INTERIOR POINT METHODS ARE NOT (MUCH) WORSE THAN SIMPLEX

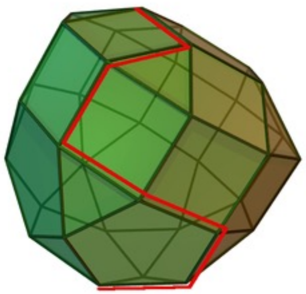
László Végh

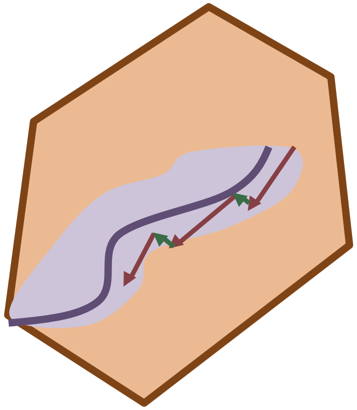
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Joint work with
Xavier Allamigeon (INRIA), Daniel Dadush (CWI), Georg Loho (Twente) and Bento Natura (LSE)

MIP 2022, Rutgers

Interior point method vs Simplex





Simplex	IPM
Dantzig, 1947	Karmarkar 1984
Fast in practice	Fast in practice
Exponential in worst case	Polynomial in the input size
Polynomial smoothed complexity	
Easy to warm start	
Numerically stable	

Complexity of LP algorithms

$$\min c^\top x$$
$$Ax = b$$
$$x \geq 0$$

- n variables, m equality constraints
- Total encoding L.
- Worst case bound for Simplex: $\binom{n}{m} \leq 2^n$
- Worst case bound for IPMs: $\text{poly}(n, L)$
  ... lots of recent improvements, see Yin Tat's talk

2^n VS $\text{poly}(n, L)$

Complexity of LP algorithms

$$\min c^\top x$$
$$Ax = b$$
$$x \geq 0$$

2^n vs $\text{poly}(n, L)$

- Some problems with polynomial encoding can be formulated as LPs with exponential entries
- Is there any function $f(n)$ and an IPM method with running time

$$\min\left(\text{poly}(n, L), f(n)\right)?$$

- Strongly polynomial algorithm: $\text{poly}(n)$ arithmetic operations

Is there a strongly polynomial algorithm for Linear Programming?

Smale's 9th question

Strongly polynomial algorithms for classes of Linear Programs

$$\min c^\top x,\ Ax = b\ \ x \geq 0$$

- Combinatorial problems: two variable per inequality systems, network flows, discounted MDPs, …
- Tardos '86: poly$(n, \log \Delta_A)$
  dependence only on A, but not on b and c.
  $$\Delta_A = \max\{|\det(B)|: B \text{ submatrix of } A\}$$
- Layered-least-squares (LLS) Interior Point Method
  Vavasis & Ye '96: poly$(n, \log \bar{\chi}_A)$ LP algorithm
  in the real model of computation
  $\bar{\chi}_A$: Dikin–Stuart–Todd condition number

Primal and dual LP

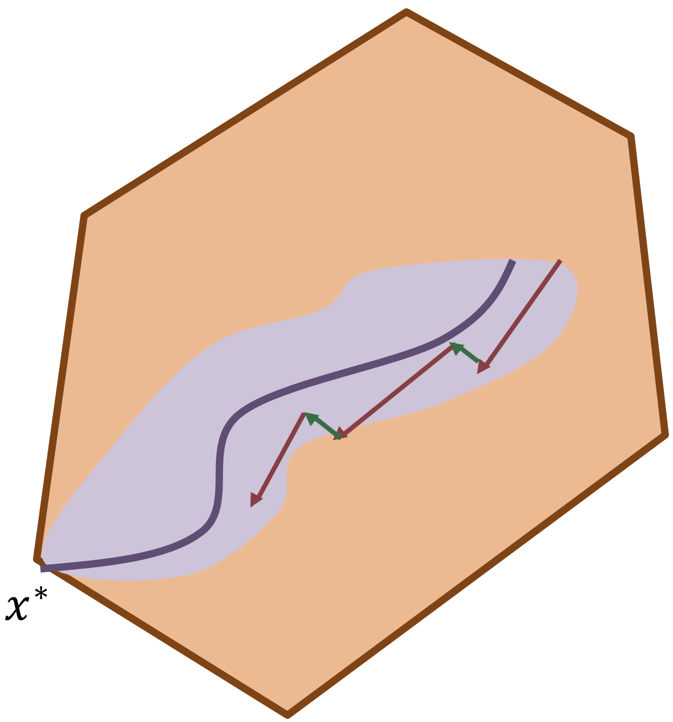
- $A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^m, b \in \mathbb{R}^n$

