A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization

by Joachim Dahl and Erling D. Andersen

CHEN Li

Friday Seminar
April 23, 2021

Outline

Preliminaries

Primal-dual scalings — generalize Nesterov-Todd scaling

Search directions — Mehrotra like predictor-corrector

Numerical Experiments

Conic Linear Optimization

$$\begin{array}{rl} \min & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array} \qquad \text{(Primal)}$$

$$\begin{array}{rl} \max & \langle b, y \rangle \\ \text{s.t.} & A^T y + s = c \\ & s \in \mathcal{K}^* \end{array} \qquad \text{(Dual)}$$

where

- $y \in \mathbb{R}^m$, $\mathcal{K} \subseteq \mathbb{R}^n$ is a proper cone (pointed, closed, convex cone with nonempty interior).
- A has full row rank
- We focus on exponential cone

  $$\mathcal{K}_{\exp} := \text{cl}\left\{x \in \mathbb{R}^3 \mid x_1 \geq x_2 \exp(x_3/x_2), x_2 > 0\right\}$$

  with its dual cone

  $$\mathcal{K}_{\exp}^* := \text{cl}\left\{z \in \mathbb{R}^3 \mid e \cdot z_1 \geq -z_3 \exp(z_2/z_3), z_3 < 0\right\}$$

Barrier Functions for Interior Point Method

- $F : \operatorname{int} \mathcal{K} \to \mathbb{R}$ is ν-**log homogeneous self-concordant barrier (ν-LHSCB)** if F is C^3 barrier function and

$$
\begin{array}{ll}
|F'''(x)[u,u,u]| \leq 2F''(x)[u,u]^{3/2} & \forall x \in \operatorname{int} \mathcal{K}, u \in \mathbb{R}^n \\
F(\tau x) = F(x) - \nu \log \tau & \forall \tau > 0
\end{array}
$$

Barrier Functions for Interior Point Method

- $F : \operatorname{int} \mathcal{K} \to \mathbb{R}$ is ν-**log homogeneous self-concordant barrier** **(ν-LHSCB)** if F is C^3 barrier function and

$$\begin{array}{ll} |F'''(x)[u,u,u]| \leq 2F''(x)[u,u]^{3/2} & \forall x \in \operatorname{int} \mathcal{K}, u \in \mathbb{R}^n \\ F(\tau x) = F(x) - \nu \log \tau & \forall \tau > 0 \end{array}$$

- Useful properties from ν-log homogeneity:

$$\begin{array}{ll} F'(\tau x) = \dfrac{1}{\tau} F'(x), & F''(\tau x) = \dfrac{1}{\tau^2} F''(x) \\ F''(x)x = -F'(x), & F'''(x)x = -2F''(x) \\ \langle F'(x), x \rangle = -\nu & \end{array} \tag{1}$$

Barrier Functions for Interior Point Method

- $F : \operatorname{int} \mathcal{K} \to \mathbb{R}$ is ν-**log homogeneous self-concordant barrier (**ν**-LHSCB)** if F is C^3 barrier function and

$$\begin{array}{ll} |F'''(x)[u,u,u]| \leq 2F''(x)[u,u]^{3/2} & \forall x \in \operatorname{int} \mathcal{K}, u \in \mathbb{R}^n \\ F(\tau x) = F(x) - \nu \log \tau & \forall \tau > 0 \end{array}$$

- Useful properties from ν-log homogeneity:

$$\begin{array}{ll} F'(\tau x) = \dfrac{1}{\tau} F'(x), & F''(\tau x) = \dfrac{1}{\tau^2} F''(x) \\ F''(x)x = -F'(x), & F'''(x)x = -2F''(x) \\ \langle F'(x), x \rangle = -\nu & \end{array} \tag{1}$$

- **Conjugate barrier**:

$$F_*(s) = \sup_{x \in \operatorname{int} \mathcal{K}} \{-\langle s, x \rangle - F(x)\}$$

is a ν-LHSCB for $\mathcal{K}^*$.

Barrier Functions for Interior Point Method

- $F : \operatorname{int} \mathcal{K} \to \mathbb{R}$ is **ν-log homogeneous self-concordant barrier (ν-LHSCB)** if F is C^3 barrier function and

$$\begin{array}{ll} |F'''(x)[u,u,u]| \leq 2F''(x)[u,u]^{3/2} & \forall x \in \operatorname{int} \mathcal{K}, u \in \mathbb{R}^n \\ F(\tau x) = F(x) - \nu \log \tau & \forall \tau > 0 \end{array}$$

- Useful properties from ν-log homogeneity:

$$\begin{array}{ll} F'(\tau x) = \dfrac{1}{\tau} F'(x), & F''(\tau x) = \dfrac{1}{\tau^2} F''(x) \\ F''(x)x = -F'(x), & F'''(x)x = -2F''(x) \\ \langle F'(x), x \rangle = -\nu & \end{array} \tag{1}$$

- **Conjugate barrier**:

$$F_*(s) = \sup_{x \in \operatorname{int} \mathcal{K}} \{-\langle s, x \rangle - F(x)\}$$

  is a ν-LHSCB for $\mathcal{K}^*$.
- 3-LHSCB for exponential cone:

$$F(x) = -\log(x_2 \log(x_1/x_2) - x_3) - \log x_1 - \log x_2$$

  Note $F_*'(s) = -x_s$ and $F_*''(s) = F''(x_s)^{-1}$ where

$$x_s = \arg \min_{x \in \operatorname{int} \mathcal{K}} \{\langle s, x \rangle + F(x)\}$$

Homogeneous Self-dual (HSD) Model

Embed KKT into the **HSD model**:

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

$$x \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_k, \quad s \in \mathcal{K}_1^* \times \cdots \times \mathcal{K}_k^*, \quad y \in \mathbb{R}^m, \quad \tau, \kappa \geq 0$$

Homogeneous Self-dual (HSD) Model

Embed KKT into the **HSD model**:

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

$$x \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_k, \quad s \in \mathcal{K}_1^* \times \cdots \times \mathcal{K}_k^*, \quad y \in \mathbb{R}^m, \quad \tau, \kappa \geq 0$$

By skew-symmetry of the matrix in HSD model, any feasible solution (x, s, y, τ, κ) must satisfy

$$\langle x, s \rangle + \tau\kappa = 0$$

Homogeneous Self-dual (HSD) Model

Embed KKT into the **HSD model**:

