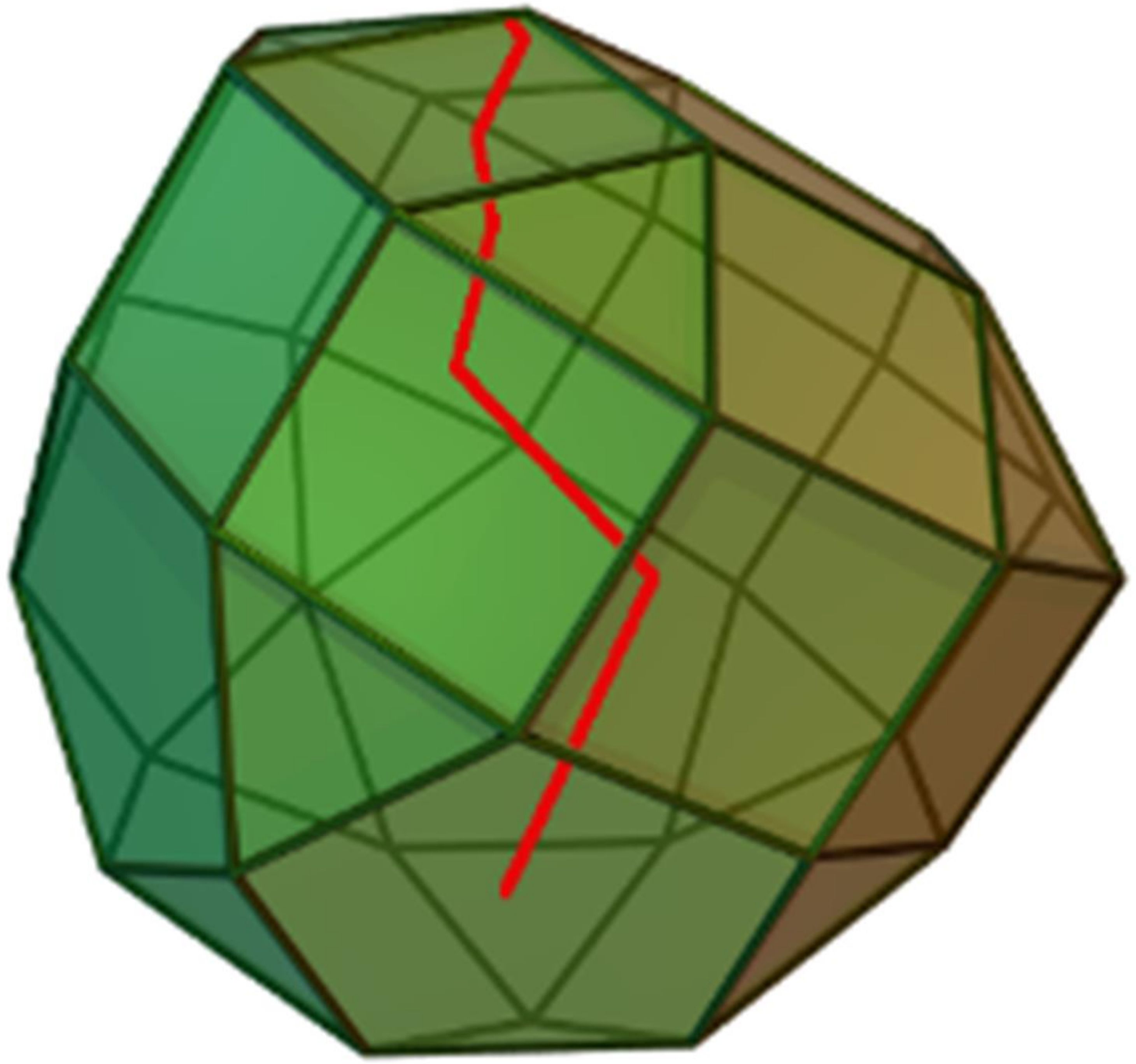
Bootcamp: Interior Point Methods II

Bento Natura, Takashi Tsuchiya, and Yinyu Ye



Friday, September 01 2023
Simons Institute

The Various Notions of ‘Solving’ LP

…or how much error ε do we allow…

$min\langle c,x\rangle : Ax = b, x \geq 0,$
m constraints, n variables

ε-approximate solution $\bar{x}$:

$$\langle c,\bar{x}\rangle \leq OPT + \varepsilon, \| A\bar{x} - b \| \leq \varepsilon, \| x^- \| \leq \varepsilon$$

Accuracy	**Low**	**High**	**Exact**
Dependency on accuracy	$\text{poly}(1/\varepsilon)$	$\text{poly}(\log(1/\varepsilon))$	$\varepsilon = 0$
Algorithms	Multiplicative weights, First order methods,…	Ellipsoid Method, Interior Point Methods,…	Simplex Method, proximity based solvers, specialized IPM

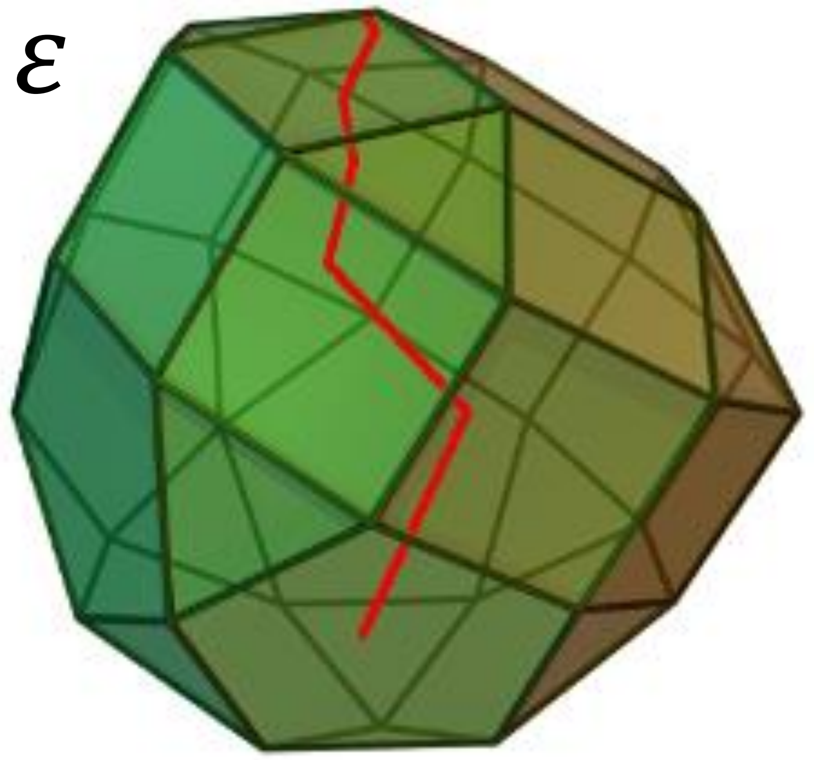
Improvement on weakly polynomial solvers for LP

…progress of recent years for high accuracy solvers…

$min\langle c, x\rangle$: $Ax = b, x \geq 0$,
m constraints, n variables

ε-approximate solution $\bar{x}$:

$\langle c, \bar{x}\rangle \leq OPT + \varepsilon, \| A\bar{x} - b \| \leq \varepsilon, \| x^{-} \| \leq \varepsilon$



Interior Point Methods

Running times ($\log(1/\varepsilon)$ terms omitted):

$\sqrt{m}(\mathrm{nnz}(A) + m^2)$ Lee-Sidford '13-'19

n^{ω} Cohen, Lee, Song '19, van den Brand '20, Jiang, Song, Weinstein, Zhang '21

$nm + m^{2.5}$ van den Brand, Lee, Liu, Saranurak, Sidford, Song, Wang '21

ω…matrix multiplication exponent: It takes $\tilde{O}(n^{\omega})$ to multiply two $n \times n$ matrices.