$$\begin{array}{c} \min c^\top x \\ Ax = b \\ x \geq 0 \end{array} \qquad \begin{array}{c} \max b^\top y \\ A^\top y + s = c \\ s \geq 0 \end{array}$$

- Complementary slackness: Primal and dual solutions (x, s) are optimal if $x^\top s = 0$:
  $x_i = 0$ or $s_i = 0$ for each $i \in [n]$.
- Optimality gap:

$$c^\top x - b^\top y = x^\top s.$$

The central path

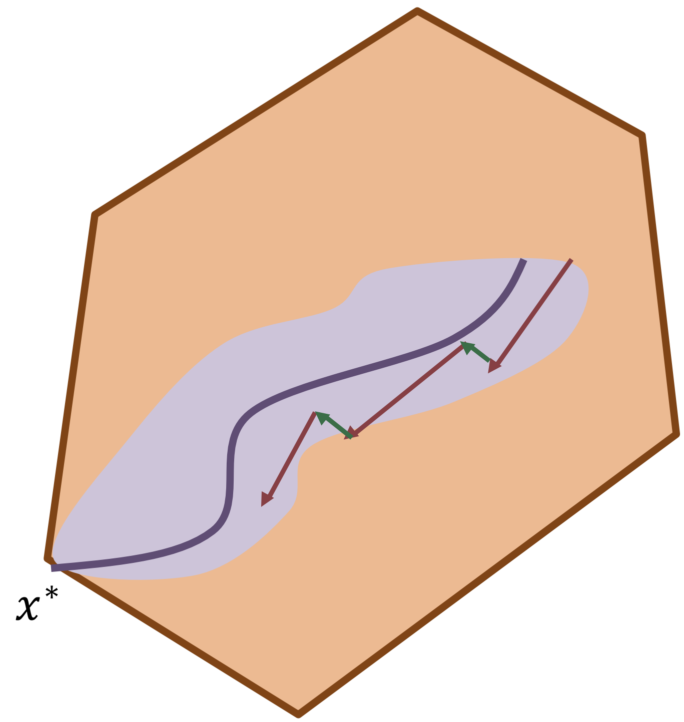


- For each $\mu > 0$, there exists a unique $z(\mu) = (x(\mu), y(\mu), s(\mu))$ such that
  $$x(\mu)_i s(\mu)_i = \mu \quad \forall i \in [n]$$
  the central path element for μ.
- The central path is the algebraic curve $\{z(\mu) : \mu > 0\}$
- For $\mu \to 0$, the limit is an optimal solution $z^* = (x^*, y^*, s^*)$.
- The duality gap is $s(\mu)^\top x(\mu) = n\mu$.
- Interior point algorithms: walk down along the central path with μ decreasing geometrically.

