Second Order Optimization Algorithms II: Homotopy/Path-Following Algorithms

Yinyu Ye

MS&E and ICME, Stanford University

http://www.stanford.edu/~yyye

Chapter 5.4-7, 6.6

Would Convexity Help SOM?

Before we answer this question, let's summarize a generic form one iteration of the Second Order Method for solving $\nabla f(\mathbf{x}) = \mathbf{g}(\mathbf{x}) = \mathbf{0}$:

$$(\nabla \mathbf{g}(\mathbf{x}^k) + \mu I)(\mathbf{x} - \mathbf{x}^k) = -\gamma \mathbf{g}(\mathbf{x}^k), \quad \text{or}$$

$$\mathbf{g}(\mathbf{x}^k) + \nabla \mathbf{g}(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \mu(\mathbf{x} - \mathbf{x}^k) = (1 - \gamma)\mathbf{g}(\mathbf{x}^k).$$

Many interpretations: when

$\frac{1}{\varepsilon^{0.5}}$ $\log(\frac{1}{\varepsilon})$ $O(\frac{1}{\varepsilon^{1.5}})$

- $\gamma = 1, \mu = 0$: pure Newton;
- γ and μ are sufficiently large: SDM;
- $\gamma = 1$ and μ decreases to 0: Homotopy or path-following method.

min $f(x)$

strictly convex | $\log(\frac{1}{\varepsilon}) \cdot \frac{1}{\sqrt{\varepsilon}}$ | convex $\log(\frac{1}{\varepsilon})$

$H \succ 0$

$\nabla^2 f(x) \succeq 0$

$\succ 0$

$x(x^0)$ $\quad x(\mu^1)$ $\quad x(\mu^2)$

A Path-Following Algorithm for Unconstrained Optimization I

min $\quad \mu^0 \uparrow$ $\quad \tilde{x}(\mu^0)$ | $\mu^1 < \mu^0$ $\quad \tilde{x}(\mu^1)$ Newton step

For any $\mu > 0$ consider the (unique) optimal solution $\mathbf{x}(\mu)$ for problem

$x(0) = x^*$

$$\mathbf{x}(\mu) = \arg\min \; f(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x}\|^2,$$

and they form a path down to $\mathbf{x}(0)$ and satisfy gradient equations with parameter μ:

$x(\mu) \to ?$ $\quad \mu \to 0^+$

$$\boxed{\mathbf{g}(\mathbf{x}) + \mu\mathbf{x} = \mathbf{0},} \quad \text{with} \quad \mu = \mu^k > 0. \qquad (1)$$

$\mu^0 \uparrow$ $\quad \mu \to 0^+$

Let the approximation path error at $\mathbf{x}^k$ with $\mu = \mu^k$ be

$x = \frac{1}{2}x^TQx + c^Tx$

$x(\mu)$ $\quad \mu^k \to \mu^{k+1}$ $\quad \mu \to 0^+$

$$\|\mathbf{g}(\mathbf{x}^k) + \mu^k\mathbf{x}^k\| \le \frac{1}{2\beta}\mu^k.$$

$Q \succeq 0$ $\quad Qx + c = 0$ $\quad Q \not\succeq 0$

Then, we like to compute a new iterate $\mathbf{x}^{k+1}$, using Newton's method with $\mathbf{x}^k$ as an initial solution, such that

$$\|\mathbf{g}(\mathbf{x}^{k+1}) + \mu^{k+1}\mathbf{x}^{k+1}\| \le \frac{1}{2\beta}\mu^{k+1}, \quad \text{where } 0 \le \mu^{k+1} < \mu^k.$$

If μ^k can be decreased at a geometric rate, independent of ϵ, and each update uses one Newton step, then this would lead to a linearly convergent algorithm.

Concordant Lipschitz Functions

We analyze the path-following algorithm when f is convex and meet a Concordant Lipschitz condition: for any point $\mathbf{x}$ and a $\beta \geq 1$

$$\|\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\| \leq \beta \mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d}, \text{ whenever } \|\mathbf{d}\| \leq O(1) < 1 \tag{2}$$

and $\mathbf{x}+\mathbf{d}$ in the function domain. Such condition can be verified using Taylor Expansion Series; basically, the third derivative of the function is bounded by its second derivative.

- All quadratic functions are concordant Lipschitz with $\beta = 0$.
- Convex function e^x is concordant Lipschitz with $\beta = O(1)$ but it is not regular Lipschitz.
- Convex function $-\log(x)$ is neither regular Lipschitz nor concordant Lipschitz.
- Function $f(\mathbf{x}) := \phi(A\mathbf{x} - \mathbf{b})$ is concordant Lipschitz if $\phi(\cdot)$ is regular Lipschitz and strictly convex.

A Path-Following Algorithm for Unconstrained Optimization II

x^k

When μ^k is replaced by μ^{k+1}, say $(1-\eta)\mu^k$ for some $\eta \in (0,\ 1]$, we aim to find a solution $\mathbf{x}$ such that

$$\mathbf{g}(\mathbf{x}) + (1-\eta)\mu^k\mathbf{x} = \mathbf{0},$$

we start from $\mathbf{x}^k$ and apply the Newton iteration:

$$\begin{array}{c} \mathbf{g}(\mathbf{x}^k) + \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d} + (1-\eta)\mu^k(\mathbf{x}^k + \mathbf{d}) = \mathbf{0}, \quad \text{or} \\ \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d} + (1-\eta)\mu^k\mathbf{d} = -\mathbf{g}(\mathbf{x}^k) - (1-\eta)\mu^k\mathbf{x}^k. \end{array} \tag{3}$$