Scaling Invariance in LP Algorithms

- Let $\mathcal{D}$ be the set of diagonal matrices with positive nonzero diagonal entries, and let $D \in \mathcal{D}$.

$$(\mathrm{P}_{\mathrm{scaled}}) \quad \min \; c^T D x' \text{ s.t. } A D x' = b, \; x' \geq 0.$$

$$(\mathrm{D}_{\mathrm{scaled}}) \quad \max \; b^T y \text{ s.t. } Dc - DA^T y = s', \; s' \geq 0.$$

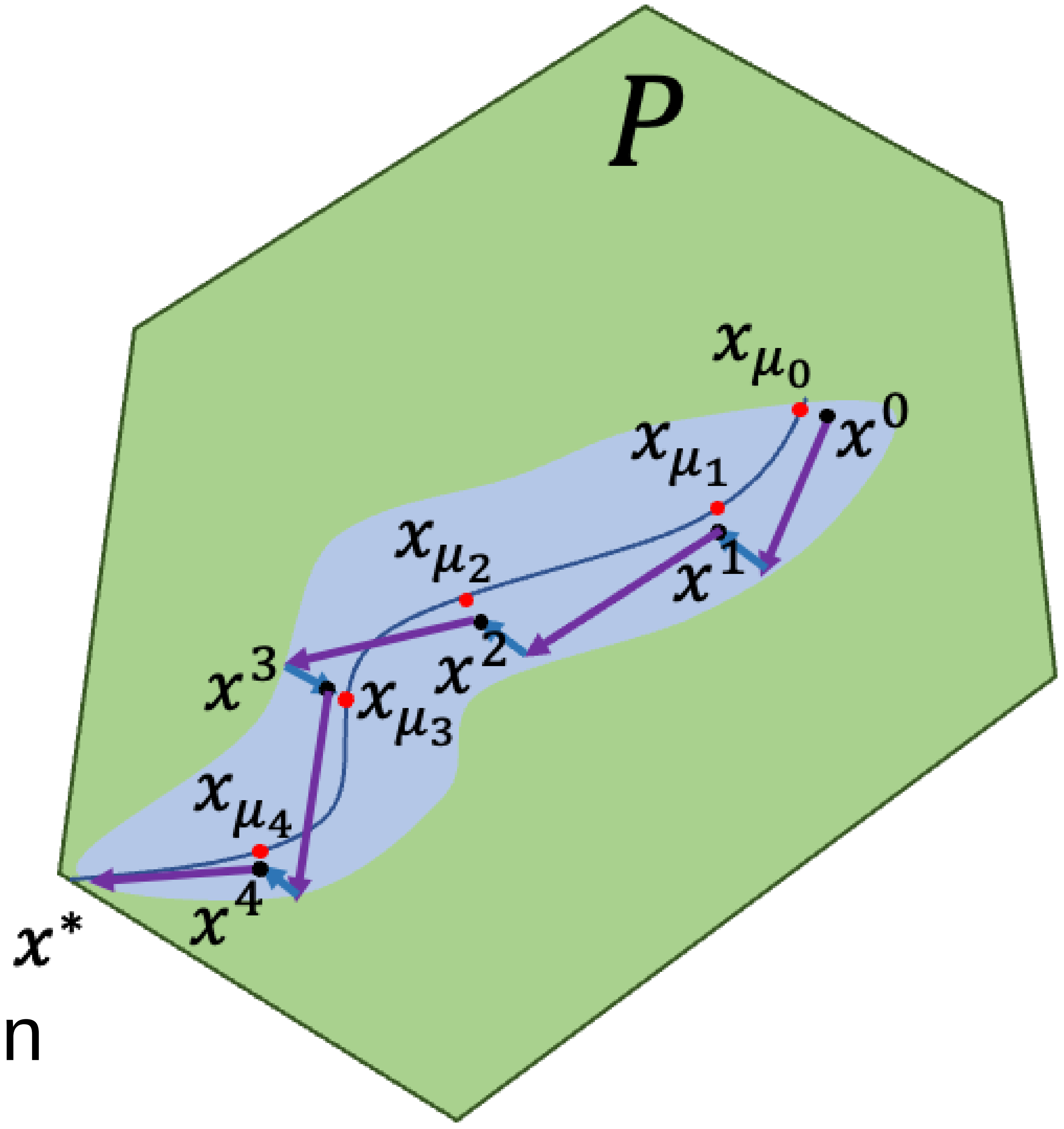
  $(\mathrm{P}_{\mathrm{scaled}})$ and $(\mathrm{D}_{\mathrm{scaled}})$ are equivalent to (P) and (D).
- An algorithm is called "scaling invariant" if it generates the (geometrically) the identical sequences when applied to (P) and (D), and $(\mathrm{P}_{\mathrm{scaled}})$ and $(\mathrm{D}_{\mathrm{scaled}})$.
- The simplex algorithm is not scaling invariant, whereas MTY-PC algorithm is scaling invariant.

Recall Predictor - Corrector Path Following

Mizuno-Todd-Y '93

$min\langle c, x\rangle: A^\top x \geq b,$
d variables, n inequalities

- Given x^0 in 'neighborhood' around x_{μ_0} for some $\mu_0 > 0$
- Compute iterates $x^1, \dots, x^t$ by alternating between
  - Predictor steps: decrease μ by moving 'down' the central path
  - Corrector steps: move back 'closer' to the central path for the same μ (Newton step).



Each iteration requires a linear system solve.

Standard analysis as seen before: Decrease μ by a factor of 2 in $O(\sqrt{n})$ iterations

Condition Numbers in LP Algorithms

- Let B be a set of indicies of columns of A, and let A_B be the submatrix associate with B.
- Let $\mathcal{B}$ be the set of index set B such that A_B is invertible.
- The condition number $\bar{\chi}_A$ of A is defined as follows:

$$\bar{\chi}_A = \max_{B \in \mathcal{B}} \|A_B^{-1} A\|.$$

Introduced by Dikin'67.
Used in Stewart '89,
Todd '90, Vavasis-Y 96',
Monteiro and Tsuchiya
…

- If the input bit length of A is L_A, then, $\bar{\chi}_A = 2^{O(L_A)}$.
- $\bar{\chi}_A$ is not scaling invariant; namely, $\bar{\chi}_{AD} \neq \bar{\chi}_A$.

$$\bar{\chi}_A^* = \min_{D \in \mathcal{D}} \bar{\chi}_{AD}$$

($\mathcal{D}$ is the set of diagonal matrices with positive diagonal entries.)
If an algorithm is scaling invariant, then we can use $\bar{\chi}_A^*$ for evaluating complexity.