The Mizuno–Todd–Ye Predictor-Corrector Algorithm

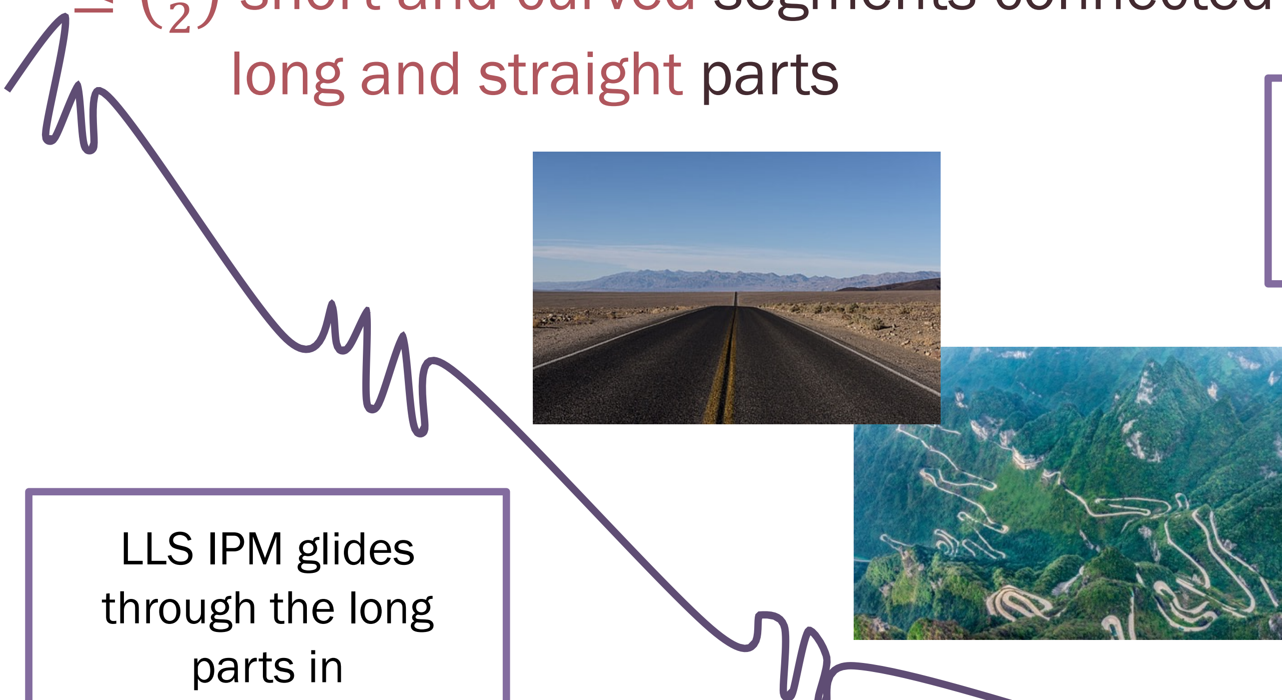
- Start from point $z_0 = (x_0, y_0, s_0)$ 'near' the central path at some $\mu_0 > 0$.
- Alternate between
  - Predictor steps: 'shoot down' the central path, decreasing μ by a factor at least $1 - 1/O(\sqrt{n})$. May move slightly 'farther' from the central path.
  - Corrector steps: do not change parameter μ, but move back 'closer' to the central path.

Within $O(\sqrt{n})$ iterations, μ decreases by a factor 2.



Layered Least Squares Interior Point Method
Vavasis-Ye '96

- $O(n^{3.5} \log \bar{\chi}_A)$ iterations
- $\bar{\chi}_A$: Dikin–Stuart–Todd condition number
- Combinatorial structure of central path:
  $\leq \binom{n}{2}$ short and curved segments connected by
  long and straight parts



Short = $O(n^{1.5} \log \bar{\chi}_A)$

LLS IPM glides through the long parts in $O(n^{1.5} \log \bar{\chi}_A)$

Scaling invariant bounds

$$\bar{\chi}_A^* = \inf\{\bar{\chi}_{AD} : D \in \mathbb{R}^{n\times n} \text{ positive diagonal}\}$$

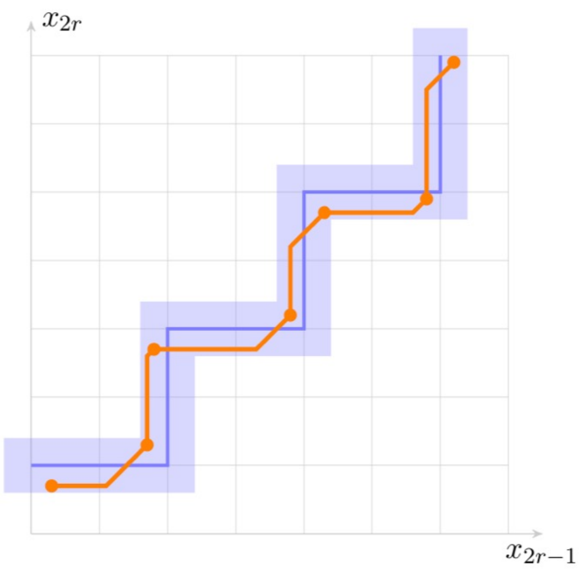
- Central path is invariant under diagonal rescalings, but $\bar{\chi}_A$ and the Vavasis–Ye algorithm are not.
- $\text{poly}(n, \log \bar{\chi}_A^*)$ algorithms
  - Predictor-Corrector Trust Region algorithm
    Lan, Monteiro & Tsuchiya '09
    computing the step directions has weakly polynomial dependence on b and c
  - Scaling invariant LLS
    Dadush, Huiberts, Natura, V. '20
    using combinatorics of circuit imbalances
- $\text{poly}(n, \log \bar{\chi}_A^*)$ bound on the Sonnevend–Stoer–Zhao curvature of the central path
  Monteiro & Tsuchiya '08