$x^{k+1} = x + d$

From the second expression, we have

$$\begin{aligned} \|\nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d} + (1-\eta)\mu^k\mathbf{d}\| &= \|-\mathbf{g}(\mathbf{x}^k) - (1-\eta)\mu^k\mathbf{x}^k\| \\ &= \|-\mathbf{g}(\mathbf{x}^k) - \mu^k\mathbf{x}^k + \eta\mu^k\mathbf{x}^k\| \\ &\le \|-\mathbf{g}(\mathbf{x}^k) - \mu^k\mathbf{x}^k\| + \eta\mu^k\|\mathbf{x}^k\| \\ &\le \tfrac{1}{2\beta}\mu^k + \eta\mu^k\|\mathbf{x}^k\|. \end{aligned} \tag{4}$$

On the other hand

$$\|\nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d} + (1-\eta)\mu^k \mathbf{d}\|^2 = \|\nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}\|^2 + 2(1-\eta)\mu^k \mathbf{d}^T \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d} + ((1-\eta)\mu^k)^2\|\mathbf{d}\|^2.$$

From convexity, $\mathbf{d}^T \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d} \ge 0$, together with (4) we have

$$\begin{array}{rcl} ((1-\eta)\mu^k)^2\|\mathbf{d}\|^2 & \le & (\frac{1}{2\beta} + \eta\|\mathbf{x}^k\|)^2(\mu^k)^2 \quad \text{and} \\ 2(1-\eta)\mu^k \mathbf{d}^T \|\nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d} & \le & (\frac{1}{2\beta} + \eta\|\mathbf{x}^k\|)^2(\mu^k)^2. \end{array}$$

The first inequality implies

$$\|\mathbf{d}\|^2 \le (\frac{1}{2\beta(1-\eta)} + \frac{\eta}{1-\eta}\|\mathbf{x}^k\|)^2.$$

Let the new iterate be $\mathbf{x}^+ = \mathbf{x}^k + \mathbf{d}$. The second inequality implies

x^{k+1}

$$\begin{array}{rl} & \|\mathbf{g}(\mathbf{x}^+) + (1-\eta)\mu^k \mathbf{x}^+\| \\ = & \|\mathbf{g}(\mathbf{x}^+) - (\mathbf{g}(\mathbf{x}^k) + \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}) + (\mathbf{g}(\mathbf{x}^k) + \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}) + (1-\eta)\mu^k(\mathbf{x}^k + \mathbf{d})\| \\ = & \|\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) + \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}\| \\ \le & \beta \mathbf{d}^T \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d} \le \underline{\frac{\beta}{2(1-\eta)}(\frac{1}{2\beta} + \eta\|\mathbf{x}^k\|)^2\mu^k}. \end{array}$$

We now just need to choose $\eta \in (0,\ 1)$ such that

$$
\begin{aligned}
\left(\frac{1}{2\beta(1-\eta)} + \frac{\eta}{1-\eta}\|\mathbf{x}^k\|\right)^2 &\le 1 \quad \text{and} \\
\frac{\beta\mu^k}{2(1-\eta)}\left(\frac{1}{2\beta} + \eta\|\mathbf{x}^k\|\right)^2 &\le \frac{1}{2\beta}(1-\eta)\mu^k = \frac{1}{2\beta}\mu^{k+1}.
\end{aligned}
$$

For example, given $\beta \ge 1$,

$x^0\ x^1\ x^2 \ldots x^*$

$(x_1+2x_2-1)^2$

$\min_{x_1,x_2}$

$$\eta = \frac{1}{2\beta(1+\|\mathbf{x}^k\|)}$$

$x_1 = .2$

$x_2 = 0.4$

would suffice.

This would give a linear convergence since $\|\mathbf{x}^k\|$ is typically bounded following the path to the optimality, while the convergence in non-convex case is only arithmetic.

Convexity, together with some types of second-order methods, make convex optimization solvers into practical technologies.

$c = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$ $\quad g = \begin{pmatrix} 2(x_1+2x_2+1) \\ 4(x_1+2x_2+1) \end{pmatrix}$ $\quad \min_{x_1,x_2}$ $\quad (x_1+2x_2+1)^2$ $\quad -.2 \quad -.4$

$H = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$

A Path-Following Algorithm for Unconstrained Optimization III

More question related to the path-following algorithm:

- For convex case, since $\mathbf{x}(\mu)$ is the unique minimizer of

$$\min \; f(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x}\|^2,$$

  what is the limit of $\mathbf{x}(\mu)$ as $\mu \to 0^+$?

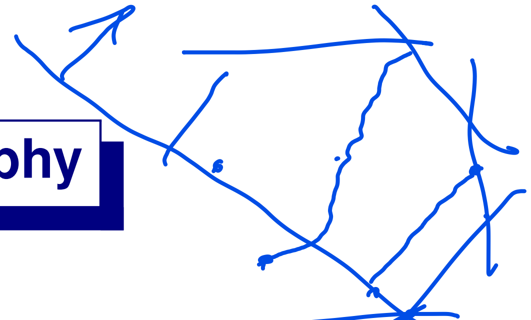
- More practical strategy to decrease μ?

- Apply first-order or 1.5-order algorithms for solving each step of the path-following, since it is to minimize a strictly convex quadratic function?

- What happen when f is bounded from below but not convex, and just meet the standard Lipschitz condition? The key is analyzing $\mathbf{x}(\mu)$, which may form multiple paths. Then can we still follow the path?

  (QPpath.m of Chapter 8)

Linear Programming Methodological Philosophy



Optimality Conditions: (1) Primal Feasibility, (2) Dual Feasibility, (3) Zero-Duality Gap/Prima-Dual Complementarity.

Recall that the (primal) Simplex Algorithm maintains the primal feasibility and complementarity while working toward dual feasibility. (The Dual Simplex Algorithm maintains dual feasibility and complementarity while working toward primal feasibility.)

In contrast, interior-point methods will move in the interior of the feasible region, hoping to by-pass many corner points on the boundary of the region. The primal-dual interior-point method maintains both primal and dual feasibility while working toward complementarity.

