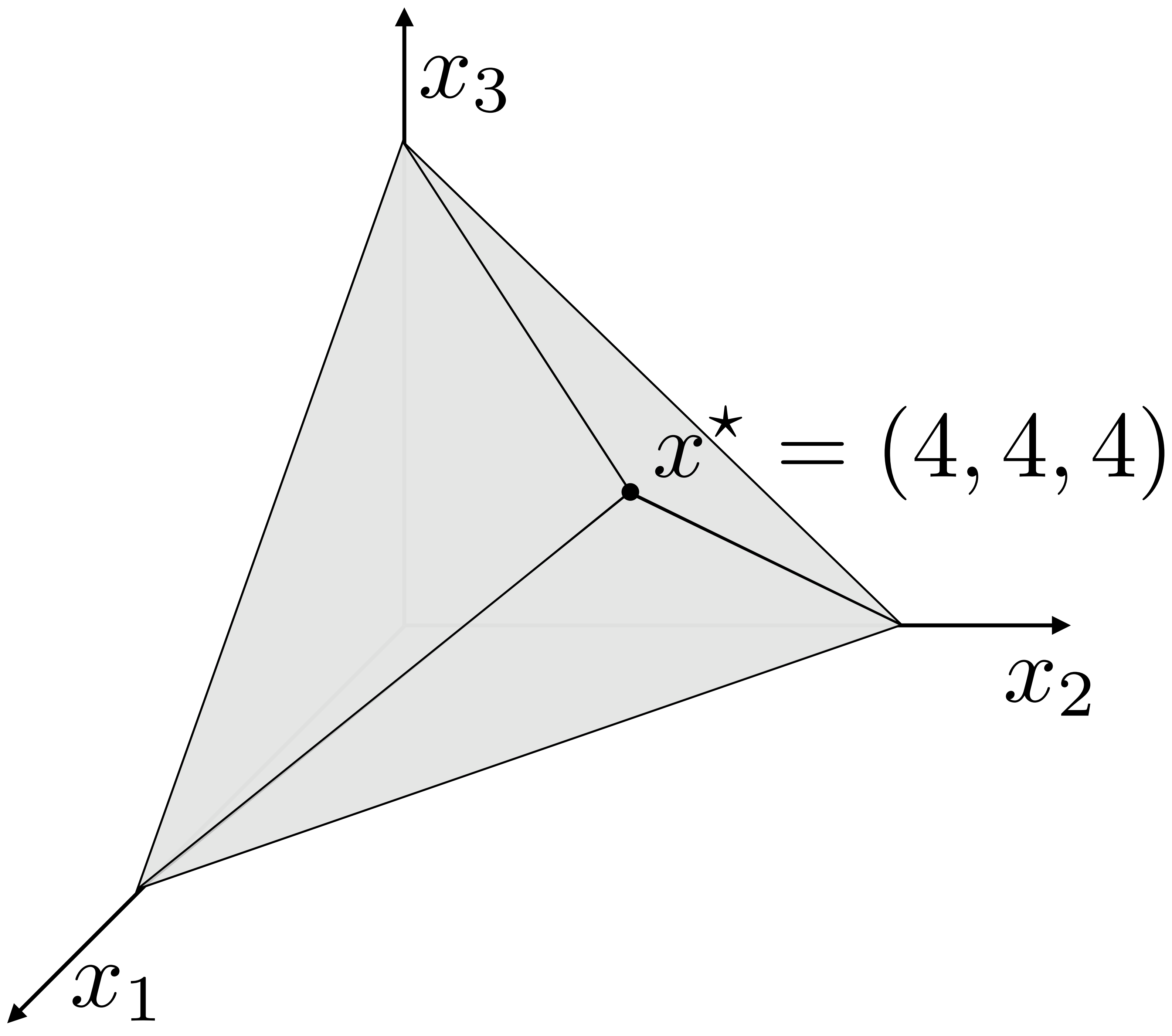
$$
(u, v) \leftarrow (u, v) + \alpha(\Delta u, \Delta v)
$$

Interior-point methods can solve **linear complementarity problems**

Interior-point vs simplex

Example

$$\begin{array}{ll} \text{minimize} & -10x_1 - 12x_2 - 12x_3 \\ \text{subject to} & x_1 + 2x_2 + 2x_3 \leq 20 \\ & 2x_1 + x_2 + x_3 \leq 20 \\ & 2x_1 + 2x_2 + x_3 \leq 20 \\ & x_1, x_2, x_3 \geq 0 \end{array}$$



$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \\ & x \geq 0 \end{array}$$

$$c = (-10, -12, -12)$$

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

$$b = (20, 20, 20)$$

Example with real solver

CVXOPT (open-source)

Code

```
import numpy as np
import cvxpy as cp

c = np.array([-10, -12, -12])
A = np.array([[1, 2, 2],
              [2, 1, 2],
              [2, 2, 1]])
b = np.array([20, 20, 20])
n = len(c)

x = cp.Variable(n)
problem = cp.Problem(cp.Minimize(c @ x),
                     [A @ x <= b, x >= 0])
problem.solve(solver=cp.CVXOPT, verbose=True)
```