2^n VS $\text{poly}(n, \log \bar{\chi}_A^*)$

- $\bar{\chi}_A^*$ can be still unbounded
- Is $\text{poly}(n, \log \bar{\chi}_A^*)$ a tight bound on what an IPM can achieve, or…
- Could even strongly polynomial IPMs exist?

Impossibility results

- THEOREM (Allamigeon, Benchimol, Gaubert & Joswig '18): No path following method can be strongly polynomial that stays in the wide neighbourhood of the standard log barrier central path.
- Proof using **tropical geometry**: studies the tropical limit of a family of parametrized linear programs.
- Allamigeon, Gaubert & Vandame '22: extension to arbitrary self-concordant barrier functions
  IPM analogue of Klee–Minty cube
- Previous work:
  Deza, Terlaky & Zinchenko '09



$$\min\left\{ \boxed{\text{poly}(n, \log \bar{\chi}_A^*)} , \ 2^n \right\}$$

THEOREM (Allamigeon, Dadush, Loho, Natura & V. '22)
There exists a primal-dual path following method where the number of iterations is

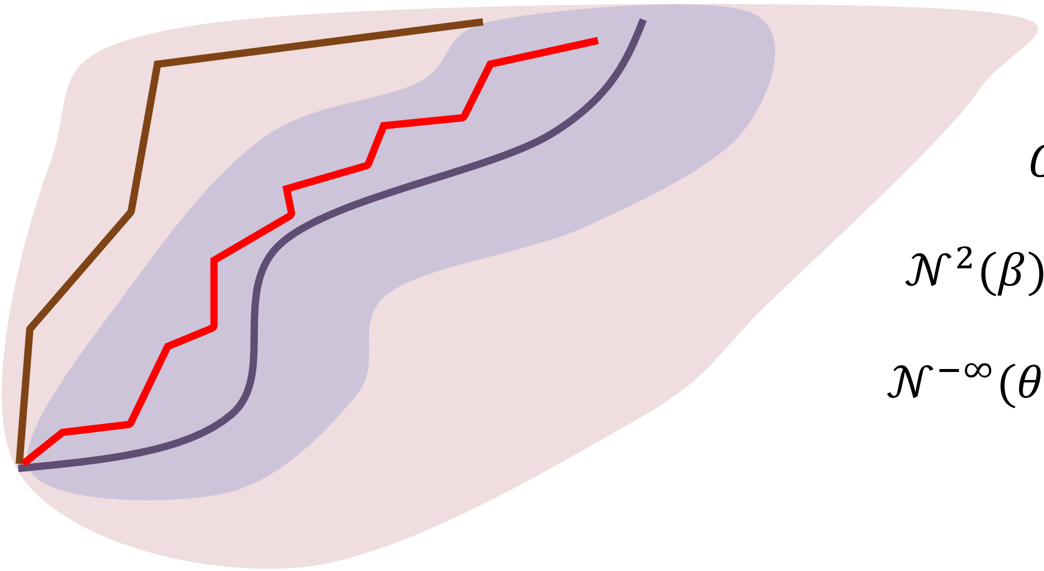
$$O(n^{1.5} \log n) \min\left\{ O(2^n), \min \begin{array}{c} \text{no. of iterations} \\ \text{of any path following method} \end{array} \right\}$$

Following a piecewise linear path in the wide neighbourhood

For any piecewise linear curve with T pieces on $(0, \mu_0)$ in the θ-wide neighbourhood, our IPM makes at most

$$O\left(n^{1.5} T \log \frac{n}{1-\theta}\right)$$

iterations is the ℓ_2-neighbourhood.