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

$$x \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_k, \quad s \in \mathcal{K}_1^* \times \cdots \times \mathcal{K}_k^*, \quad y \in \mathbb{R}^m, \quad \tau, \kappa \geq 0$$

By skew-symmetry of the matrix in HSD model, any feasible solution (x, s, y, τ, κ) must satisfy

$$\langle x, s \rangle + \tau\kappa = 0$$

1. If $\tau > 0$, then x/τ, $(y, s)/\tau$ are optimal solutions for primal and dual problems.
2. If $\kappa > 0$, then either $\langle b, y \rangle > 0$ and primal is infeasible; or $\langle c, x \rangle < 0$ and dual is infeasible; or both.
3. If $\kappa = \tau = 0$, then no conclusion

Notations

For notation convenience, we suppress τ and κ and define new variables

$$x := (x; \tau) \in \mathcal{K} := \mathcal{K}_1 \times \cdots \times \mathcal{K}_k \times \mathbb{R}_+$$
$$s := (s; \kappa) \in \mathcal{K}^* := \mathcal{K}_1^* \times \cdots \times \mathcal{K}_k^* \times \mathbb{R}_+$$

and barrier

$$F(x) := \sum_{i=1}^{k} F(x_i) - \log \tau$$

with complexity $\nu := \sum_{i=1}^{k} \nu_i + 1$.

Notations

For notation convenience, we suppress τ and κ and define new variables

$$x := (x; \tau) \in \mathcal{K} := \mathcal{K}_1 \times \cdots \times \mathcal{K}_k \times \mathbb{R}_+$$
$$s := (s; \kappa) \in \mathcal{K}^* := \mathcal{K}_1^* \times \cdots \times \mathcal{K}_k^* \times \mathbb{R}_+$$

and barrier

$$F(x) := \sum_{i=1}^{k} F(x_i) - \log \tau$$

with complexity $\nu := \sum_{i=1}^{k} \nu_i + 1$.

Let $z := (x, s, y) \in \mathcal{D} := \mathcal{K} \times \mathcal{K}^* \times \mathbb{R}^m$ and define

$$G(z) := \begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} - \begin{bmatrix} 0 \\ s \end{bmatrix}$$

Then HSD model is

$$G(z) = 0, \quad z \in \mathcal{D}$$

Central Path

Given an initial point $z^0 \in \operatorname{int} \mathcal{D}$, consider central path (parameterized by $\mu \in (0, 1]$) defined by the optimal solution

$$z_\mu := \arg\min_{z \in \mathcal{D}} \left\{ \Phi(z) : \ G(z) = \mu G(z^0) \right\} \tag{2}$$

where

$$\Phi(z) := F(x) + F_*(s) + \langle x^0, s \rangle + \langle x, s^0 \rangle.$$

[1] Yurii Nesterov, Michael J Todd, and Yinyu Ye. "Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems". In: *Mathematical Programming* 84.2 (1999), pp. 227–268.

Central Path

Given an initial point $z^0 \in \operatorname{int} \mathcal{D}$, consider central path (parameterized by $\mu \in (0, 1]$) defined by the optimal solution

$$z_\mu := \arg\min_{z \in \mathcal{D}} \left\{ \Phi(z) : \; G(z) = \mu G(z^0) \right\} \tag{2}$$

where

$$\Phi(z) := F(x) + F_*(s) + \langle x^0, s \rangle + \langle x, s^0 \rangle.$$

It is known[1] that the z_μ is characterized by

$$G(z_\mu) = \mu G(z^0) \tag{3}$$

$$s_\mu = -\mu F'(x_\mu), \quad x_\mu = -\mu F'_*(s_\mu) \tag{4}$$

[1] **Nesterov, Todd, and Ye**, "Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems".

Central Path

Given an initial point $z^0 \in \operatorname{int} \mathcal{D}$, consider central path (parameterized by $\mu \in (0, 1]$) defined by the optimal solution

$$z_\mu := \arg\min_{z \in \mathcal{D}} \left\{ \Phi(z) : \; G(z) = \mu G(z^0) \right\} \tag{2}$$

where

$$\Phi(z) := F(x) + F_*(s) + \langle x^0, s \rangle + \langle x, s^0 \rangle.$$

It is known[1] that the z_μ is characterized by

$$G(z_\mu) = \mu G(z^0) \tag{3}$$

$$s_\mu = -\mu F'(x_\mu), \quad x_\mu = -\mu F'_*(s_\mu) \tag{4}$$

Note that on the central path

$$\langle x_\mu, s_\mu \rangle / \nu = \mu.$$

[1] **Nesterov, Todd, and Ye**, "Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems".

Central Path Neighborhood

- Let
$$\tilde{x} := -F_*'(s), \quad \tilde{s} := -F'(x), \quad \tilde{\mu} := \langle \tilde{x}, \tilde{s} \rangle / \nu$$
for any iterate $(x, s, y) \in \mathcal{D}$.

[2] Yu E Nesterov and Michael J Todd. "Primal-dual interior-point methods for self-scaled cones". In: *SIAM Journal on optimization* 8.2 (1998), pp. 324–364.

Central Path Neighborhood

- Let
$$\tilde{x} := -F_*'(s), \quad \tilde{s} := -F'(x), \quad \tilde{\mu} := \langle \tilde{x}, \tilde{s} \rangle / \nu$$
for any iterate $(x, s, y) \in \mathcal{D}$.
- It is known[2] that
$$\mu\tilde{\mu} \geq 1$$
with equality only on the central path. So
$$\inf \{\beta : \quad \beta\mu\tilde{\mu} \leq 1, \ \beta \in (0, 1]\}$$
indicates how much the iterate deviates the central path.

[2] **Nesterov and Todd**, "Primal-dual interior-point methods for self-scaled cones".