Condition Number Based Complexity Analyses

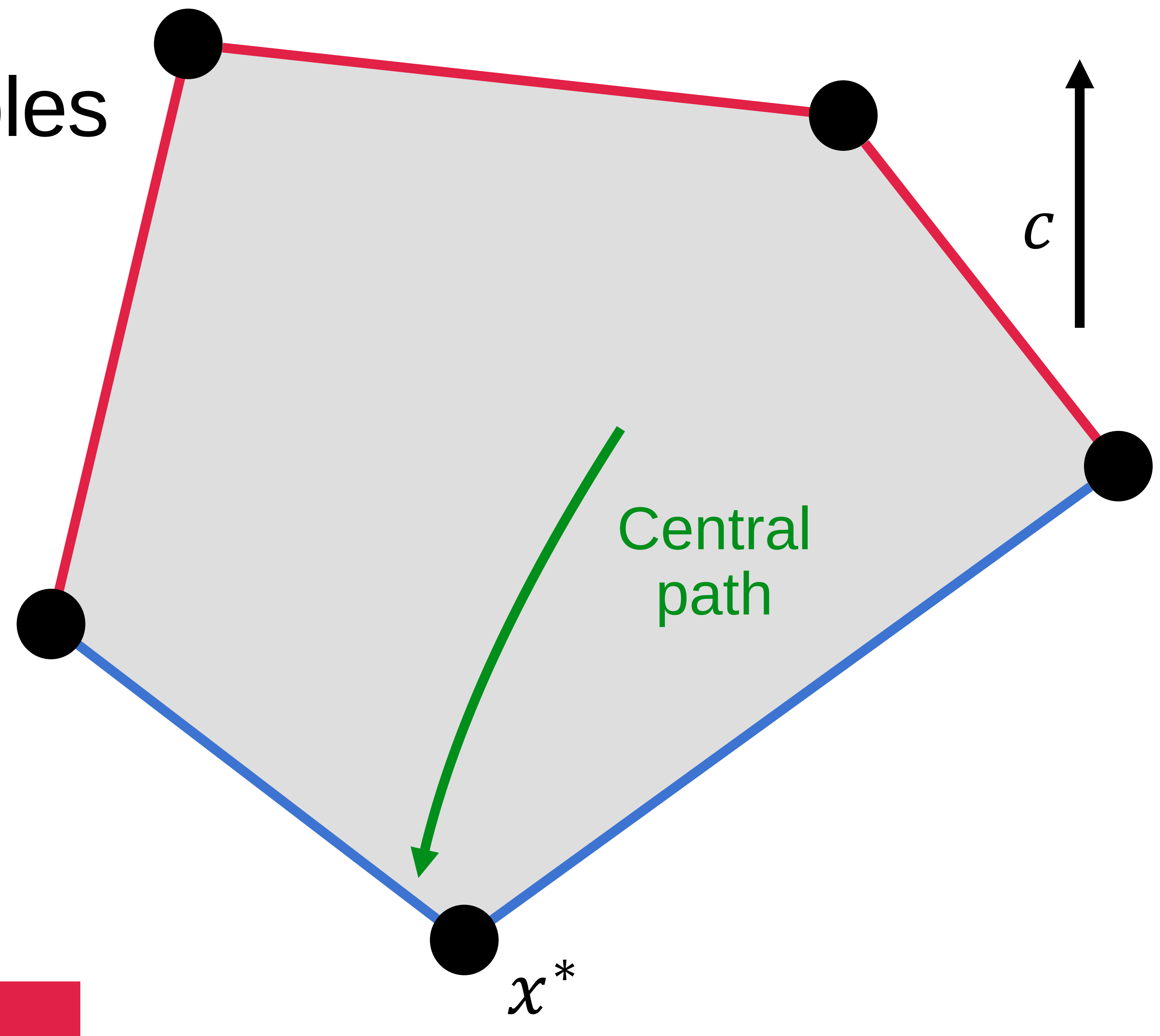
Theorem (Vavasis and Y '96):
LP of the form $min\langle c, x\rangle, Ax = b, x \geq 0$ can be solved ***exactly*** within $O(n^{3.5}\log(\bar{\chi}_A)$ many iterations, each of which requires $O(n)$ linear system solves.

Note: Number of iterations independent of bit-encoding b and c. This captures many combinatorial problems, with ***nice*** constraint matrix A but arbitrary b and c

The key technique is the departure from affine-scaling to the layered-least-squares linear system solver

Layered-Least-Squares Linear System

$min\langle c,x\rangle: Ax = b, x \geq 0$, m constrains, n variables



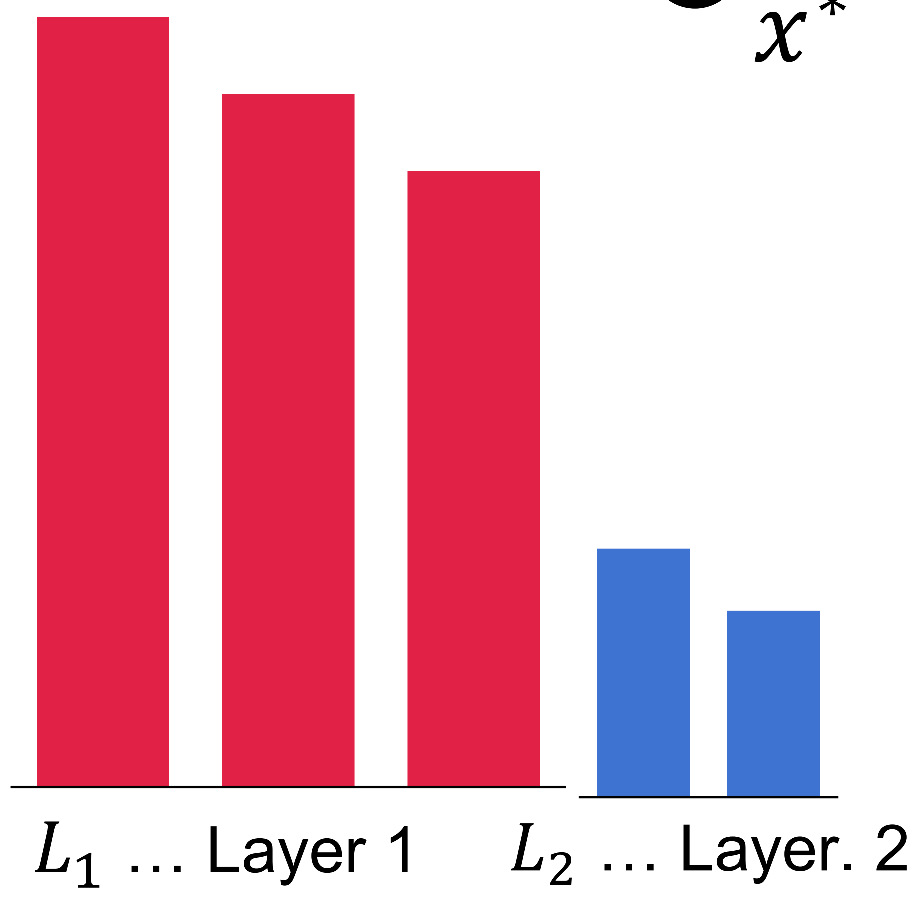
Standard affine scaling:

$$\Delta x := \text{argmin} \sum_{i=1}^{n} \left(\frac{x_i + \Delta x_i}{x_i}\right)^2 \; s.t.\; A\Delta x = 0$$

Vavasis-Y Layered-Least-Squares step:

Step 1: $z := \text{argmin}_{\Delta x} \sum_{i \in L_2} \left(\frac{x_i+\Delta x_i}{x_i}\right)^2$ s.t. $A\Delta x = 0$

Step 2: $\Delta x := \text{argmin}_{\Delta x} \sum_{i \in [n]} \left(\frac{x_i+\Delta x_i}{x_i}\right)^2 \; s.t. A\Delta x = 0, \Delta x_{L_2} = z_{L_2}$