$$\bar{\mu}(z) = \frac{x^\top s}{n}$$

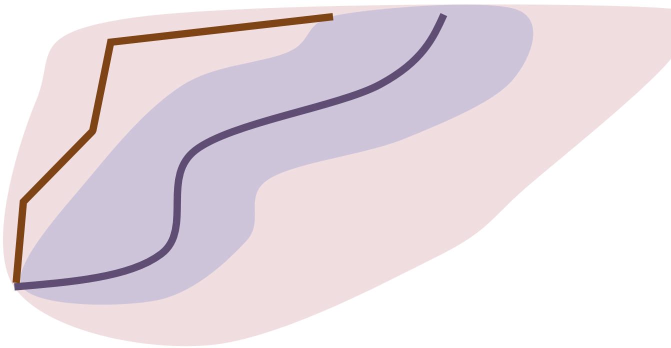
$$CP = \left\{ z = (x, y, s): \frac{xs}{\bar{\mu}(z)} = \mathbf{1} \right\}$$

$$\mathcal{N}^2(\beta) = \left\{ z = (x, y, s): \left\| \frac{xs}{\bar{\mu}(z)} - \mathbf{1} \right\| \leq \beta \right\}$$

$$\mathcal{N}^{-\infty}(\theta) = \left\{ z = (x, y, s): \frac{xs}{\bar{\mu}(z)} \geq (1-\theta)\mathbf{1} \right\}$$

The Max Central Path

GOAL: Show the existence of a piecewise linear curve with $\leq 2^n$ pieces in the wide neighbourhood



Max Central Path

$\bar{x}(\mu) = \big(\bar{x}_1(\mu), \bar{x}_2(\mu), \ldots, \bar{x}_n(\mu)\big),$

$\bar{s}(\mu) = (\bar{s}_1(\mu), \bar{s}_2(\mu), \ldots, \bar{s}_n(\mu))$

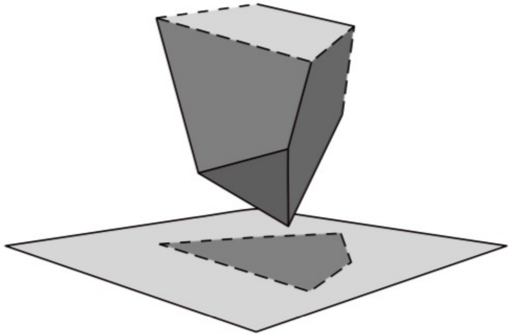
$$\begin{aligned} \bar{x}_i &= \max x_i \\ Ax &= b \\ x &\geq 0 \\ c^\top x &\leq OPT + n\mu \end{aligned} \qquad\qquad \begin{aligned} \bar{s}_i &= \max s_i \\ A^\top y + s &= c \\ s &\geq 0 \\ b^\top y &\geq OPT - n\mu \end{aligned}$$

LEMMA: $\frac{\bar{x}(\mu)}{2n} \leq x(\mu) \leq \bar{x}(\mu)$ and $\frac{\bar{s}(\mu)}{2n} \leq s(\mu) \leq \bar{s}(\mu)$

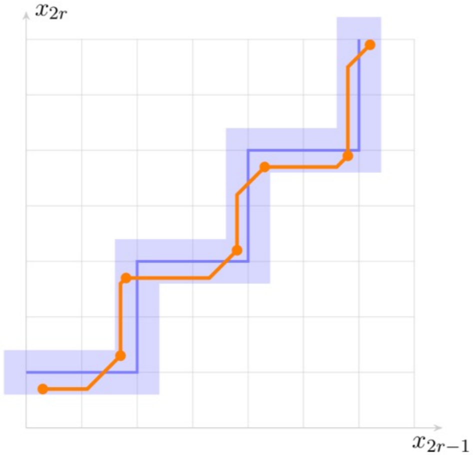
$\bar{x}(\mu)$ and $\bar{s}(\mu)$ not primal & dual feasible, but averaging the corresponding solutions gives a PL curve in the wide neighbourhood

The Max Central Path

- The Max Central Path has $O(2^n)$ segments
- It is also bounded by the total length of $2n$ shadow vertex simplex paths
- Shadow Vertex Simplex:
  average case analysis Borgwardt '87
  smoothed complexity Spielman & Teng '04