Central Path Neighborhood

- Let
$$\tilde{x} := -F_*'(s), \quad \tilde{s} := -F'(x), \quad \tilde{\mu} := \langle \tilde{x}, \tilde{s} \rangle / \nu$$
for any iterate $(x, s, y) \in \mathcal{D}$.
- It is known[2] that
$$\mu \tilde{\mu} \geq 1$$
with equality only on the central path. So
$$\inf \{\beta : \quad \beta \mu \tilde{\mu} \leq 1, \ \beta \in (0, 1]\}$$
indicates how much the iterate deviates the central path.
- Define the central path neighborhood
$$\mathcal{N}(\beta) = \left\{(x, s) \in \mathcal{K} \times \mathcal{K}^* \mid \nu_i \langle F'(x_i), F_*'(s_i) \rangle^{-1} \geq \beta \mu, \forall i = 1, ..., k+1\right\}$$
It is the generalization of one-sided ∞-norm neighborhood in LP.

[2] **Nesterov and Todd**, "Primal-dual interior-point methods for self-scaled cones".

Central Path Neighborhood

- Let
$$\tilde{x} := -F_*'(s), \quad \tilde{s} := -F'(x), \quad \tilde{\mu} := \langle \tilde{x}, \tilde{s} \rangle / \nu$$
for any iterate $(x, s, y) \in \mathcal{D}$.
- It is known[2] that
$$\mu\tilde{\mu} \geq 1$$
with equality only on the central path. So
$$\inf \{\beta : \quad \beta\mu\tilde{\mu} \leq 1, \ \beta \in (0, 1]\}$$
indicates how much the iterate deviates the central path.
- Define the central path neighborhood
$$\mathcal{N}(\beta) = \left\{(x, s) \in \mathcal{K} \times \mathcal{K}^* \mid \nu_i \langle F'(x_i), F_*'(s_i) \rangle^{-1} \geq \beta\mu, \forall i = 1, ..., k+1\right\}$$
It is the generalization of one-sided ∞-norm neighborhood in LP.
- Typically β is small, say $\beta = 10^{-6}$.

[2] **Nesterov and Todd**, "Primal-dual interior-point methods for self-scaled cones".

The Primal-Dual Interior Point Algorithm

1. Initialization:

$$y^0 = 0, \quad x^0 = s^0 = -F'(x^0),$$

   then $\langle x^0, s^0 \rangle = \nu$.

   - For exponential cone, $x^0 = (1.290928, 0.805102, -0.827838)$.

The Primal-Dual Interior Point Algorithm

1. Initialization:
$$y^0 = 0, \quad x^0 = s^0 = -F'(x^0),$$
   then $\langle x^0, s^0 \rangle = \nu$.
   - For exponential cone, $x^0 = (1.290928, 0.805102, -0.827838)$.

2. In each iteration, given current iterate $z = (x, s, y)$, do
   - Compute scaling matrix W
   - Compute affine search direction Δz^a, the corrector term, the stepsize α_a and the centering parameter σ
   - Compute the combined centering-corrector search direction Δz and update $z = z + \alpha \Delta z$ with the largest stepsize inside the central path neighborhood $\mathcal{N}_\beta$
   - Check termination criteria

Outline

Preliminaries

Primal-dual scalings — generalize Nesterov-Todd scaling

Search directions — Mehrotra like predictor-corrector

Numerical Experiments

Primal-Dual Scaling Matrix

Tunçel[3] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T : T \succ 0, T^2 s = x, T^2 F'(x) = F'_*(s)\}$$

and a set of bounded scalings parameterized by ξ:

$$\mathcal{T}_2(x,s,\xi) := \{T \in \mathcal{T}_1(x,s) : (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F F''(x)^{-1}\}$$

where $\delta_F := (\nu(\mu\tilde{\mu} - 1) + 1)/\mu$.

[3] Levent Tunçel. "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form". In: *Foundations of computational mathematics* 1.3 (2001), pp. 229–254.

Primal-Dual Scaling Matrix

Tunçel[3] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T : T \succ 0, T^2 s = x, T^2 F'(x) = F'_*(s)\}$$

and a set of bounded scalings parameterized by ξ:

$$\mathcal{T}_2(x,s,\xi) := \{T \in \mathcal{T}_1(x,s) : (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F F''(x)^{-1}\}$$

where $\delta_F := (\nu(\mu\tilde{\mu} - 1) + 1)/\mu$.

- If $\inf\{\xi : \mathcal{T}_2(x,s,\xi) \neq \emptyset\} \in O(1)$ over all $(x,s) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}^*$, then $O(\sqrt{\nu}\log(1/\epsilon))$ iteration complexity is proved for an infeasible-start primal-dual interior point method

[3] **Tunçel**, "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form".

Primal-Dual Scaling Matrix

Tunçel[3] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T : T \succ 0, T^2 s = x, T^2 F'(x) = F'_*(s)\}$$

and a set of bounded scalings parameterized by ξ:

$$\mathcal{T}_2(x,s,\xi) := \{T \in \mathcal{T}_1(x,s) : (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F F''(x)^{-1}\}$$

where $\delta_F := (\nu(\mu\tilde{\mu} - 1) + 1)/\mu$.

- If $\inf\{\xi : \mathcal{T}_2(x,s,\xi) \neq \emptyset\} \in O(1)$ over all $(x,s) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}^*$, then $O(\sqrt{\nu}\log(1/\epsilon))$ iteration complexity is proved for an infeasible-start primal-dual interior point method
- A self-scaled cone has a unique Nesterov-Todd scaling point w s.t.

  $$s = F''(w)x, \quad F'(x) = F''(w)F'_*(s)$$

  and $F''(w)^{-1} \in \mathcal{T}_2(x,s,4/3)$

---

[3] Tunçel, "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form".

Primal-Dual Scaling Matrix

Tunçel[3] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T : T \succ 0, T^2 s = x, T^2 F'(x) = F'_*(s)\}$$

and a set of bounded scalings parameterized by ξ:

$$\mathcal{T}_2(x,s,\xi) := \{T \in \mathcal{T}_1(x,s) : (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F F''(x)^{-1}\}$$

where $\delta_F := (\nu(\mu\tilde{\mu} - 1) + 1)/\mu$.

- If $\inf\{\xi : \mathcal{T}_2(x,s,\xi) \neq \emptyset\} \in O(1)$ over all $(x,s) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}^*$, then $O(\sqrt{\nu}\log(1/\epsilon))$ iteration complexity is proved for an infeasible-start primal-dual interior point method
- A self-scaled cone has a unique Nesterov-Todd scaling point w s.t.

  $$s = F''(w)x, \quad F'(x) = F''(w)F'_*(s)$$

  and $F''(w)^{-1} \in \mathcal{T}_2(x,s,4/3)$
- But exponential cone is NOT a self-scaled cone

[3] Tunçel, "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form".