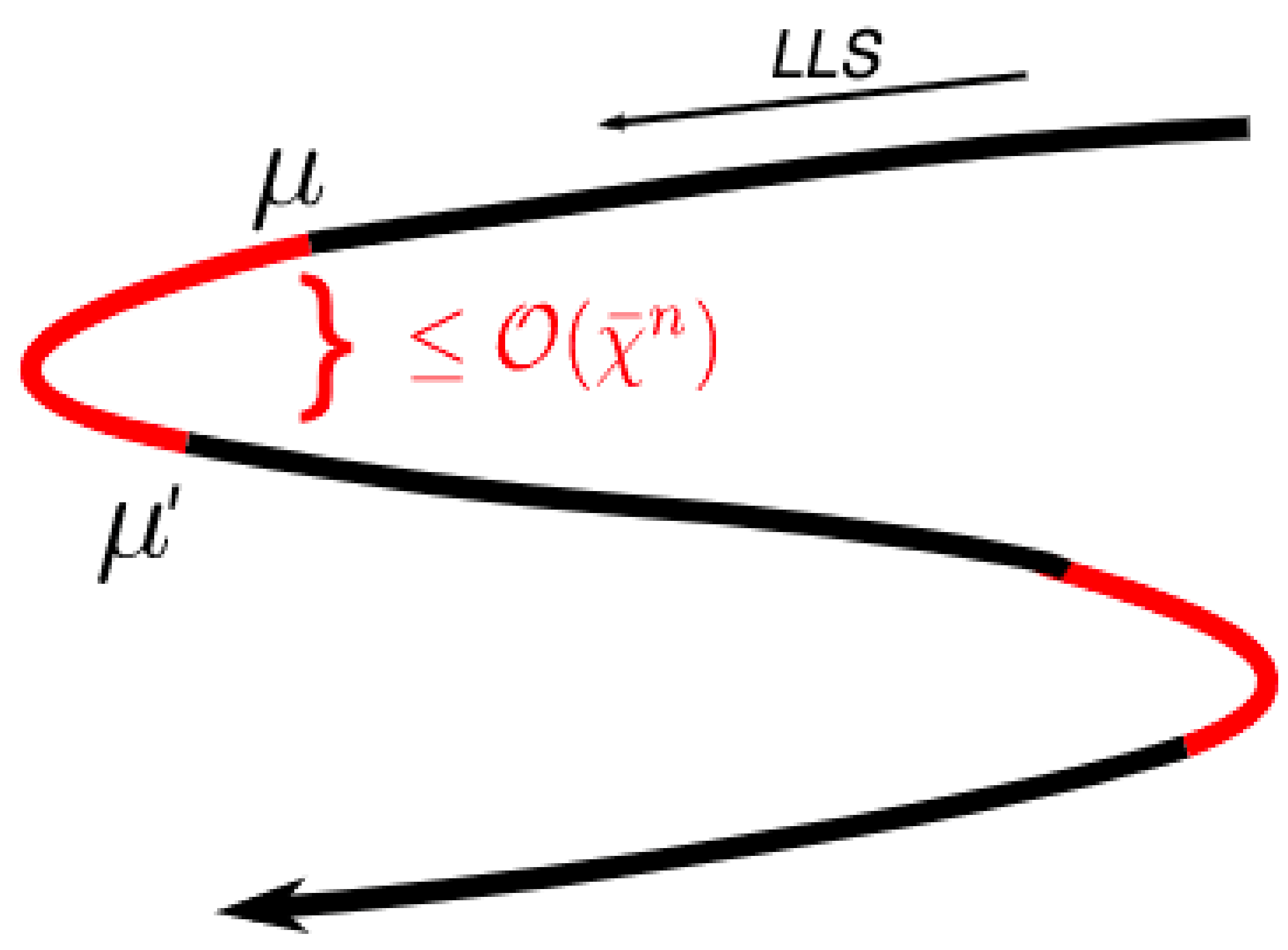


Key Idea of the V-Y Analysis

Central path consists of

- $O(n^2)$ ***short curved segments*** of length $O(\overline{\chi}_A)$ in μ
  $\Rightarrow \sqrt{n}\cdot n\log(\overline{\chi}_A)$ iterations required to traverse each.

- $O(n^2)$ long straight segments
  $\Rightarrow$ single iteration of LLS step sufficient to traverse each, even if unbounded length in μ.



The algorithm is not scale-invariant.

Scale-Invariant Improvements

There is an exact algorithm (Vavasis and YE 1996) that

- Finds an exact optimal solution in $O(n^{3.5}\log(\bar{\chi}_A + n))$ iterations. Thus, the computational complexity does not depend on b nor c.
- The algorithm is not scaling invariant.
- The number of iterations to reduce the duality gap by a factor of ε by MTY-PC algorithm [Monteiro and Tsuchiya 2004]:

$$O(n^2 \log\log(1/\varepsilon) + n^{3.5}\log(\bar{\chi}^*_A + n))$$

- Analysis is based on Vavasis-Ye framework. Compared with Vavasis-Ye algorithm, $\bar{\chi}_A$ is replaced by $\bar{\chi}^*_A$ (scaling invariance), but $n^2\log\log(1/\varepsilon)$ is there (no finite termination).

Theorem (Dadush, Huiberts, Natura, Végh '20):

LP of the form $min\langle c, x\rangle, Ax = b, x \geq 0$ can be solved ***exactly*** within $\tilde{O}(n^{2.5}\log(\bar{\chi}^*_A))$ many iterations, each of which requires $O(n)$ linear system solves,

Key Question: How ‘Curved’ is the Central Path?

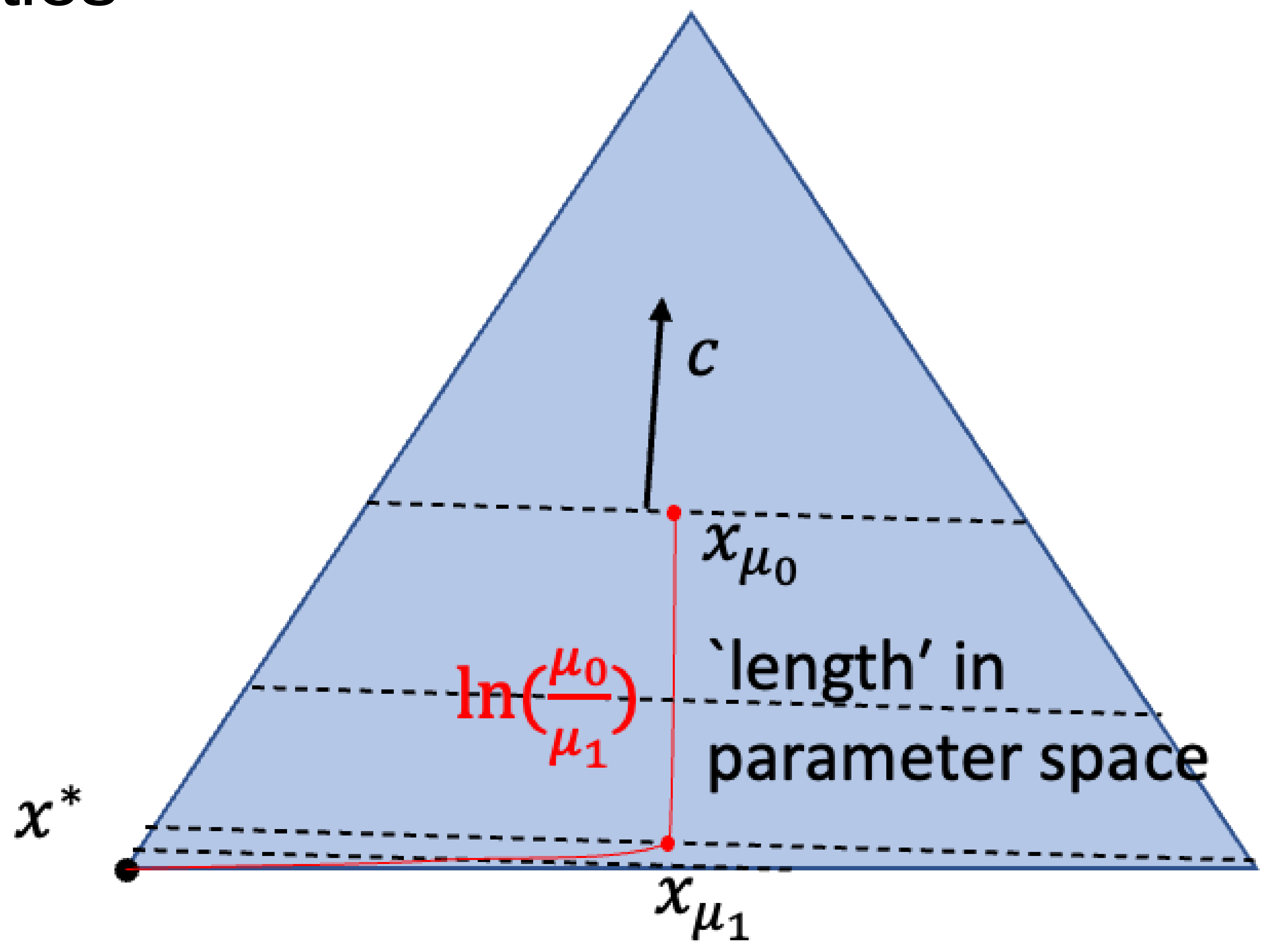
$min\langle c,x\rangle: A^\top x \geq b$, d variables, n inequalities