The key for the simplex method is to make computer see corner points; and the key for interior-point methods is to stay in the interior of the feasible region.

Interior-Point Algorithms for LP

$$(LP)\ \min\ \mathbf{c}^T\mathbf{x}\ \text{s.t.}\ A\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq\mathbf{0}\quad <=>\quad (LD)\ \max\ \mathbf{b}^T\mathbf{y}\ \text{s.t.} A^T\mathbf{y}+\mathbf{s}=\mathbf{c},\ \mathbf{s}\geq\mathbf{0}.$$

$$\text{int}\,\mathcal{F}_p=\{\mathbf{x}:\ A\mathbf{x}=\mathbf{b},\ \mathbf{x}>\mathbf{0}\}\neq\emptyset$$

x^k

$$\text{int}\,\mathcal{F}_d=\{(\mathbf{y},\mathbf{s}):\ \mathbf{s}=\mathbf{c}-A^T\mathbf{y}>\mathbf{0}\}\neq\emptyset.$$

(y^k, s^k)

Let z^* denote the optimal value and

$$\mathcal{F}=\mathcal{F}_p\times\mathcal{F}_d.$$

We are interested in finding an ϵ-approximate solution for the LP problem:

$$\mathbf{x}^T\mathbf{s}=\mathbf{c}^T\mathbf{x}-\mathbf{b}^T\mathbf{y}\leq\epsilon.$$

For simplicity, we assume that an interior-point pair $(\mathbf{x}^0,\mathbf{y}^0,\mathbf{s}^0)$ is known, and we will use it as our initial point pair.

Barrier Functions and Analytic Center

Consider the barrier function optimization problems:

AC

$$(PB) \quad \text{minimize} \quad -\sum_{j=1}^{n} \log x_j \qquad \text{s.t.} \quad \mathbf{x} \in \text{int}\, \mathcal{F}_p$$

and

$$(DB) \quad \text{maximize} \quad \sum_{j=1}^{n} \log s_j \quad \log\left(\prod s_j\right) \qquad \text{s.t.} \quad (\mathbf{y}, \mathbf{s}) \in \text{int}\, \mathcal{F}_d$$

The maximizer $\mathbf{x}$ (or $(\mathbf{y}, \mathbf{s})$) of (PB) (or (BD)) is called the analytic center of bounded polyhedron $\mathcal{F}_p$ (or $\mathcal{F}_d$). Applying the KKT conditions and using $X = \text{diag}(\mathbf{x})$, we have

$$-X^{-1}\mathbf{e} - A^T\mathbf{y} = \mathbf{0} \quad \text{or} \quad -\mathbf{e} - XA^T\mathbf{y} = \mathbf{0}, \; A\mathbf{x} = \mathbf{b}, \; \mathbf{x} > \mathbf{0}.$$

After introducing auxiliary vector $\mathbf{s} = X^{-1}\mathbf{e}$, the conditions become

$$\begin{array}{rcl} X\mathbf{s} & = & \mathbf{e} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & \mathbf{0} \\ \mathbf{x} & > & \mathbf{0}. \end{array} \qquad \text{or} \left(\begin{array}{rcl} \|S\mathbf{x} & = & \mathbf{e}\| \\ A\mathbf{x} & = & \mathbf{0} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c} \\ \mathbf{s} & > & \mathbf{0}. \end{array} \right) \le O(n) \quad O(n^2)$$

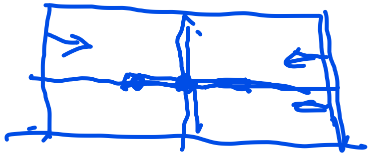
(A, b)