Primal-Dual Scaling Matrix

Tunçel[3] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T : T \succ 0, T^2 s = x, T^2 F'(x) = F'_*(s)\}$$

and a set of bounded scalings parameterized by ξ:

$$\mathcal{T}_2(x,s,\xi) := \{T \in \mathcal{T}_1(x,s) : (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F F''(x)^{-1}\}$$

where $\delta_F := (\nu(\mu\tilde{\mu} - 1) + 1)/\mu$.

- If $\inf\{\xi : \mathcal{T}_2(x,s,\xi) \neq \emptyset\} \in O(1)$ over all $(x,s) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}^*$, then $O(\sqrt{\nu}\log(1/\epsilon))$ iteration complexity is proved for an infeasible-start primal-dual interior point method
- A self-scaled cone has a unique Nesterov-Todd scaling point w s.t.

  $$s = F''(w)x, \quad F'(x) = F''(w)F'_*(s)$$

  and $F''(w)^{-1} \in \mathcal{T}_2(x,s,4/3)$
- But exponential cone is NOT a self-scaled cone
- For general nonsymmetric cone,, Tunçel finds $T \in \mathcal{T}_1(x,s)$ by BFGS updates.

[3] **Tunçel**, "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form".

A Quasi-Newton Method with Multiple Secant Equations

Theorem 1 (Existence Theorem[4])

Let $S, Y \in \mathbb{R}^{n\times p}$ have full rank p. Then there exists $H \succ 0$ s.t. $HS = Y$ if and only if $Y^T S \succ 0$.

Implications:

- $H = Y(Y^T S)^{-1} Y^T + ZZ^T, \quad S^T Z = 0, \quad \mathrm{rank} Z = n - p$
- $H^{-1} = S(Y^T S)^{-1} S^T + RR^T, \quad Y^T R = 0, \quad \mathrm{rank} R = n - p$

[4] Robert B Schnabel. ***Quasi-Newton Methods Using Multiple Secant Equations***. Tech. rep. COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE, 1983.

A Quasi-Newton Method with Multiple Secant Equations

Theorem 1 (Existence Theorem[4])

Let $S, Y \in \mathbb{R}^{n \times p}$ have full rank p. Then there exists $H \succ 0$ s.t. $HS = Y$ if and only if $Y^T S \succ 0$.

Implications:

- $H = Y(Y^T S)^{-1} Y^T + ZZ^T, \quad S^T Z = 0, \quad \mathrm{rank} Z = n - p$
- $H^{-1} = S(Y^T S)^{-1} S^T + RR^T, \quad Y^T R = 0, \quad \mathrm{rank} R = n - p$

A result on Quasi-Newton method:

- Given $H \succ 0$, the solution to

$$\begin{array}{ll} \min & \|\Omega^{1/2}(H_+^{-1} - H^{-1})\Omega^{1/2}\|_F \\ \text{s.t.} & H_+^{-1} Y = S \\ & H_+^{-1} \succ 0 \end{array}$$

  with any $\Omega \succ 0$ satisfying $\Omega S = Y$ is

$$H_{BFGS} = Y(Y^T S)^{-1} Y^T + H - HS(S^T HS)^{-1} S^T H$$

[4] **Schnabel**, *Quasi-Newton Methods Using Multiple Secant Equations.*

Compute H_{BFGS} by Sequential Updates

Recall $H_{BFGS} = Y(Y^T S)^{-1} Y^T + H - HS(S^T HS)^{-1} S^T H$

Theorem 2

Given $Y, S \in \mathbb{R}^{n\times p}$ with $Y^T S \succ 0$, then

$$Y(Y^T S)^{-1} Y^T = VV^T, \quad S(Y^T S)^{-1} S^T = UU^T$$

where $V = (v_1, v_2, ..., v_p)$, $U = (u_1, u_2, ..., u_p)$ and

$$\begin{aligned} v_k &:= \frac{Y_{k-1}e_k}{\langle Y_{k-1}e_k, S_{k-1}e_k\rangle^{1/2}}, \quad Y_k := Y_{k-1} - v_k v_k^T S_{k-1} \\ u_k &:= \frac{S_{k-1}e_k}{\langle Y_{k-1}e_k, S_{k-1}e_k\rangle^{1/2}}, \quad S_k := S_{k-1} - u_k u_k^T Y_{k-1} \end{aligned}$$

for any $k = 1, ..., p$ and $Y_0 = Y, S_0 = S$.

Proof. Let $L = \left(\frac{Y_0^T S_0 e_1}{\langle Y_0 e_1, S_0 e_1\rangle^{1/2}}, ..., \frac{Y_{p-1}^T S_{p-1} e_p}{\langle Y_{p-1} e_p, S_{p-1} e_p\rangle^{1/2}} \right)$, one can check $LL^T = Y^T S$ is a Cholesky factorization and $LV^T = Y^T$.

Compute BFGS Scalings

Recall the **double secant equation** $(W^T W)^{-1/2} \in \mathcal{T}_1(x, s)$, i.e.,

$$W^{-T} s = W x, \quad W^{-T} \tilde{s} = W \tilde{x}$$

Let

$$Y := [s, \tilde{s}], \quad H := \mu F''(x), \quad S := [x, \tilde{x}]$$

The goal is find W satisfying double secant equation and $W^T W \approx H$.

Compute BFGS Scalings

Recall the **double secant equation** $(W^TW)^{-1/2} \in \mathcal{T}_1(x,s)$, i.e.,

$$W^{-T}s = Wx, \quad W^{-T}\tilde{s} = W\tilde{x}$$

Let

$$Y := [s, \tilde{s}], \quad H := \mu F''(x), \quad S := [x, \tilde{x}]$$

The goal is find W satisfying double secant equation and $W^TW \approx H$.