$$x_\mu := \operatorname{argmin}\langle c,x\rangle - \mu\sum_{i=1}^{n}\log(\langle a_i,x\rangle - b_i)$$

- Parameter μ - optimality gap
- Multiplicative decrease in μ in each iteration
- x_μ is ‘as far away as possible’ from constraints subject to having optimality gap μ

Question: How many iterations to solve an LP exactly?

Can the condition numbers be bounded Polynomially in dimensions?



As $c \to e_2$ convergence to x^* is arbitrarily slow

Path can be ***arbitrarily curved or “long”*** in parameter space

The Central Path Curvature and Iteration # I

Let $(\dot{x}, \dot{s}, \dot{y})$ be the derivative of $(x(\nu), s(\nu), y(\nu))$, which satisfies

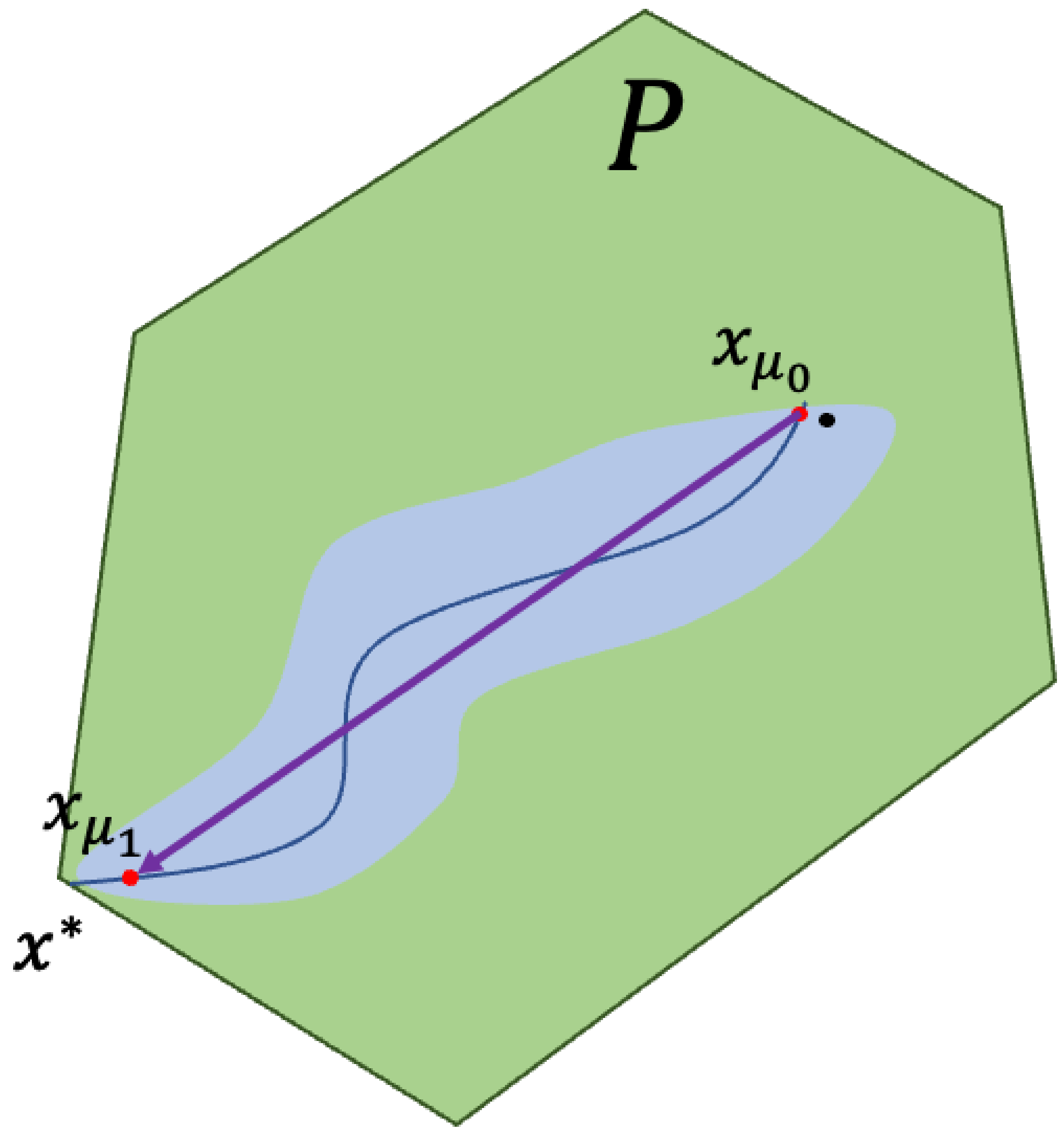
$$\dot{x} \circ s + x \circ \dot{s} = -e, \ A\dot{x} = 0, \ A^T \dot{y} + \dot{s} = 0.$$

Sonnevend Curvature [Sonnevend, Stoer, Zhao 1991]:

$$\kappa(\nu) = \sqrt{\nu |\dot{x} \circ \dot{s}|}$$

Sonnevend Curvature integral:

$$I_{PD}(\nu_{ini}, \nu_{fin}) = \int_{\nu_{ini}}^{\nu_{fin}} \frac{\kappa(\nu)}{\nu} d\nu.$$



The Central Path Curvature and Iteration # II for CP-Based IPMs

- The iteration number of IPM following $\mathcal{C}$ from ν_{ini} to ν_{fin} is approximated as follows:

$$\#\text{ of iterations} \sim \frac{I_{PD}(\nu_{ini}, \nu_{fin})}{\sqrt{\beta}}$$