- The Max Central Path is also related to the tropical central path used in the lower bounds

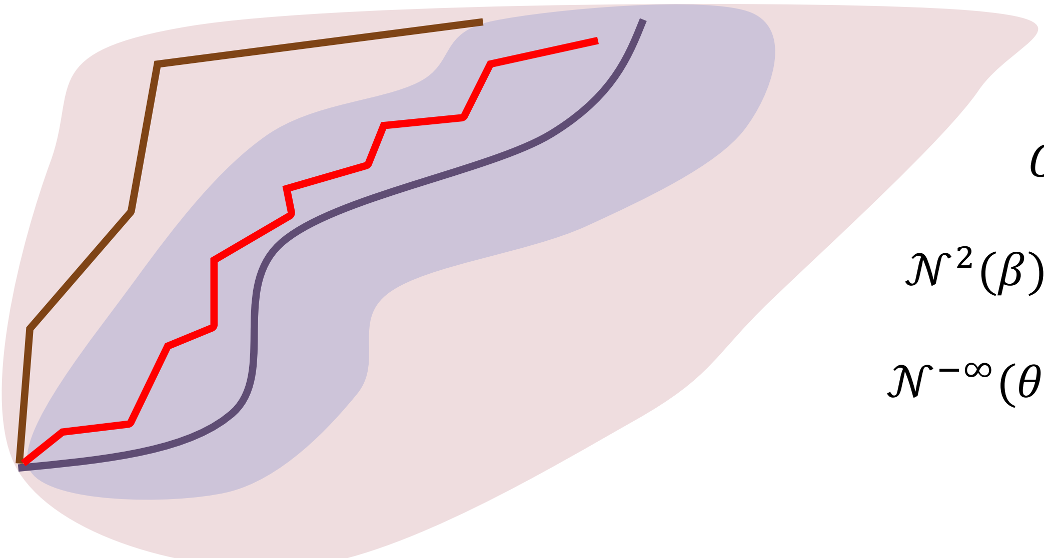


Following a piecewise linear path in the wide neighbourhood

For any piecewise linear curve with T pieces on $(0, \mu_0)$ in the θ-wide neighbourhood, our IPM makes at most

$$O\left(n^{1.5} T \log \frac{n}{1-\theta}\right)$$

iterations is the ℓ_2-neighbourhood.



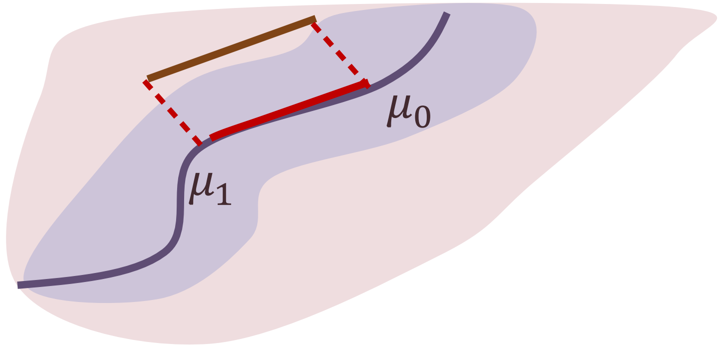
$$\bar{\mu}(z) = \frac{x^\top s}{n}$$

$$CP = \left\{z = (x, y, s): \frac{xs}{\bar{\mu}(z)} = \mathbf{1}\right\}$$

$$\mathcal{N}^2(\beta) = \left\{z = (x, y, s): \left\|\frac{xs}{\bar{\mu}(z)} - \mathbf{1}\right\| \leq \beta\right\}$$

$$\mathcal{N}^{-\infty}(\theta) = \left\{z = (x, y, s): \frac{xs}{\bar{\mu}(z)} \geq (1-\theta)\mathbf{1}\right\}$$