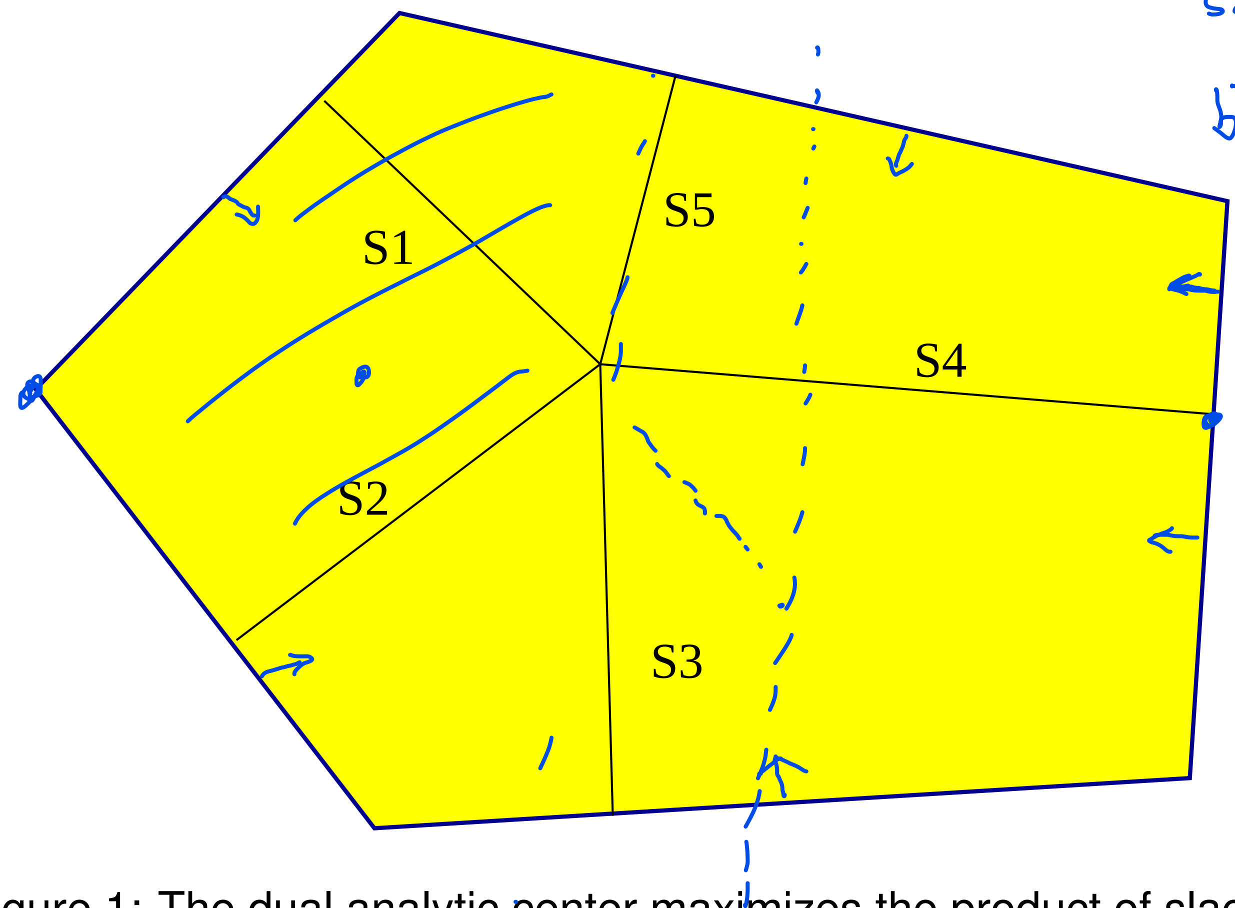
CG

$\max \min_j (s_j)$

s.t. $(y, s) \in F_p$

$b^T y$





$b^T y \geq b^T y_0$

y_0

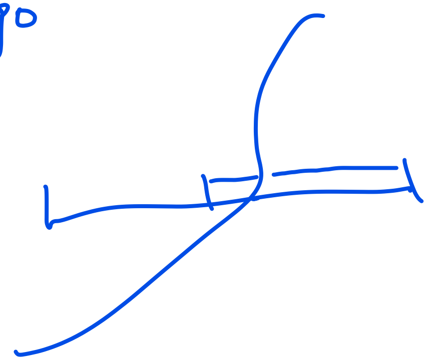


Figure 1: The dual analytic center maximizes the product of slacks.

Examples

$$\mathcal{F}_p = \{\mathbf{x} : \sum_j \mathbf{x}_j = 1,\ \mathbf{x} \geq \mathbf{0}\}.$$

The analytic center of $\mathcal{F}_p$ would be

$$\mathbf{x}^c = (\frac{1}{n};\ ...;\ \frac{1}{n}),\ y = -n,\ \mathbf{s} = (n;\ ...;\ n).$$

$$\mathcal{F}_d = \{\mathbf{y} : \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}.$$

The analytic center of $\mathcal{F}_d$ would be

$$\mathbf{y}^c = \arg\max \sum_i (\log(y_i) + \log(1 - y_i)) = \arg\max \sum_i \log(y_i(1 - y_i))$$

that is

$$\mathbf{y}^c = (\frac{1}{2};\ ...;\ \frac{1}{2}),\ \mathbf{s} = \frac{1}{2}\mathbf{e},\ \mathbf{x} = 2\mathbf{e}.$$

Why “analytic”: depending on the analytical representation data.

Logarithmic Function and Scaled Concordant Lipschitz

Lemma 1 *Let $B(\mathbf{x}) = -\sum_{j=1}^{n} \log(x_j)$. Then, for any point $\mathbf{x} > \mathbf{0}$ and direction vector $\mathbf{d}$ such that $\|X^{-1}\mathbf{d}\|_\infty \leq \alpha (< 1)$,*

$$-\mathbf{e}^T X^{-1}\mathbf{d} \leq B(\mathbf{x}+\mathbf{d}) - B(\mathbf{x}) \leq -\mathbf{e}^T X^{-1}\mathbf{d} + \frac{\|X^{-1}\mathbf{d}\|^2}{2(1-\alpha)}.$$

The Barrier function property can be generalized to the so-called Second-Order Scaled Concordant Lipschitz Condition: for any $\mathbf{x} > \mathbf{0}$ and $\mathbf{x} + \mathbf{d}$ in the function domain:

$$\|X\left(\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\right)\| \leq \beta_\alpha \mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d}, \text{ whenever } \|X^{-1}\mathbf{d}\| \leq \alpha(< 1).$$

Such condition can be verified using Taylor Expansion Series; basically, the scaled third derivative of the function is bounded by its (unscaled) second derivative.

- All quadratic functions are scaled concordant Lipschitz with $\beta_\alpha = 0$.
- Convex function $-\log(x)$ is scaled concordant Lipschitz with $\beta_\alpha = \frac{1}{(1-\alpha)}$.
- All power functions $\{x^p :\ x > 0\}$ with integer p are scaled concordant Lipschitz with $\beta_\alpha = \frac{O(p)}{(1-\alpha)}$.

Affine-Scaling Gradient Projection

To compute the analytic center, we consider the affine-scaling GPM from any feasible $\mathbf{x} > \mathbf{0}$:

$$\begin{array}{ll} \text{minimize} & -\mathbf{e}^T X^{-1}\mathbf{d} \\ \text{s.t.} & A\mathbf{d} = \mathbf{0},\ \|X^{-1}\mathbf{d}\| \le \alpha \end{array} \quad \text{or} \quad \begin{array}{ll} \text{minimize} & -\mathbf{e}^T \mathbf{d}' \\ \text{s.t.} & AX\mathbf{d}' = \mathbf{0},\ \|\mathbf{d}'\| \le \alpha \end{array}$$

which has a close-form solution

$$\mathbf{d}' = \alpha(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}/\|(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}\|.$$