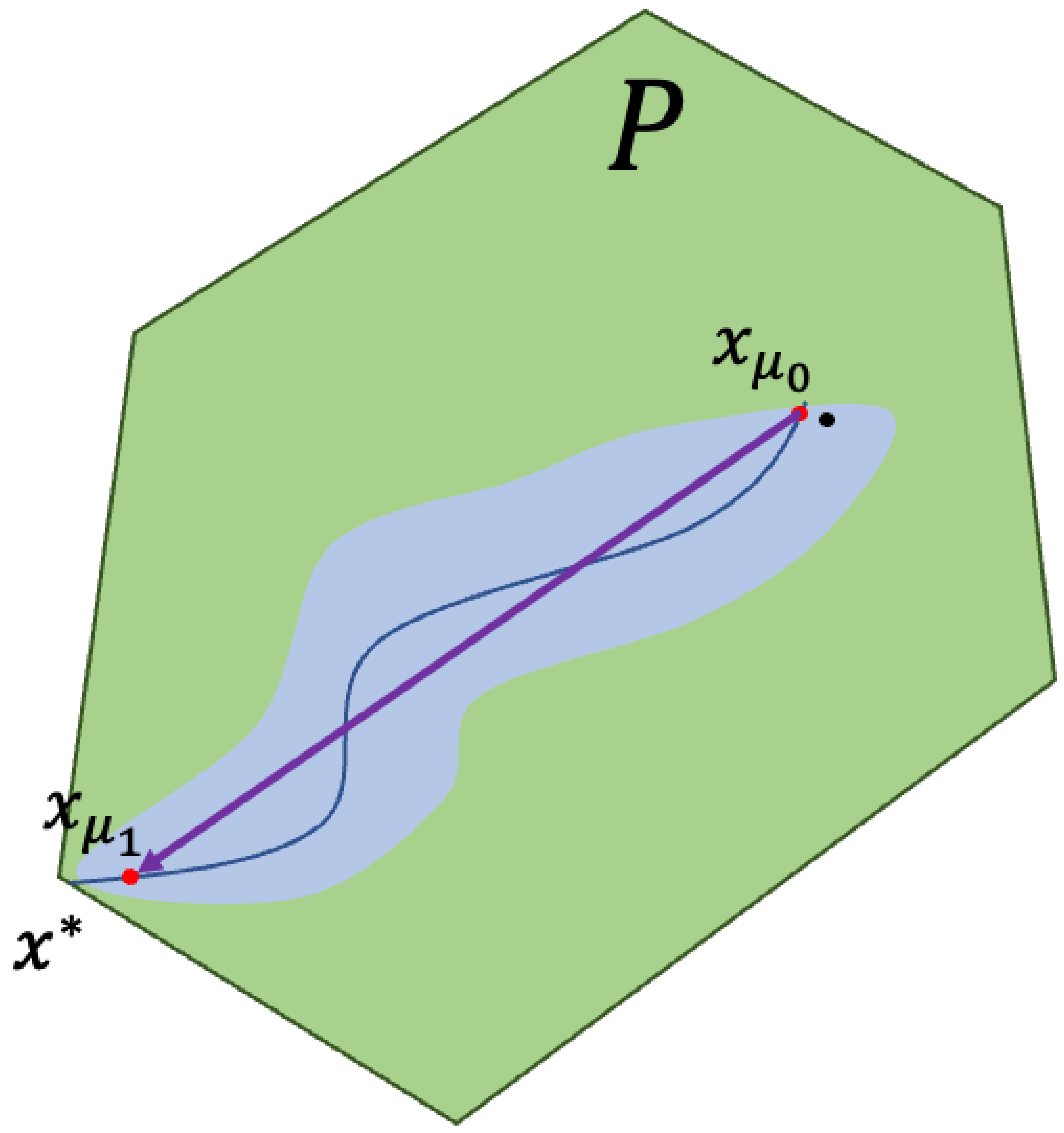
((The value of integral) $\sim$ (# of iterations when $\beta = 1$))

- By using Vavasis-Ye analysis, it can be shown that

$$I_{PD}(0, \infty) = O(n^{3.5} \log(\bar{\chi}_A^* + n)).$$

[Monteiro and Tsuchiya 2008]

- $I_{PD}(\nu_{ini}, \nu_{fin})$ is rigorously represented as differential geometric quantity [Kakihara, Ohara and Tsuchiya 2013] by using information geometry.

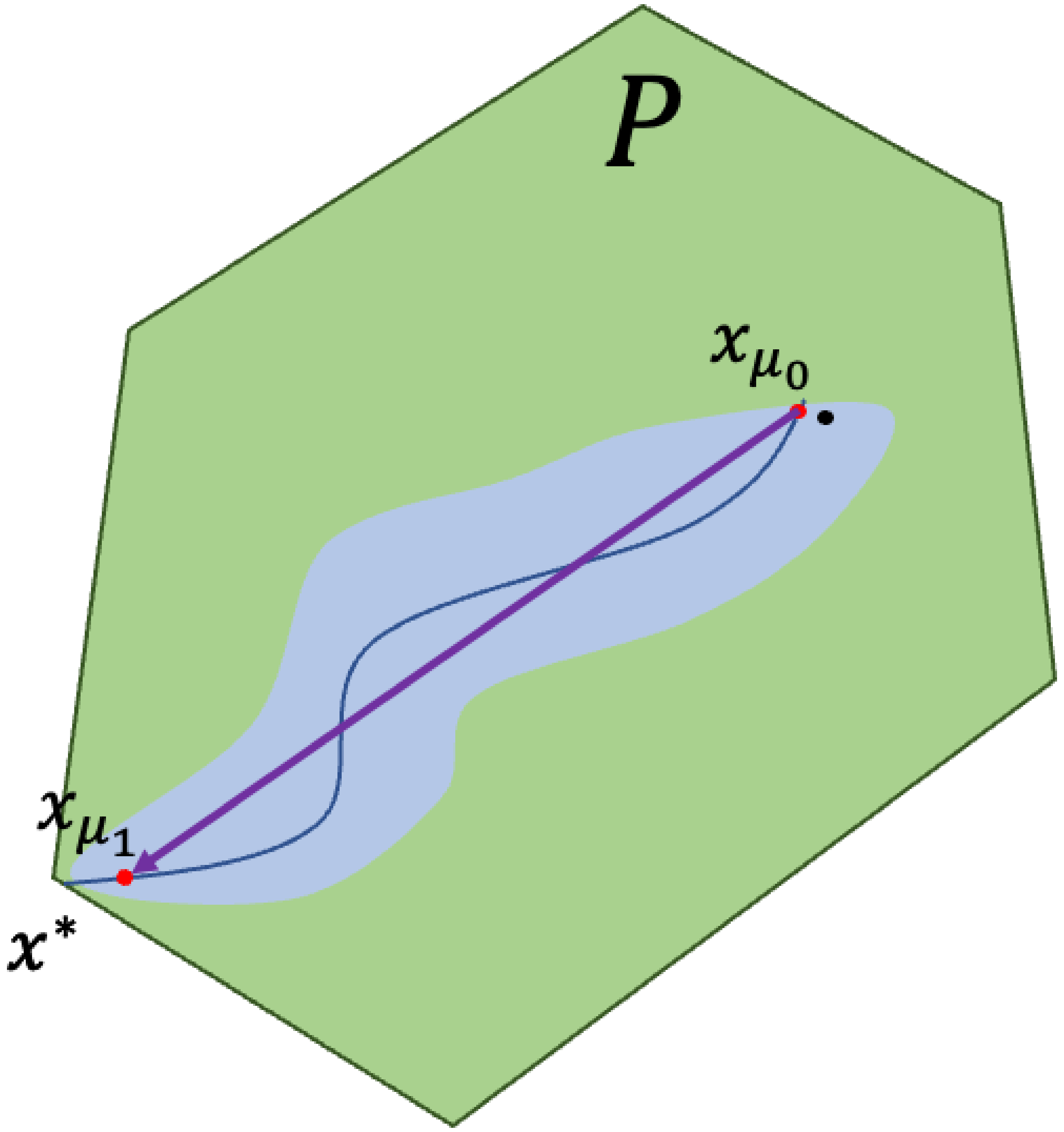


Lower Bounds I

Condition-Number vs Iteration Count

$min\langle c, x\rangle: A^{\top} x \geq b,$
d variables, n inequalities

How many iterations needed to go from parameter μ_0 to parameter μ_1?