- If (x,s) is on central path, then $W^TW = H$ with $H^{-1} \in \mathcal{T}_2(x,s,1)$
- Otherwise, $Y^TS \succ 0$. Based on the quasi-Newton method with multiple secant equation, we get

$$\begin{aligned} W^TW = H_{BFGS} \quad &= Y(Y^TS)^{-1}Y^T + H - HS(S^THS)^{-1}S^TH \\ \text{(by Theorem 2)} \quad &= \left(\tfrac{ss^T}{\langle x,s\rangle} + \tfrac{\delta_s\delta_s^T}{\langle \delta_s,\delta_x\rangle}\right) + H - \left(\tfrac{Hx(Hx)^T}{\langle x,Hx\rangle} + \tfrac{H\rho_x(H\rho_x)^T}{\langle \rho_x,H\rho_x\rangle}\right) \end{aligned}$$

where

$$\delta_x := x - \mu\tilde{x}, \quad \delta_s := s - \mu\tilde{s}, \quad \rho_x := \tilde{x} - \frac{\langle x, H\tilde{x}\rangle}{\langle x, Hx\rangle}x.$$

Compute BFGS Scalings (cont.)

Recall the rank 4 update

$$W^TW = H + \frac{ss^T}{\langle x, s\rangle} + \frac{\delta_s\delta_s^T}{\langle \delta_s, \delta_x\rangle} - \frac{Hx(Hx)^T}{\langle x, Hx\rangle} - \frac{H\rho_x(H\rho_x)^T}{\langle \rho_x, H\rho_x\rangle}$$

is the solution to

$$\begin{array}{ll} \min\limits_{H_+} & \|\Omega^{1/2}(H_+^{-1} - H^{-1})\Omega^{1/2}\|_F \\ \text{s.t.} & H_+^{-1}Y = S \\ & H_+^{-1} \succ 0 \end{array}$$

for any $\Omega \succ 0$ satisfying $\Omega S = Y$. One can assume $\Omega = W^TW$ for simplicity.

Compute BFGS Scalings (cont.)

Recall the rank 4 update

$$W^T W = H + \frac{ss^T}{\langle x, s\rangle} + \frac{\delta_s \delta_s^T}{\langle \delta_s, \delta_x\rangle} - \frac{Hx(Hx)^T}{\langle x, Hx\rangle} - \frac{H\rho_x (H\rho_x)^T}{\langle \rho_x, H\rho_x\rangle}$$

is the solution to

$$\begin{array}{ll} \min\limits_{H_+} & \|\Omega^{1/2}(H_+^{-1} - H^{-1})\Omega^{1/2}\|_F \\ \text{s.t.} & H_+^{-1} Y = S \\ & H_+^{-1} \succ 0 \end{array}$$

for any $\Omega \succ 0$ satisfying $\Omega S = Y$. One can assume $\Omega = W^T W$ for simplicity.

The rank 4 update can be reduced to a rank 3 update after algebraic manipulation.

Compute Optimal Scalings in 3-dimensional Case

If $n = 3$, $p = 2$, we have a method for computing optimal scaling.

By Theorem 1, we know

$W^T W = Y(Y^T S)^{-1} Y^T + t\gamma\gamma^T, \;\; W^{-1} W^{-T} = S(Y^T S)^{-1} S^T + t^{-1} r r^T$

where

$$S^T \gamma = 0, \quad Y^T r = 0, \quad \langle r, \gamma \rangle = 1.$$

Compute Optimal Scalings in 3-dimensional Case

If $n = 3$, $p = 2$, we have a method for computing optimal scaling.

By Theorem 1, we know

$$W^T W = Y(Y^T S)^{-1} Y^T + t\gamma\gamma^T, \quad W^{-1} W^{-T} = S(Y^T S)^{-1} S^T + t^{-1} r r^T$$

where

$$S^T \gamma = 0, \quad Y^T r = 0, \quad \langle r, \gamma \rangle = 1.$$

Assume $\|\gamma\| = 1$ w.l.o.g, then we have cross product formulas

$$\gamma = \frac{x \otimes \tilde{x}}{\|x \otimes \tilde{x}\|}, \quad r = \frac{s \otimes \tilde{s}}{\langle s \otimes \tilde{s}, \gamma \rangle}$$

Compute Optimal Scalings in 3-dimensional Case

If $n = 3$, $p = 2$, we have a method for computing optimal scaling.

By Theorem 1, we know

$W^T W = Y(Y^T S)^{-1} Y^T + t\gamma\gamma^T, \;\; W^{-1}W^{-T} = S(Y^T S)^{-1} S^T + t^{-1} r r^T$

where

$$S^T\gamma = 0, \;\; Y^T r = 0, \;\; \langle r, \gamma \rangle = 1.$$

Assume $\|\gamma\| = 1$ w.l.o.g, then we have cross product formulas

$$\gamma = \frac{x \otimes \tilde{x}}{\|x \otimes \tilde{x}\|}, \;\; r = \frac{s \otimes \tilde{s}}{\langle s \otimes \tilde{s}, \gamma \rangle}$$

Let $Q^T := [r, S]$, one can solve $\inf_\xi \{\xi : (W^T W)^{-1} \in \mathcal{T}_2(x, s, \xi)\}$, or

$$\inf_\xi \left\{ \xi : (\xi\delta_F)^{-1} Q F''(x) Q^T \preceq \left[\begin{array}{cc} t & 0 \\ 0 & Y^T S \end{array} \right] \preceq \xi\delta_F Q F_*''(s)^{-1} Q^T \right\}$$

Compute Optimal Scalings in 3-dimensional Case (cont.)

If $n = 3$, $p = 2$, we have a method for computing optimal scaling.

Solve

$$\inf_{\xi} \left\{ \xi : (\xi\delta_F)^{-1} Q F''(x) Q^T \preceq \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \preceq \xi\delta_F Q F''_*(s)^{-1} Q^T \right\}$$

Define monotonically decreasing function

$$\xi^l(t) := \inf_{\xi} \left\{ \xi : (\xi\delta_F)^{-1} Q F''(x) Q^T \preceq \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \right\}$$

and monotonically increasing function

$$\xi^u(t) := \inf_{\xi} \left\{ \xi : \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \preceq \xi\delta_F Q F''_*(s)^{-1} Q^T \right\}$$

Solve $\xi^l(t) = \xi^u(t)$ by bisection given upper and lower bounds on t.