Polarization of central path segments



If there is a linear segment in the wide neighbourhood of the central path segment

$$CP[\mu_1, \mu_0] = \{x(\mu): \mu_1 \leq \mu \leq \mu_0\},$$

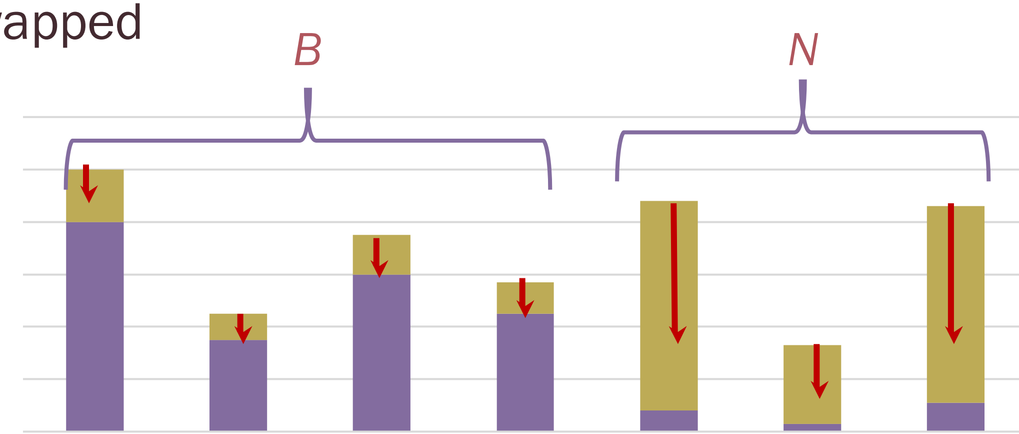
then this segment is polarized

$$\exists B \cup N = \{1,2, \dots, n\}$$

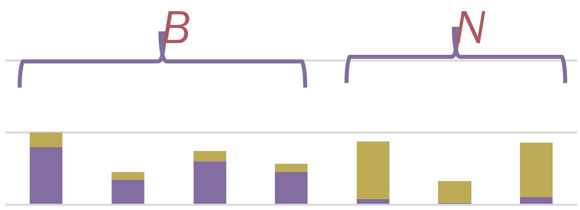
$$\gamma x_i(\mu_0) \leq x(\mu) \leq n x_i(\mu_0) \;\; \forall i \in B$$

$$\frac{1}{n}\frac{\mu}{\mu_0} x_i(\mu_0) \leq x(\mu) \leq \frac{1}{\gamma}\frac{\mu}{\mu_0} x_i(\mu_0) \forall i \in N$$

same for $s(\mu)$ with B and N swapped



Subspace Layered Least Squares IPM



New LLS step direction that can traverse any polarized segment $CP[\mu_1, \mu_0]$ in $O(n^{1.5} \log n)$ iterations, no matter the length

STEP 1: guess the polarizing partition $B \cup N$

$(\Delta x^a, \Delta s^a)$: Standard affine scaling step

$$\tilde{B} := \left\{ i: \left| \frac{\Delta x_i^a}{x_i} \right| < \left| \frac{\Delta s_i^a}{s_i} \right| \right\}, \tilde{N} := [n] \setminus \tilde{B}$$

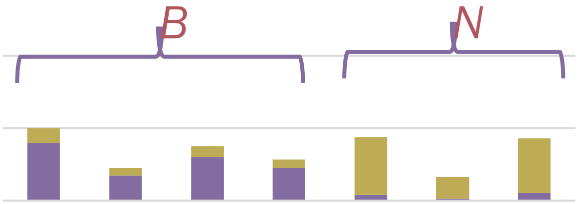
LEMMA (Roughly): If we are still far from the end of the polarized segment, this reveals the polarizing partition:

$$\tilde{B} = B, \qquad \tilde{N} = N$$

Predictor-Corrector Trust Region algorithm
Lan, Monteiro & Tsuchiya '09

Given (B, N), select the primal direction $\Delta x \in \ker(A)$ such that

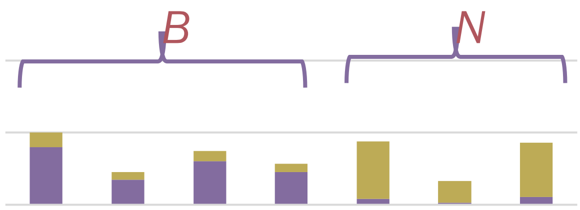
- make the most progress in decreasing variables in N, while
- barely change the variables in B