Lower bound:
min #pieces of any piecewise linear curve from x_{μ_0} to x_{μ_1} that stays inside some *neighborhood*.

…lower bound depends on which neighborhood we use.

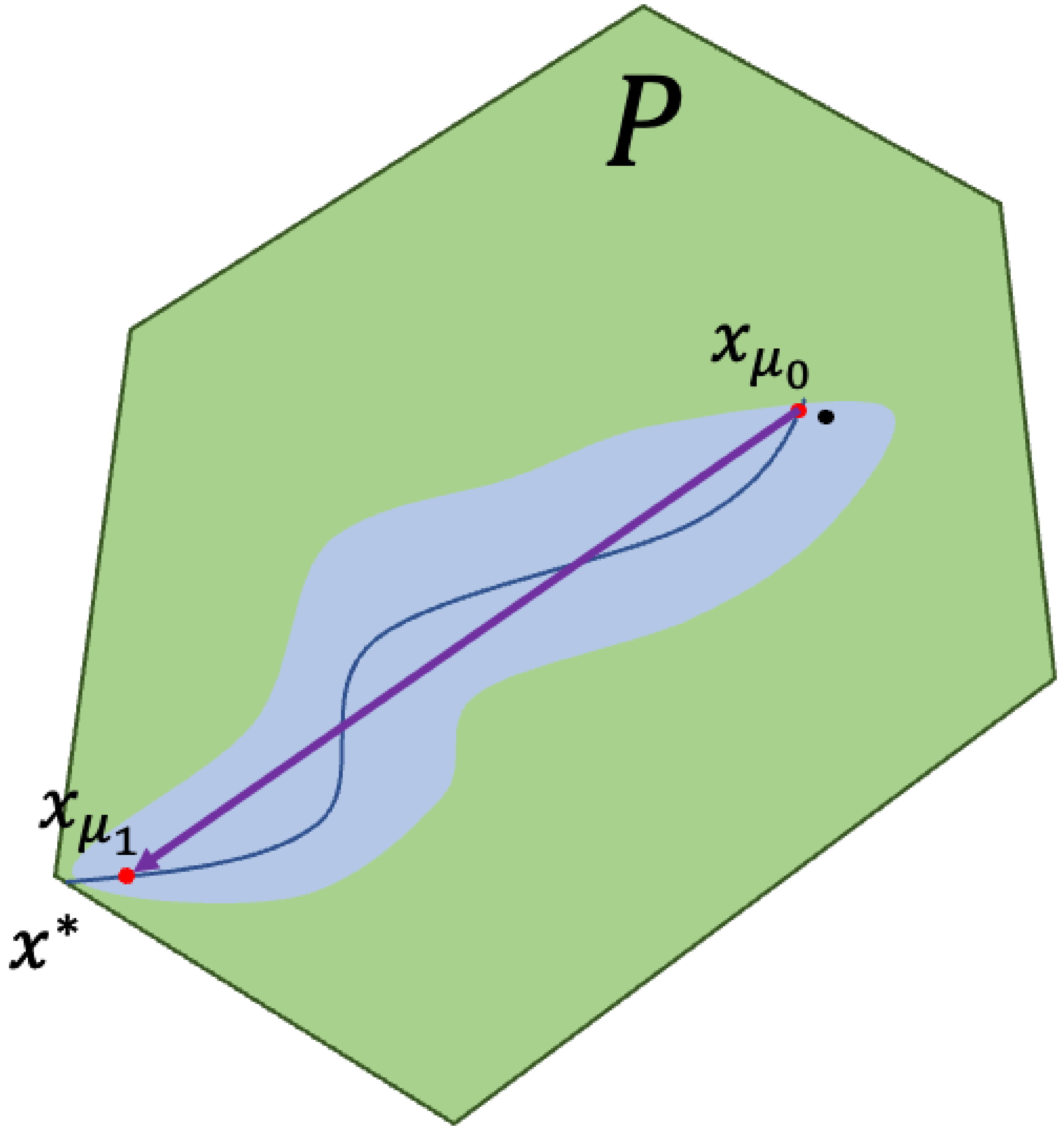
Lower Bounds II

Condition-Number vs Iteration Count

$min\langle c, x\rangle : A^\top x \geq b,$

d variables, n inequalities

- Central path can visit all (small) neighborhood of a variant of the Klee-Minty cube [Deza, Nematollahi and Terlaky 2008] ($n \sim m^3 2^{3m}$).
- Sonnevend curvature integral of a similar Klee-Minty type instance is exponential order [Mut and Terlaky 2014].
- It was an open problem to construct an instance with exponentially many sharp turns (in m) of central path with $n = O(poly(m))$. This problem is solved by using Tropical geometry [Allamigeon, Benchimol, Gaubert, Joswig 2018].



New Prospect: Straight Line IPM Complexity (SLC)

- Initial $(x_0, \mu_0) \in N$ (a neighborhood of the path).

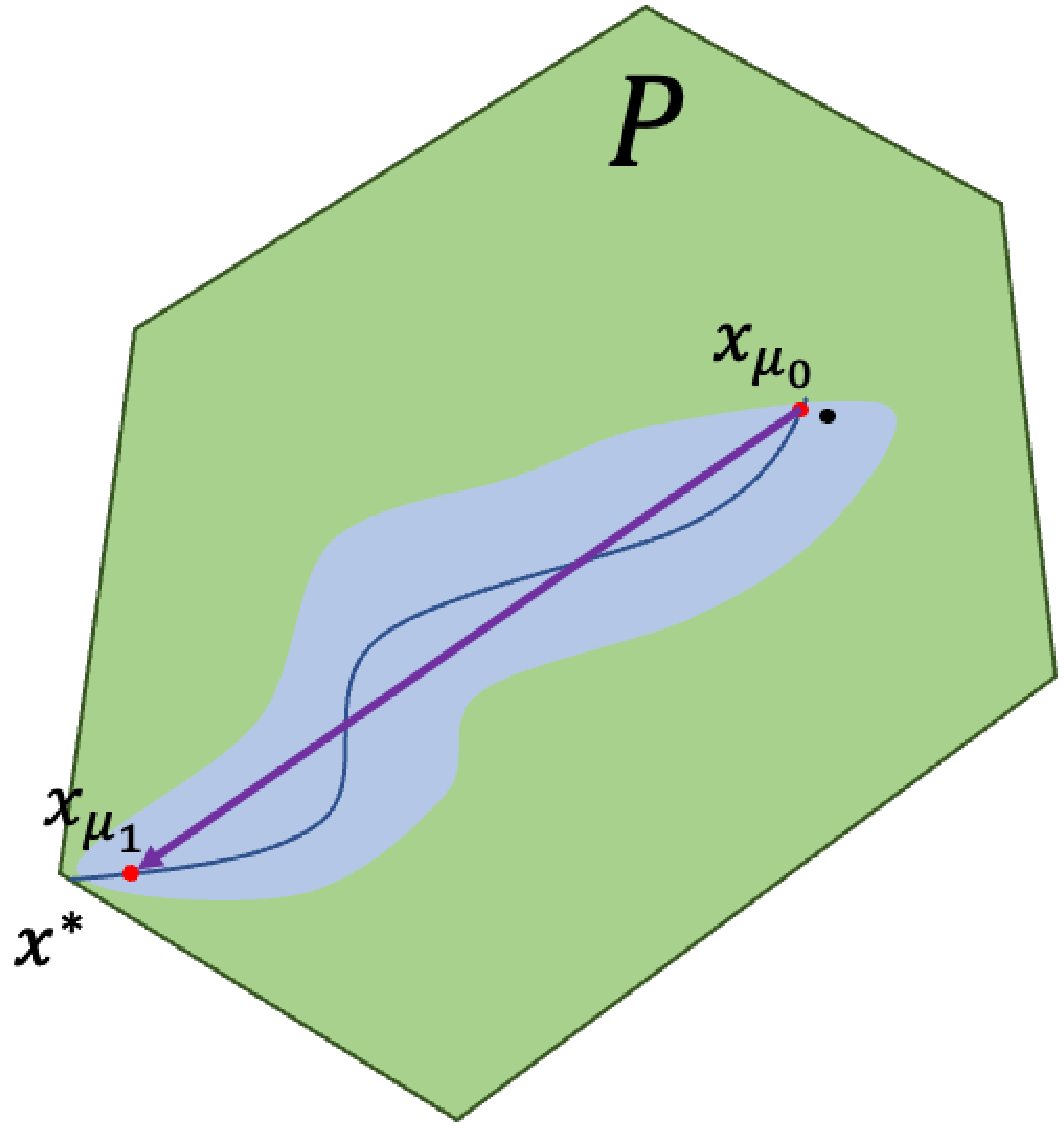
$min\langle c, x\rangle: A^\top x \geq b$,
d variables, n inequalities