Comment on the Scalings

- BFGS scaling corresponds to

$$t = \mu \left\| F''(x) - \frac{\tilde{s}\tilde{s}^T}{\nu} - \frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x} \rangle - \nu\tilde{\mu}^2} \right\|_F .$$

Comment on the Scalings

- BFGS scaling corresponds to

$$t = \mu \left\| F''(x) - \frac{\tilde{s}\tilde{s}^T}{\nu} - \frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x} \rangle - \nu\tilde{\mu}^2} \right\|_F .$$

- They tried both optimal scaling and BFGS scaling and report **no** significance difference. Hence they only report BFGS scaling.

Comment on the Scalings

- BFGS scaling corresponds to

$$t = \mu \left\| F''(x) - \frac{\tilde{s}\tilde{s}^T}{\nu} - \frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x} \rangle - \nu\tilde{\mu}^2} \right\|_F.$$

- For exponential cone, they use BFGS scaling by

$$W = \left[\frac{s}{\sqrt{\langle x, s \rangle}}, \frac{\delta_s}{\sqrt{\langle \delta_x, \delta_s \rangle}}, \sqrt{t}\gamma \right]^T, \; W^{-1} = \left[\frac{x}{\sqrt{\langle x, s \rangle}}, \frac{\delta_x}{\sqrt{\langle \delta_x, \delta_s \rangle}}, \frac{r}{\sqrt{t}} \right]^T$$

where

$$\gamma = \frac{x \otimes \tilde{x}}{\|x \otimes \tilde{x}\|}, \quad r = \frac{s \otimes \tilde{s}}{\langle s \otimes \tilde{s}, \gamma \rangle}.$$

Comment on the Scalings

- BFGS scaling corresponds to

$$t = \mu \left\| F''(x) - \frac{\tilde{s}\tilde{s}^T}{\nu} - \frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x} \rangle - \nu\tilde{\mu}^2} \right\|_F .$$

- For exponential cone, they use BFGS scaling by

$$W = \left[\frac{s}{\sqrt{\langle x, s \rangle}}, \frac{\delta_s}{\sqrt{\langle \delta_x, \delta_s \rangle}}, \sqrt{t}\gamma \right]^T, \; W^{-1} = \left[\frac{x}{\sqrt{\langle x, s \rangle}}, \frac{\delta_x}{\sqrt{\langle \delta_x, \delta_s \rangle}}, \frac{r}{\sqrt{t}} \right]^T$$

  where

$$\gamma = \frac{x \otimes \tilde{x}}{\|x \otimes \tilde{x}\|}, \quad r = \frac{s \otimes \tilde{s}}{\langle s \otimes \tilde{s}, \gamma \rangle}.$$

- For all the problems they solved the largest ξ is 1.72 for BFGS scaling

Outline

Affine Search Direction with Scaling

Recall the affine direction is the Newton direction of the KKT system

- Given scaling matrix W satisfying the double secant equations

$$Wx = W^{-T}s, \quad W\tilde{x} = W^{-T}\tilde{s}$$

The linearization of the centrality condition $s + \mu F'(x) = 0$:

$$s + \Delta s + \mu(F'(x) + F''(x)\Delta x) = 0$$

is scaled to (as if $W^TW = \mu F''(x)$)

$$W\Delta x + W^{-T}\Delta s = \mu W^{-T}\tilde{s} - W^{-T}s$$

Affine Search Direction with Scaling

Recall the affine direction is the Newton direction of the KKT system

- Given scaling matrix W satisfying the double secant equations

$$Wx = W^{-T}s, \quad W\tilde{x} = W^{-T}\tilde{s}$$

The linearization of the centrality condition $s + \mu F'(x) = 0$:

$$s + \Delta s + \mu(F'(x) + F''(x)\Delta x) = 0$$

is scaled to (as if $W^TW = \mu F''(x)$)

$$W\Delta x + W^{-T}\Delta s = \mu W^{-T}\tilde{s} - W^{-T}s$$

- The **affine search direction** $\Delta z^a = (\Delta x^a, \Delta s^a, \Delta y^a)$ is defined as the solution of

$$G(\Delta z^a) = -G(z), \quad W\Delta x^a + W^{-T}\Delta s^a = -W^{-T}s$$

Affine Direction

For convenience, let

$$v := Wx = W^{-T}s, \quad \tilde{v} := W\tilde{x} = W^{-T}\tilde{s}$$

The **affine search direction** Δz^a is the solution of

$$G(\Delta z^a) = -G(z), \quad W\Delta x^a + W^{-T}\Delta s^a = -v$$

Lemma 1

The affine direction Δz^a satisfies

$$\langle s, \Delta x^a\rangle + \langle \Delta s^a, x\rangle = -\langle x, s\rangle, \quad \langle \Delta x^a, \Delta s^a\rangle = 0$$

Proof.

$$\begin{aligned} \langle s, \Delta x^a\rangle + \langle \Delta s^a, x\rangle &= \langle W^{-T}s, W\Delta x^a\rangle + \langle W^{-T}\Delta s^a, Wx\rangle \\ &= \langle v, W\Delta x^a + W^{-T}\Delta s^a\rangle = -\langle v, v\rangle \\ &= \langle Wv, W^{-T}v\rangle = -\langle x, s\rangle \end{aligned}$$

$\langle x + \Delta x^a, s + \Delta s^a\rangle = 0$ by skew-symmetry of matrix in HSD

Corrector Direction: High-order error of Linearization

Consider the 1st and 2nd order derivatives of centrality condition

$$s_\mu + \mu F'(x_\mu) = 0$$

w.r.t. μ:

$$\begin{aligned} &\dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu + F'(x_\mu) = 0 \\ &\ddot{s}_\mu + 2F''(x_\mu)\dot{x}_\mu + \mu(F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] + F''(x_\mu)\ddot{x}_\mu) = 0 \end{aligned} \tag{5}$$

which together with equation (1) implies

$$\mu\dot{x}_\mu = -F''(x_\mu)^{-1}(\dot{s}_\mu + F'(x_\mu)) = x_\mu - F''(x_\mu)^{-1}\dot{s}_\mu$$