Note that $\mathbf{d} = X\mathbf{d}'$ so that we let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}$, which should remain positive:

$$\mathbf{x}^+ = \mathbf{x} + \mathbf{d} = \mathbf{x} + X\mathbf{d}' = X(\mathbf{e} + \mathbf{d}') > \mathbf{0}$$

as long as $\mathbf{x} > \mathbf{0}$ and $\|\mathbf{d}'\| < 1$. Then, from Lemma 1 the Barrier function value would be decreased at least by

$$B(\mathbf{x}^+) - B(\mathbf{x}) \le -\alpha\|(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}\| + \frac{\alpha^2}{2(1-\alpha)}.$$

$x_j s_j = 1 \Leftrightarrow \frac{1}{x_j} - s_j = 0$

Convergence Speed Analysis

For simplicity, let $\mathbf{y}(\mathbf{x}) = (AX^2A^T)^{-1}AX\mathbf{e}$ and $\mathbf{s}(\mathbf{x}) = A^T\mathbf{y}(\mathbf{s})$ so that

$$(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e} = \mathbf{e} - X\mathbf{s}(\mathbf{x}).$$

Note that $\mathbf{y}(\mathbf{x})$ minimizes $\min_{\mathbf{y}} \ \|\mathbf{e} - XA^T\mathbf{y}\|^2$.

Thus, as long as $\|\mathbf{e} - X\mathbf{s}(\mathbf{x})\| \geq 1$, the Barrier function can be decreased by a universal constant $-\alpha + \frac{\alpha^2}{2(1-\alpha)} = -3/4$ when we set $\alpha = 1/2$.

If the quantity $\|\mathbf{e} - X\mathbf{s}(\mathbf{x})\| < 1$, then we simply let $\mathbf{x}^+ = \mathbf{x} + X(\mathbf{e} - X\mathbf{s}(\mathbf{x}))$, in which case we now prove $\|\mathbf{e} - X^+\mathbf{s}(\mathbf{x}^+)\| \leq \|\mathbf{e} - X\mathbf{s}(\mathbf{x})\|^2$ (quadratic convergence)!

$$\begin{aligned}
\|\mathbf{e} - X^+\mathbf{s}(\mathbf{x}^+)\|^2 &\leq \|\mathbf{e} - X^+\mathbf{s}(\mathbf{x})\|^2, \quad \text{(because } \mathbf{y}(\mathbf{x}^+) \text{ minimizes the squares)} \\
&= \|\mathbf{e} - (2X - X^2S(\mathbf{x})\mathbf{s}(\mathbf{x})\|^2 \\
&= \sum_{j=1}^n \left(1 - 2x_js_j(\mathbf{x}) + x_j^2(s_j(\mathbf{x}))^2\right)^2 \\
&= \sum_{j=1}^n (1 - x_js_j(\mathbf{x}))^4 \\
&\leq \left(\sum_{j=1}^n (1 - x_js_j(\mathbf{x}))^2\right)^2 = \|\mathbf{e} - X\mathbf{s}(\mathbf{x})\|^4.
\end{aligned}$$

Analytic Volume and Cutting Plane for LP: Geometric Interpretation

$$AV(\mathcal{F}_d) := \prod_{j=1}^{n} \bar{s}_j = \prod_{j=1}^{n} (c_j - \mathbf{a}_j^T \bar{\mathbf{y}})$$

can be viewed as the analytic volume of polytope $\mathcal{F}_d$ or simply $\mathcal{F}$ in the rest of discussions.

If one inequality in $\mathcal{F}$, say the first one, needs to be translated, change $\mathbf{a}_1^T \mathbf{y} \le c_1$ to $\mathbf{a}_1^T y \le \mathbf{a}_1^T \bar{\mathbf{y}}$; i.e., the first inequality is parallelly moved and it now cuts through $\bar{\mathbf{y}}$ and divides $\mathcal{F}$ into two bodies. Analytically, c_1 is replaced by $\mathbf{a}_1^T \bar{\mathbf{y}}$ and the rest of data are unchanged. Let

$$\mathcal{F}^+ := \{\mathbf{y} : \ \mathbf{a}_j^T \mathbf{y} \le c_j^+, \ j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = 2, ..., n$ and $c_1^+ = \mathbf{a}_1^T \bar{\mathbf{y}}$.