Straight Line Complexity:

$\mathrm{SLC}(N, \mu_0)$:= minimum # of pieces of any piecewise-linear traversing N from (x_0, μ_0) to $(x^*, 0)$

Multiplicative neighborhood $N_\infty(\theta)$ for $\theta \in (0,1)$:

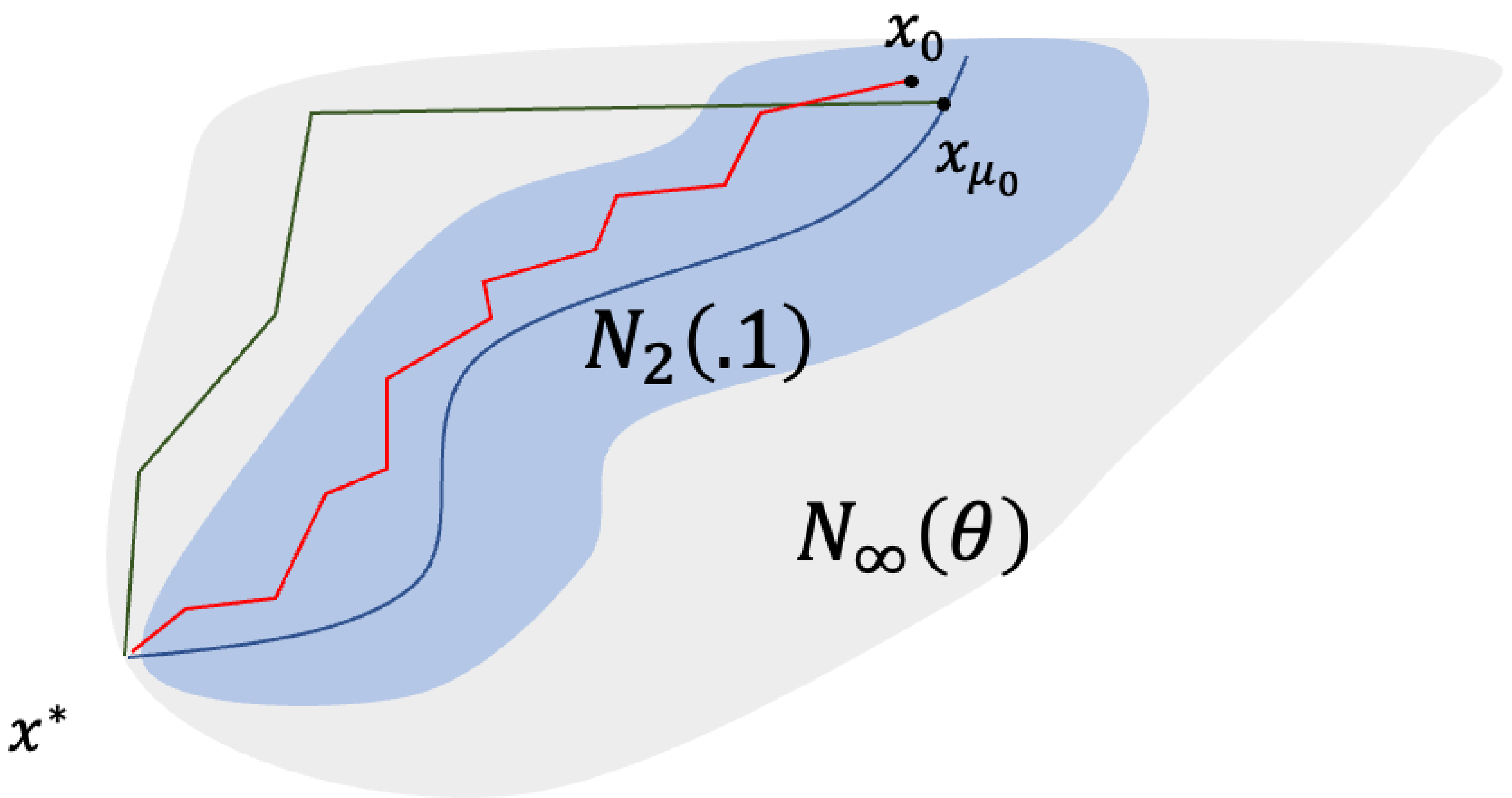
$$N_\infty(\theta) := \{(x, \mu): \frac{a_i^\top x_\mu - b_i}{a_i^\top x - b_i} \in [1-\theta, (1-\theta)^{-1}], i \in [n]\}$$



Straight Line Complexity to IPM complexity

THEOREM : (Allamigeon, Dadush, Loho, Natura, Végh '22):
Given $(x_0, \mu_0) \in N_2(0.1)$ there exists an IPM that stays in $N_2(0.1)$ and solves the LP ***exactly*** using a number of iterations that is bounded by

$$O(n^{1.5}\log(\frac{n}{1-\theta})\mathrm{SLC}(N_\infty(\theta), \mu_0)), \forall \theta \in (0,1)$$



- Every 'reasonable' IPM traverses $N_\infty(1 - 1/\mathrm{poly}(n))$
- How large can the straight line complexity $\mathrm{SLC}(N_\infty(1 - 1/\mathrm{poly}(n)), \mu_0)$ be ?
  - Vavasis-Y '96 : $O(n^{3.5}\log(n\bar{\chi}_A))$
  - Dadush, Huiberts, N., Végh '20: $O(n^{2.5}\log(n\bar{\chi}_A))$

  - vertices of the polytope, i.e. $\binom{n}{d}$

Which LPs are strongly polynomially solvable?

Strongly polynomial (known before 2022)

LP in small dimension $d = O(\log^2(n)/\mathrm{loglog}n)$

Strongly Polynomial Straight Line Complexity (SLC)

Combinatorial LP: A integral, $\| A \|_\infty = 2^{O(\mathrm{poly}(n))}$

- Shortest Path
- Bipartite Matching
- Maximum flow
- Minimum-cost flow
- Multi-commodity flow
- lattice polytopes

- Maximum generalized flow
- 2-variable-per-inequality feasibility systems
- Discounted Markov Decision Processes (MDP)
- Deterministic MDP

Klee-Minty cubes

Markov Decision Processes

Minimum-cost generalized flow

LP

LP remains an open research field…