Corrector Direction: High-order error of Linearization

Consider the 1st and 2nd order derivatives of centrality condition

$$s_\mu + \mu F'(x_\mu) = 0$$

w.r.t. μ:

$$\begin{aligned} &\dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu + F'(x_\mu) = 0 \\ &\ddot{s}_\mu + 2F''(x_\mu)\dot{x}_\mu + \mu(F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] + F''(x_\mu)\ddot{x}_\mu) = 0 \end{aligned} \tag{5}$$

which together with equation (1) implies

$$\mu\dot{x}_\mu = -F''(x_\mu)^{-1}(\dot{s}_\mu + F'(x_\mu)) = x_\mu - F''(x_\mu)^{-1}\dot{s}_\mu$$

and

$$\begin{aligned} \mu F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] &= F'''(x_\mu)[\dot{x}_\mu, x_\mu] - F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu] \\ &= -2F''(x_\mu)\dot{x}_\mu - F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu] \end{aligned}$$

Corrector Direction: High-order error of Linearization

Consider the 1st and 2nd order derivatives of centrality condition

$$s_\mu + \mu F'(x_\mu) = 0$$

w.r.t. μ:

$$\begin{aligned} &\dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu + F'(x_\mu) = 0 \\ &\ddot{s}_\mu + 2F''(x_\mu)\dot{x}_\mu + \mu(F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] + F''(x_\mu)\ddot{x}_\mu) = 0 \end{aligned} \tag{5}$$

which together with equation (1) implies

$$\mu\dot{x}_\mu = -F''(x_\mu)^{-1}(\dot{s}_\mu + F'(x_\mu)) = x_\mu - F''(x_\mu)^{-1}\dot{s}_\mu$$

and

$$\begin{aligned} \mu F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] &= F'''(x_\mu)[\dot{x}_\mu, x_\mu] - F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu] \\ &= -2F''(x_\mu)\dot{x}_\mu - F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu] \end{aligned}$$

Hence

$$\ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu = F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu]$$

Corrector Direction (cont.)

Recall centrality condition

$$s_\mu + \mu F'(x_\mu) = 0$$

and

$$\begin{aligned}\dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu &= -F'(x_\mu)\\ \ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu &= F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu]\end{aligned}$$

Linearization $s + \Delta s + \mu F'(x) + \mu F''(x)\Delta x$ leads to high-order error:

$$-\frac{1}{2}F'''(x)[\Delta x, F''(x)^{-1}\Delta s]$$

Corrector Direction (cont.)

Recall centrality condition

$$s_\mu + \mu F'(x_\mu) = 0$$

and

$$\begin{aligned} \dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu &= -F'(x_\mu) \\ \ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu &= F'''(x_\mu)[\dot{x}_\mu, F''(x_\mu)^{-1}\dot{s}_\mu] \end{aligned}$$

Linearization $s + \Delta s + \mu F'(x) + \mu F''(x)\Delta x$ leads to high-order error:

$$-\frac{1}{2}F'''(x)[\Delta x, F''(x)^{-1}\Delta s]$$

The **high-order pure corrector direction** Δz^c is obtained by solving

$$G(\Delta z^c) = 0, \quad W\Delta x^c + W^{-T}\Delta s^c = -W^{-T}\eta$$

where $\eta := -\frac{1}{2}F'''(x)[\Delta x^a, F''(x)^{-1}\Delta s^a]$

Corrector Direction with Scaling

The **high-order pure corrector direction** Δz^c is obtained by

$$G(\Delta z^c) = 0, \quad W\Delta x^c + W^{-T}\Delta s^c = -W^{-T}\eta$$

where $\eta = -\frac{1}{2}F'''(x)[\Delta x^a, F''(x)^{-1}\Delta s^a]$.

Lemma 2

The pure corrector direction Δz^c satisfies

$$\langle s, \Delta x^c\rangle + \langle \Delta s^c, x\rangle = 0, \quad \langle \Delta x^c, \Delta s^c\rangle = 0$$

Proof.

$$\begin{aligned}
\langle s, \Delta x^c\rangle + \langle \Delta s^c, x\rangle \quad &= \langle W^{-T}s, W\Delta x^c\rangle + \langle W^{-T}\Delta s^c, Wx\rangle \\
&= \langle v, W\Delta x^c + W^{-T}\Delta s^c\rangle \\
&= \langle v, -\tfrac{1}{2}W^{-T}F'''(x)[\Delta x^a, F''(x)^{-1}\Delta s^a]\rangle \\
&= \langle x, -\tfrac{1}{2}F'''(x)[\Delta x^a, F''(x)^{-1}\Delta s^a]\rangle \\
&= \langle -2F''(x)\Delta x^a, -\tfrac{1}{2}F''(x)^{-1}\Delta s^a\rangle \\
&= \langle \Delta x^a, \Delta s^a\rangle = 0
\end{aligned}$$

Skew-symmetry implies $\langle \Delta x^c, \Delta s^c\rangle = 0$.

Combined Centering Search Direction

Given centering parameter $\sigma > 0$, the combined centering search direction is obtained by

$$G(\Delta z) = -(1-\sigma)G(z), \quad W\Delta x + W^{-T}\Delta s = -v + \sigma\mu\tilde{v} - W^{-T}\eta$$

Lemma 3

The combined search direction Δz *satisfies*

$$\langle s, \Delta x\rangle + \langle \Delta s, x\rangle = -(1-\sigma)\langle x, s\rangle, \quad \langle \Delta x, \Delta s\rangle = 0$$

Proof. Recall $\mu = \langle x, s\rangle/\nu$.

$$\begin{aligned}
\langle s, \Delta x\rangle + \langle \Delta s, x\rangle \quad &= \langle W^{-T}s, W\Delta x\rangle + \langle W^{-T}\Delta s, Wx\rangle \\
&= \langle v, -v + \sigma\mu\tilde{v} - W^{-T}\eta\rangle \\
&= -\langle x, s\rangle + \sigma\mu\langle x, \tilde{s}\rangle - \langle v, W^{-T}\eta\rangle \\
&= -(1-\sigma)\langle x, s\rangle + 0
\end{aligned}$$

Skew-symmetry implies $\langle (1-\sigma)x + \Delta x, (1-\sigma)s + \Delta s\rangle = 0$. Hence $\langle \Delta x, \Delta s\rangle = 0$.