Analogously for $\Delta s \in \mathrm{im}(A^\top)$

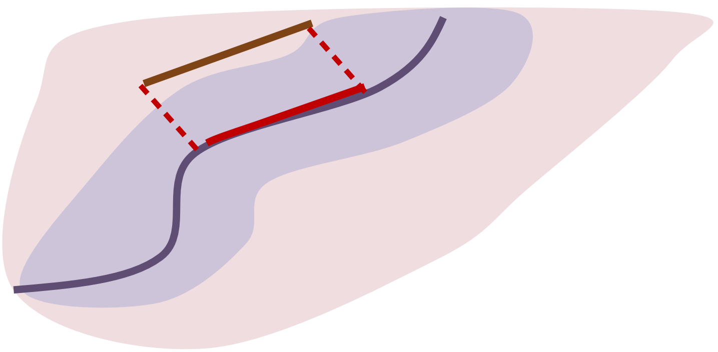
PROBLEM: cannot compute in strongly polynomial time

Subspace Layered Least Squares IPM

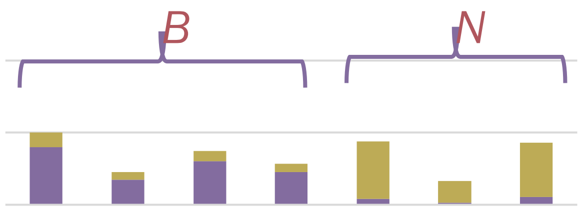


Given (B, N), select the primal direction $\Delta x \in \ker(A)$ such that

- make the most progress in decreasing variables in N,
  restricting $\Delta x_N \in V$ for a subspace $V \subseteq \mathrm{Proj}_N(\ker(A))$
- V is chosen such that any $\Delta x_N \in V$ can be extended to $(\Delta x_B, \Delta x_N) \in \ker(A)$ with small Δx_B
- V obtained using an (approximate) singular value decomposition



Analysis

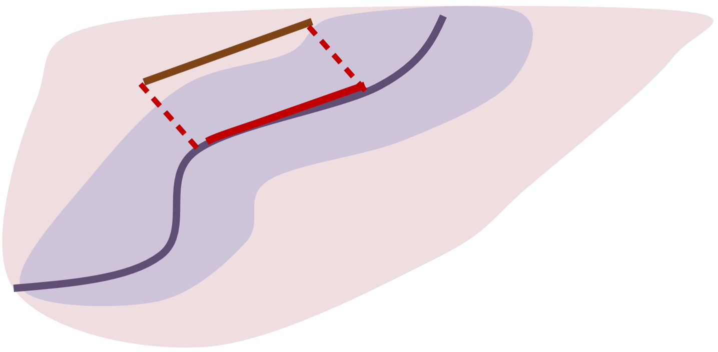


Focus on subspaces $V \subseteq \mathrm{Proj}_{\mathrm{N}}(\ker(A))$ and $U \subseteq \mathrm{Proj}_{\mathrm{B}}(\mathrm{im}(A^{\top}))$ used in the algorithm.

In $O(\sqrt{n}\log n)$ iterations

- We reach the end of the polarized segment, or
- $\dim(V)$ and $\dim(U)$ increase

Can happen at most n times



$$\min\left\{ \boxed{\mathrm{poly}(n, \log \bar{\chi}^*_A)} , \quad 2^n \right\}$$

THEOREM (Allamigeon, Dadush, Loho, Natura & V. '22)
There exists a primal-dual path following method where the number of iterations is

$$O(n^{1.5} \log n) \min\left\{ O(2^n), \min \begin{array}{c}\text{no. of iterations}\\ \text{of any path following method}\end{array} \right\}$$

The same holds for the Lan–Monteiro–Tsuchiya trust region algorithm.

Summary and open questions

- Have IPM beaten Simplex yet? Not quite:
  - Subexponential randomized simplex: Kalai '92, Matoušek, Sharir & Welzl '92
  - IPM has exponential lower bounds
- The number of iterations of our algorithm is universally the best for any log barrier IPM—up to a factor $O(n^{1.5}\log n)$

QUESTIONS

- Improving on $\bar{\chi}_A^*$, can we find a tighter condition number that is always bounded as $O(2^n)$?
- Does our algorithm obtain new strongly polynomial algorithms for some problems?
- Find a tighter analysis of our algorithm
- Find such "universal" IPM for any self-concordant barrier
- Develop a strongly polynomial variant of the trust region step