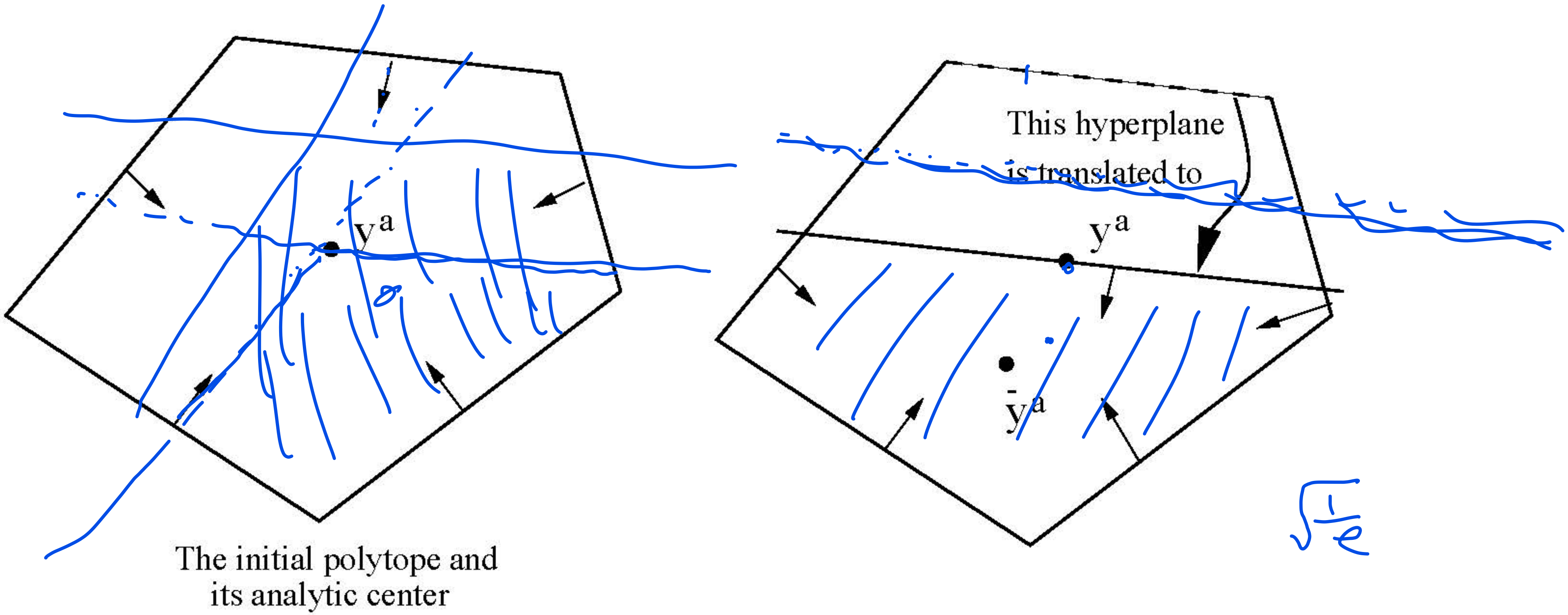


Figure 2: Translation of a hyperplane to the AC.

Analytic Volume Reduction of the New Polytope

Let $\bar{\mathbf{y}}^+$ be the analytic center of $\mathcal{F}^+$. Then, the analytic volume of $\mathcal{F}^+$

$$AV(\mathcal{F}^+) = \prod_{j=1}^{n}(c_j^+ - \mathbf{a}_j^T\bar{\mathbf{y}}^+) = (\mathbf{a}_1^T\bar{\mathbf{y}} - \mathbf{a}_1^T\bar{\mathbf{y}}^+)\prod_{j=2}^{n}(c_j - \mathbf{a}_j^T\bar{\mathbf{y}}^+).$$

We have the following volume reduction theorem:

Theorem 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \le \exp(-1). \quad \frac{1}{e}$$

2.718

Proof

Since $\bar{\mathbf{y}}$ is the analytic center of $\mathcal{F}$, there exists $\bar{\mathbf{x}} > \mathbf{0}$ such that

$$\bar{X}\bar{\mathbf{s}} = \bar{X}(\mathbf{c} - A^T\bar{\mathbf{y}}) = \mathbf{e} \quad \text{and} \quad A\bar{\mathbf{x}} = \mathbf{0}.$$

Thus,

$$\bar{\mathbf{s}} = (\mathbf{c} - A^T\bar{\mathbf{y}}) = \bar{X}^{-1}\mathbf{e} \quad \text{and} \quad \mathbf{c}^T\bar{\mathbf{x}} = (\mathbf{c} - A^T\bar{\mathbf{y}})^T\bar{\mathbf{x}} = \mathbf{e}^T\mathbf{e} = n.$$

We have

$$\begin{aligned}\mathbf{e}^T\bar{X}\bar{\mathbf{s}}^+ &= \mathbf{e}^T\bar{X}(\mathbf{c}^+ - A^T\bar{\mathbf{y}}^+) = \mathbf{e}^T\bar{X}\mathbf{c}^+ \\ &= \mathbf{c}^T\bar{\mathbf{x}} - \bar{x}_1(c_1 - \mathbf{a}_1^T\bar{\mathbf{y}}) = n - 1.\end{aligned}$$

$$
\begin{aligned}
\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} &= \prod_{j=1}^{n} \frac{\bar{s}_j^+}{\bar{s}_j} \\
&= \prod_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \\
&\le \left(\frac{1}{n} \sum_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \right)^n \\
&= \left(\frac{1}{n} \mathbf{e}^T \bar{X} \bar{\mathbf{s}}^+ \right)^n \\
&= \left(\frac{n-1}{n} \right)^n \le \exp(-1).
\end{aligned}
$$

Analytic Volume of Polytope and Multiple Cutting Planes

Now suppose we translate $k(< n)$ hyperplanes, say $1, 2, ..., k$, moved to cut the analytic center $\bar{\mathbf{y}}$ of $\mathcal{F}$, that is,

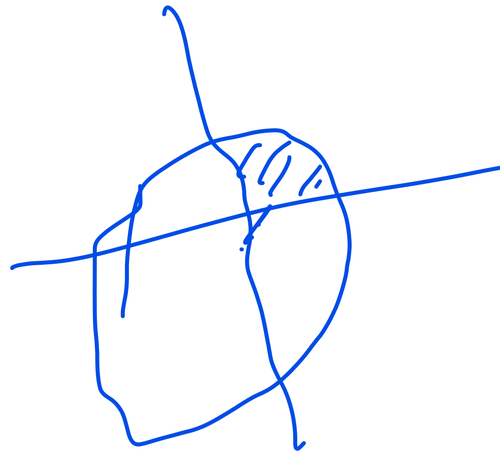
$$\mathcal{F}^+ := \{\mathbf{y} : \ \mathbf{a}_j^T \mathbf{y} \leq c_j^+, \ j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = k+1, ..., n$ and $c_j^+ = \mathbf{a}_j^T \bar{\mathbf{y}}$ for $j = 1, ..., k$.