Implication of Combined Centering Search Direction

For all $\alpha \in \mathbb{R}$, we have

$$\begin{aligned} G(z + \alpha \Delta z) &= (1 - \alpha(1 - \sigma))G(z) \\ \langle x + \alpha \Delta x, s + \alpha \Delta s \rangle &= (1 - \alpha(1 - \sigma))\langle x, s \rangle \end{aligned}$$

The residuals and complementarity gap decrease at the same rate:
In iteration k,

$$\begin{aligned} G(z^k) &= \mu^k G(z^0) \\ \langle x^k, s^k \rangle &= \mu^k \nu \end{aligned}$$

where $\mu^k := \langle x^k, s^k \rangle / \nu$.

Implication of Combined Centering Search Direction

For all $\alpha \in \mathbb{R}$, we have

$$\begin{aligned} G(z + \alpha \Delta z) &= (1 - \alpha(1 - \sigma))G(z) \\ \langle x + \alpha \Delta x, s + \alpha \Delta s \rangle &= (1 - \alpha(1 - \sigma))\langle x, s \rangle \end{aligned}$$

The residuals and complementarity gap decrease at the same rate:
In iteration k,

$$\begin{aligned} G(z^k) &= \mu^k G(z^0) \\ \langle x^k, s^k \rangle &= \mu^k \nu \end{aligned}$$

where $\mu^k := \langle x^k, s^k \rangle / \nu$.

Tunçel proves the algorithm (without corrector) has polynomial complexity if the scaling matrix is bounded, which is not proved in the paper.

Mehrotra Like Search Direction

- Compute affine search direction Δz^a via

$$G(\Delta z^a) = -G(z), \quad W\Delta x^a + W^{-T}\Delta s^a = -Wx,$$

the corrector term

$$\eta := -\frac{1}{2}F'''(x)[\Delta x^a, F''(x)^{-1}\Delta s^a],$$

the stepsize α_a

$$\alpha_a := \sup\{\alpha \mid x + \alpha\Delta x^a \in \mathcal{K}, s + \alpha\Delta s^a \in \mathcal{K}^*, \alpha \in [0,1]\}$$

and the centering parameter

$$\sigma := (1-\alpha_a)\min\{(1-\alpha_a)^2, 1/4\}$$

- Compute the combined centering-corrector search direction Δz via

$$G(\Delta z) = -(1-\sigma)G(z), \; W\Delta x + W^{-T}\Delta s = -Wx + \sigma\mu W\tilde{x} - W^{-T}\eta,$$

and update $z = z + \alpha\Delta z$ with the largest stepsize inside the central path neighborhood $\mathcal{N}_\beta$ where $\beta = 10^{-6}$.

Outline

Implementation

In this numerical part, we shall release τ and κ from x and s.

- Dualization, Presolve and Scaling
- Compute Search Direction
  - Eliminate Δs and $\Delta\kappa$ from linearized centrality condition and get

$$\begin{bmatrix} W^TW & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & \tau^{-1}\kappa \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta\tau \end{bmatrix} = \text{RHS}$$

  - Solve the system based on LDLT factorization of symmetric matrix $\begin{bmatrix} -W^TW & A^T \\ A & 0 \end{bmatrix}$ with possible diagonal modification.

Stopping Criteria

- Optimality condition:

$$\begin{cases} \dfrac{\left\|Ax^k/\tau^k - b\right\|_\infty}{1+\|b\|_\infty} \le \epsilon_p \\ \dfrac{\left\|A^T y^k/\tau^k + s^k/\tau^k - c\right\|_\infty}{1+\|c\|_\infty} \le \epsilon_d \\ \dfrac{\min\left\{\langle x^k/\tau^k, s^k/\tau^k\rangle, \left|\langle c, x^k/\tau^k\rangle - \langle b, y^k/\tau^k\rangle\right|\right\}}{\max\left\{1, \min\left\{\left|\langle c, x^k/\tau^k\rangle\right|, \left|\langle b, y^k/\tau^k\rangle\right|\right\}\right\}} \le \epsilon_g \end{cases}$$

- Infeasibility measure:

$$\frac{\|A^T y^k + s^k\|}{\langle b, y^k\rangle} \le \epsilon_i, \quad \langle b, y^k\rangle > 0$$
$$\frac{\|Ax^k\|_\infty}{\langle c, x^k\rangle} \le -\epsilon_i, \quad \langle c, x^k\rangle < 0$$

- Ill-posedness:

$$\left\| \begin{array}{c} A^T y^k + s^k \\ Ax^k \end{array} \right\|_\infty \le \epsilon_i \left\| \begin{array}{c} y^k \\ s^k \\ x^k \end{array} \right\|_\infty, \quad \left\| \begin{array}{c} y^k \\ s^k \\ x^k \end{array} \right\|_\infty > 0$$

Comparision with ECOS[5]

Numnerical instances are from http://cblib.zib.de/

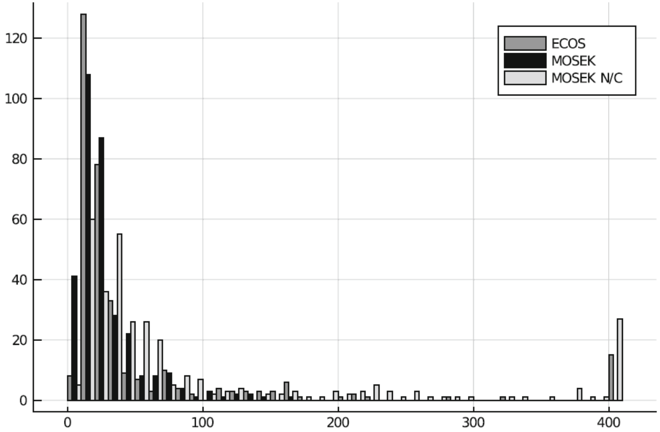

Fig. 2 Histograms of solver iterations for 326 test problems with exponential cones. MOSEK (including corrector) succesfully solved all instances and generally required the fewest iterations. MOSEK N/C (no corrector) solved 300 instances, and ECOS solved 201 instances. The number of iterations was limited to 400 in all solvers

[5] Santiago Akle Serrano. *Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone*. Stanford University, 2015.

Concluding Remarks

- Combination of infeasible-start primal-dual interior point methods[6] and scaling matrix[7] for nonsymmetric cone
- A new high-order corrector for nonsymmetric cones — generalization of Mehrotra like predictor-corrector
- Theoretical complexity analysis is not complete.

Thank You! Questions?

[6] **Nesterov, Todd, and Ye**, "Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems".

[7] **Tunçel**, "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form".