Output

```
     pcost       dcost       gap    pres   dres   k/t
 0: -1.3077e+02 -2.3692e+02  2e+01  1e-16  6e-01  1e+00
 1: -1.3522e+02 -1.4089e+02  1e+00  2e-16  3e-02  4e-02
 2: -1.3599e+02 -1.3605e+02  1e-02  2e-16  3e-04  4e-04
 3: -1.3600e+02 -1.3600e+02  1e-04  1e-16  3e-06  4e-06
 4: -1.3600e+02 -1.3600e+02  1e-06  1e-16  3e-08  4e-08
Optimal solution found.
```

Solution

```
In [3]: x.value
Out[3]: array([3.99999999, 4.        , 4.        ])
```

[The CVXOPT linear and quadratic cone program solvers, L. Vandenberghe 2010]

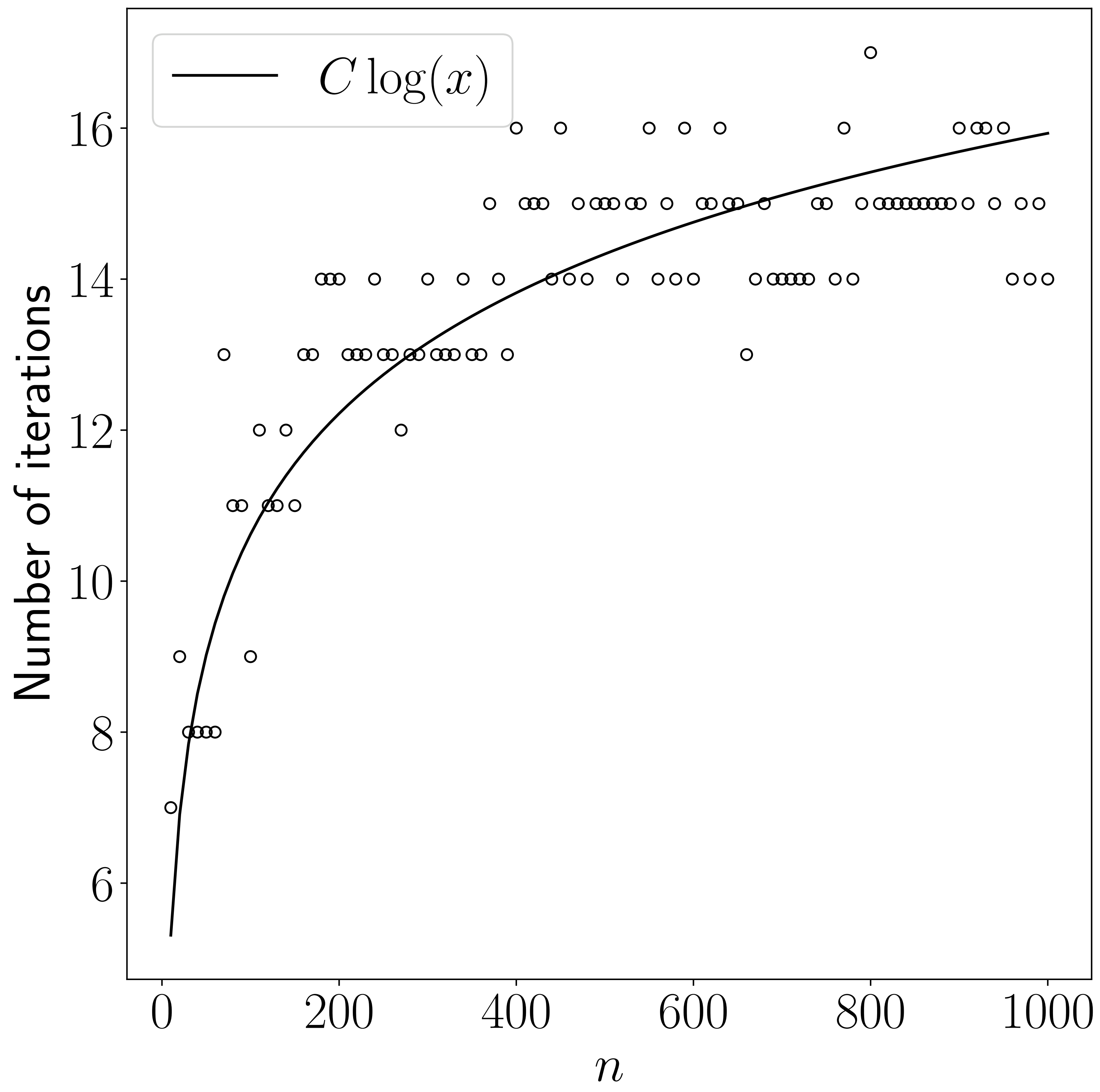
Average interior-point complexity

Random LPs

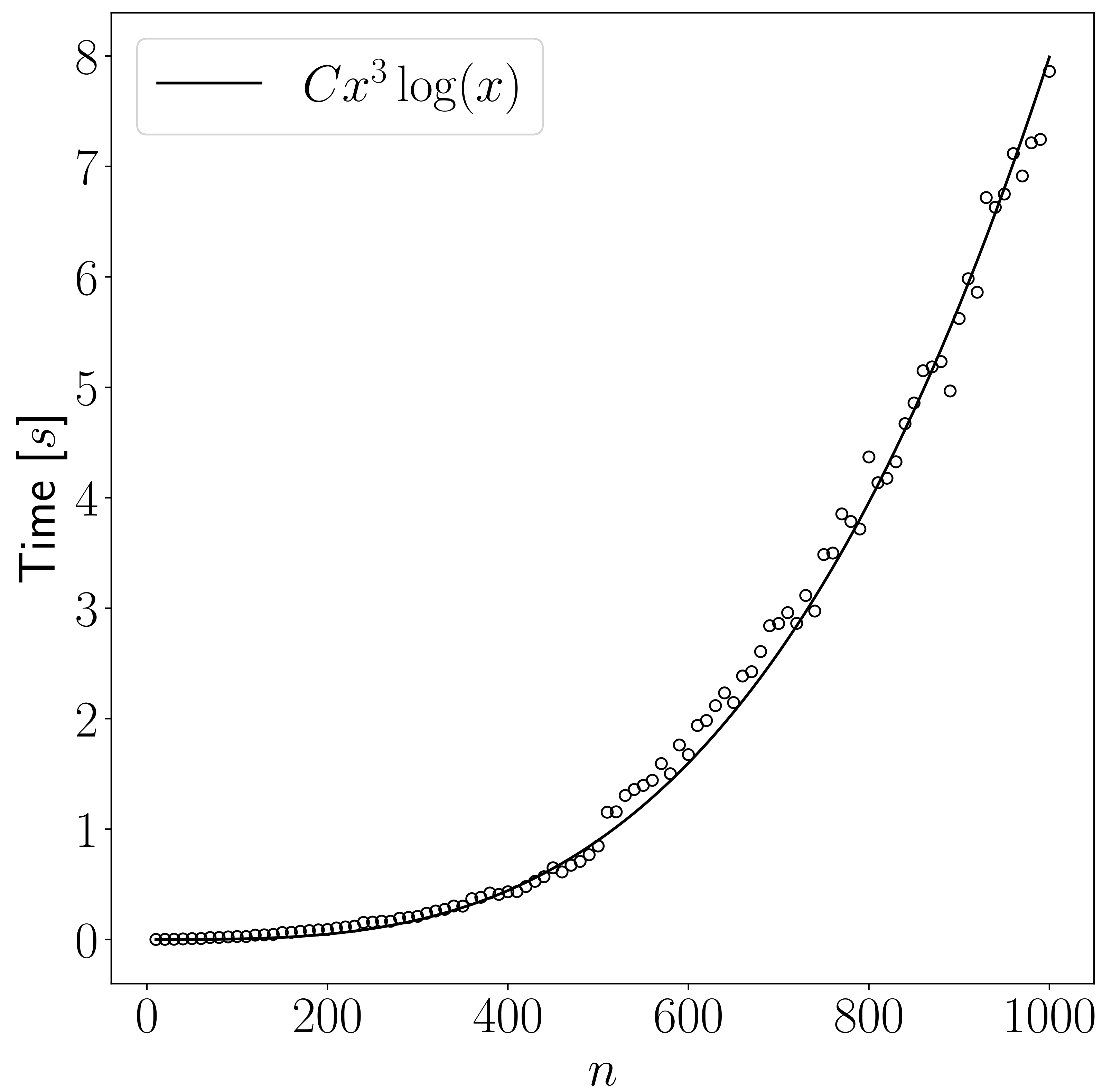
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \le b \end{array}$$

n variables
$3n$ constraints

Iterations: $O(\log n)$



Time: $O(n^3 \log n)$



Comparison between interior-point method and simplex

Primal simplex

- Primal feasibility
- Zero duality gap

↓

Dual feasibility

Dual simplex

- Dual feasibility
- Zero duality gap

↓

Primal feasibility

Primal-dual interior-point

- Interior condition

↓

- Primal feasibility
- Dual feasibility
- Zero duality gap

Exponential worst-case complexity

Requires feasible point

Can be warm-started

Polynomial worst-case complexity

Allows infeasible start

Cannot be warm-started

Which algorithm should I use?

Dual simplex

- Small-to-medium problems
- Repeated solves with varying data

Interior-point (barrier)

- Medium-to-large problems
- Sparse structured problems

How do solvers with multiple options decide?

<u>Concurrent Optimization</u>

Why not both? (crossover)

Interior-point → Few simplex steps

Interior-point methods implementation

Today, we learned to:

- **Apply** Mehrotra predictor-corrector algorithm
- **Exploit** linear algebra to speedup computations
- **Detect** infeasibility/unboundedness with homogeneous self-dual embedding
- **Analyze** empirical complexity
- **Compare** interior-point and simplex methods

Next lecture

- Introduction to nonlinear optimization