Corollary 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \leq \exp(-k).$$

$\left(\frac{1}{e}\right)^k$



Barrier Regularization Function for LP: Algebraic Implementation

$b^T y(90) > b^T y(100)$

Consider the LP pair with the barrier function

$x(\mu)$

$$(LPB) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^n \log x_j \quad \text{s.t.} \quad \mathbf{x} \in \text{int}\,\mathcal{F}_p$$

$<=>$

$$(LDB) \quad \text{maximize} \quad \mathbf{b}^T\mathbf{y} + \mu\sum_{j=1}^n \log s_j \quad \text{s.t.} \quad (\mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}_d,$$

and they are primal-dual to each other and share a common set of KKT Optimality Conditions:

$x(100)$ 100

$x(90)$ 90

$c^T x(90) < c^T x(100)$

$\boxed{s - \mu./x = 0}$ P

$$\begin{cases} X\mathbf{s} = \mu\mathbf{e} \\ A\mathbf{x} = \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} = -\mathbf{c}; \end{cases} \tag{5}$$

$\boxed{\mu^{k+1} = \left(1 - \frac{1}{2\sqrt{n}}\right)\mu^k}$

$x - \mu./s = 0$

where barrier parameter

$$\mu = \frac{\mathbf{x}^T\mathbf{s}}{n} = \frac{\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}}{n},$$

so that it's the average of complementarity or duality gap. As μ varies, the optimizers form the LP central paths in the primal and dual feasible regions, respectively.

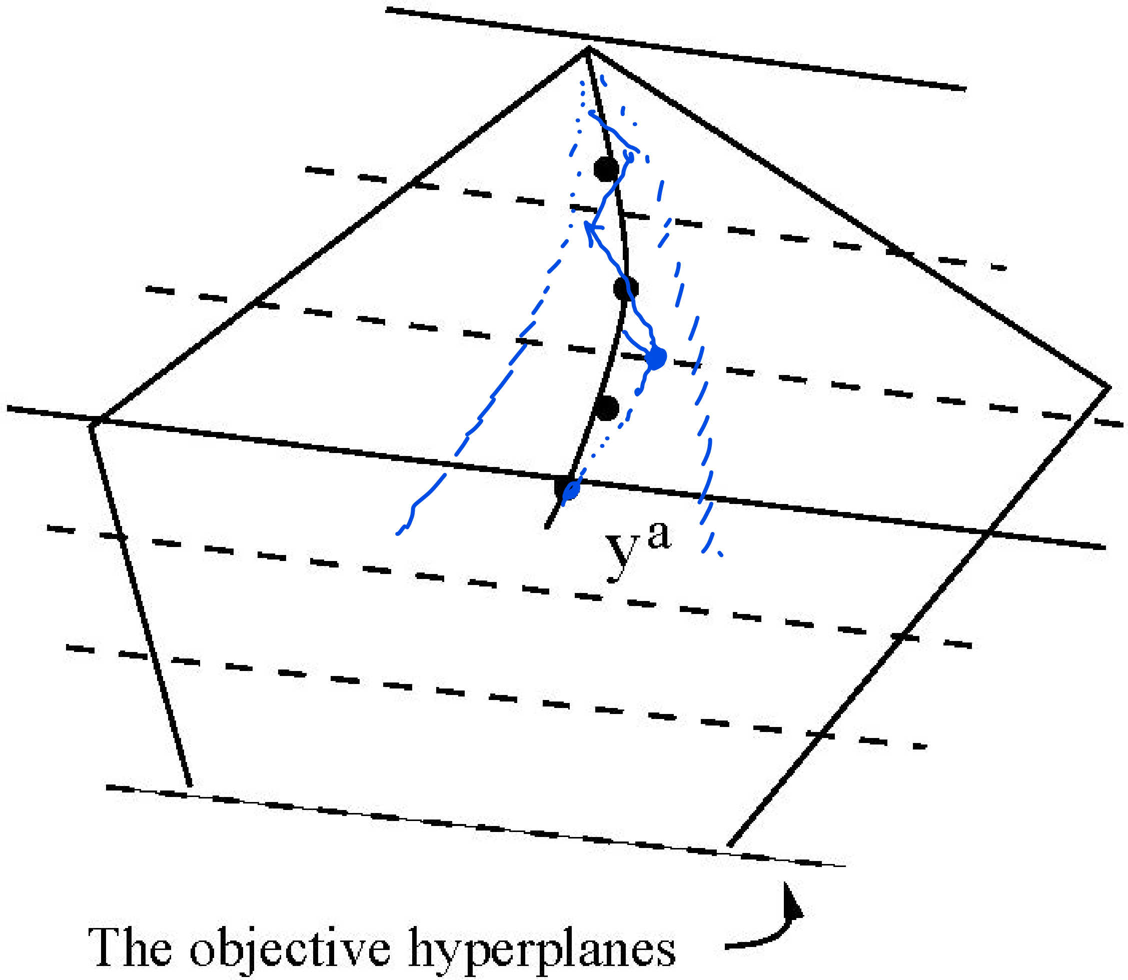


Figure 3: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Examples

$$\min \sum_j c_j \mathbf{x}_j - \mu \sum_j \log(x_j) \text{ s.t. } \sum_j x_j = 1.$$

$$c_j - \frac{\mu}{x_j} = y, \; x_j > 0, \; \forall j,$$

thus, $x_j = \frac{\mu}{c_j - y}, \; \forall j$. Then, from

$$\sum_j \frac{\mu}{c_j - y} = 1, \; c_j - y > 0, \; \forall j,$$

we can solve $y(\mu)$ and $\mathbf{x}(\mu)$ as the roots of polynomials.

Central Path for Linear Programming

$$\mathcal{C} = \{(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu)) \in \operatorname{int} \mathcal{F} : \; X\mathbf{s} = \mu\mathbf{e}, \; 0 < \mu < \infty\};$$

is called the (primal and dual) central path of linear programming.

Theorem 2 *Let both (LP) and (LD) have interior feasible points for the given data set* $(A, \mathbf{b}, \mathbf{c})$. *Then for any* $0 < \mu < \infty$, *the central path point pair* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *exists and is unique. Moreover, the followings hold.*

i) *The central path point* $(\mathbf{x}(\mu), \mathbf{s}(\mu))$ *is bounded for* $0 < \mu \le \mu^0$ *and any given* $0 < \mu^0 < \infty$.

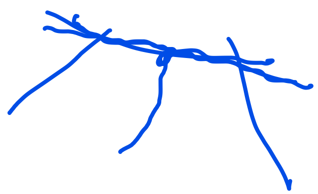
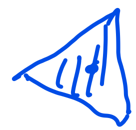
ii) *For* $0 < \mu' < \mu$,

$$\mathbf{c}^T\mathbf{x}(\mu') < \mathbf{c}^T\mathbf{x}(\mu) \quad \textit{and} \quad \mathbf{b}^T\mathbf{y}(\mu') > \mathbf{b}^T\mathbf{y}(\mu)$$

$x(\mu) \xrightarrow[\mu \to 0^+]{} x(0)$

if both primal and dual have no constant objective values.

iii) $(\mathbf{x}(\mu), \mathbf{s}(\mu))$ *converges to an optimal solution pair for (LP) and (LD). Moreover, the limit point* $\mathbf{x}(0)_{P^*} > \mathbf{0}$ *and the limit point* $\mathbf{y}(0), \mathbf{s}(0)_{Z^*} > \mathbf{0}$ *are the analytic centers of the optimal solution sets of primal and dual, respectively; where* (P^*, Z^*) *is the strictly complementarity partition if variable index set* $\{1, 2, ..., n\}$.

Proof of (iii)

Since $\mathbf{x}(\mu)$ and $\mathbf{s}(\mu)$ are both bounded, they have at least one limit point which we denote by $\mathbf{x}(0)$ and $\mathbf{s}(0)$. Let $\mathbf{x}^*_{P*} > \mathbf{0}$ ($\mathbf{x}^*_{Z*} = \mathbf{0}$) and $\mathbf{s}^*_{Z*} > \mathbf{0}$ ($\mathbf{s}^*_{P*} = \mathbf{0}$), be the analytic centers on the optimal sets of on the primal and dual optimal faces, respectively, that is, they are the maximizers of
$\{\prod_{j\in P*} x_j : \; A_{P*}\mathbf{x}_{P*} = \mathbf{b}, \; \mathbf{x}_{P*} \geq \mathbf{0}\}$ and
$\{\prod_{j\in Z*} s_j : \; \mathbf{s}_{Z*} = \mathbf{c}_{Z*} - A^T_{Z*}\mathbf{y} \geq \mathbf{0}, \; \mathbf{c}_{P*} - A^T_{P*}\mathbf{y} = \mathbf{0}\}$, respectively. Note
$(\mathbf{x}(\mu) - \mathbf{x}^*)^T(\mathbf{s}(\mu) - \mathbf{s}^*) = 0,$ so that

$$\sum_j^n \left(s^*_j x(\mu)_j + x^*_j s(\mu)_j\right) = n\mu, \quad \text{or} \quad \sum_{j\in P*}\left(\frac{x^*_j}{x(\mu)_j}\right) + \sum_{j\in Z*}\left(\frac{s^*_j}{s(\mu)_j}\right) = n.$$

Therefore, from the arithmetic-geometric mean inequality we have

$$\prod_{j\in P*}\frac{x^*_j}{x(\mu)_j}\prod_{j\in Z*}\frac{s^*_j}{s(\mu)_j} \leq 1, \quad \text{or} \quad \left(\prod_{j\in P*} x(\mu)_j\right)\left(\prod_{j\in Z*} s(\mu)_j\right) \geq \left(\prod_{j\in P*} x^*_j\right)\left(\prod_{j\in Z*} s^*_j\right)$$

The limit points must also satisfy the inequality which implies $\prod_{j\in P*} x(0)_j \geq \prod_{j\in P*} x^*_j$ and $\prod_{j\in Z*} s(0)_j \geq \prod_{j\in Z*} s^*_j$. But the analytic center is unique so that the claim is true.

$s = c - A^T y$

The Primal-Dual Path-Following Algorithm for LP

$x(\mu)\left(1 - \frac{1}{2\sqrt{n}}\right)$

In general, we start from an (approximate) central path point $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \mathcal{F}$ such that

$\begin{pmatrix} x_1^k s_1^k \\ \vdots \\ x_n^k s_n^k \end{pmatrix}$ $=0$

$$\|X^k \mathbf{s}^k - \mu^k \mathbf{e}\| \leq \sigma \mu^k, \quad \text{for some } \sigma \in [0, 1).$$

Then, let $\mu^{k+1} = (1 - \eta)\mu^k$ for some $\eta \in (0,\ 1]$, we aim to find a new pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}$ such that

$$X\mathbf{s} = \mu^{k+1}\mathbf{e}.$$

We start from $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \mathcal{F}$ and apply the Newton iteration for direction vectors $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$:

$(S^k)^{-1}$ $(S^k)^{-1}$

$x^{k+1} = x^k + d_x$
$y^{k+1} = y^k + d_y$
$s^{k+1} = s^k + d_s$

$$\begin{aligned} \cancel{S^k}\mathbf{d}_x + X^k \mathbf{d}_s &= \mu^{k+1}\mathbf{e} - X^k \mathbf{s}^k \\ A\mathbf{d}_x &= \mathbf{0} \\ A^T \mathbf{d}_y + \mathbf{d}_s &= \mathbf{0} \end{aligned},$$

$\boxed{\text{Path}}$ μ

then let $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x, \quad \mathbf{y}^{k+1} = \mathbf{y}^k + \mathbf{d}_y, \quad \mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{d}_s$. Carefully choosing $\sigma = O(1)$ and $\eta = O(\frac{1}{\sqrt{n}})$ guarantees $(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) > \mathbf{0}$ and

$$\|X^{k+1}\mathbf{s}^{k+1} - \mu^{k+1}\mathbf{e}\| \leq \sigma\mu^{k+1}, \quad \text{for the same } \sigma \in [0, 1).$$

Too many restrictions when following a path... Is a function-driven interior-point algorithm?