Primal-dual IPM with Asymmetric Barrier

Yurii Nesterov, CORE/INMA (UCL)

September 29, 2008 (IFOR, ETHZ)

Outline

Nonlinear Conic Optimization Problem

Primal-Dual Problem:

$$\min_x\{\langle c, x\rangle : \; Ax = b, \; x \in K\} \stackrel{?}{=} \max_{s,y}\{\langle b, y\rangle : \; s + A^* y = c, \; s \in K^*\},$$

where K is a *normal* cone (convex, pointed, solid), and

$$K^* = \{s : \langle s, x\rangle \geq 0, x \in K\}.$$

Main Assumptions:

1. $\exists \, x_0 \in \mathrm{int}\, K, \; s_0 \in \mathrm{int}\, K^*, \; y_0 \in H : \; Ax = b, \; s + A^* y_0 = c.$
2. $F_P(x)$ is a ν_{F_P}-normal barrier for K (self-concordant, log-homogeneous: $F_P(\tau x) = F(x) - \nu_{F_P} \ln \tau$, $x \in \mathrm{int}\, K$)

Nonlinear Conic Optimization Problem

Primal-Dual Problem:

$$\min_x\{\langle c,x\rangle :\ Ax=b,\ x\in K\} \overset{?}{=} \max_{s,y}\{\langle b,y\rangle :\ s+A^*y=c,\ s\in K^*\},$$

where K is a *normal* cone (convex, pointed, solid), and

$$K^* = \{s : \langle s,x\rangle \geq 0, x\in K\}.$$

Main Assumptions:

1. $\exists\, x_0 \in \mathrm{int}\, K,\ s_0 \in \mathrm{int}\, K^*,\ y_0 \in H :\ Ax=b,\ s+A^*y_0=c.$
2. $F_P(x)$ is a ν_{F_P}-normal barrier for K (self-concordant, log-homogenerous: $F_P(\tau x) = F(x) - \nu_{F_P}\ln\tau$, $x\in\mathrm{int}\, K$)
3. **Important:** we can compute the *dual barrier*

$$F_D(s) \quad = \quad F_P^*(s) \overset{\mathrm{def}}{=} \max_x[-\langle s,x\rangle - F_P(x)]$$

Examples

Positive orthant

$$
\begin{aligned}
K &= \{x \in R^n : \; x \geq 0\} \quad (= K^*), \\
F_P(x) &= -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu_P = n.
\end{aligned}
$$

Examples

Positive orthant

$$\begin{aligned} K &= \{x \in R^n : \ x \geq 0\} \quad (= K^*), \\ F_P(x) &= -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu_P = n. \end{aligned}$$

Lorentz cone

$$\begin{aligned} K &= \{(\tau, x) \in R^{n+1} : \ \tau \geq \|x\|_2\} \quad (= K^*), \\ F_P(\tau, x) &= -\ln(\tau^2 - \|x\|_2^2), \quad \nu_P = 2. \end{aligned}$$

Examples

Positive orthant

$$\begin{aligned} K &= \{x \in R^n : \; x \geq 0\} \quad (= K^*), \\ F_P(x) &= -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu_P = n. \end{aligned}$$

Lorentz cone

$$\begin{aligned} K &= \{(\tau, x) \in R^{n+1} : \; \tau \geq \|x\|_2\} \quad (= K^*), \\ F_P(\tau, x) &= -\ln(\tau^2 - \|x\|_2^2), \quad \nu_P = 2. \end{aligned}$$

Cone of positive-semidefinite matrices

$$\begin{aligned} K &= \{X \in \mathcal{S}_n : \; X \succeq 0\} \quad (= K^*), \\ F_P(X) &= -\ln \det X, \quad \nu_P = n. \end{aligned}$$

Symmetric primal-dual IPM $\quad (F_D = F_P^*)$

Primal-dual central path: $(x(t), s(t), y(t)) \in K \times K^* \times H$:

$$\begin{aligned} x(t) &= \arg\min_{Ax=b} [t\langle c, x\rangle + F_P(x)], \quad t > 0, \\ y(t) &= \arg\max_{y} [t\langle b, y\rangle - F_D(c - A^T y)]. \end{aligned}$$

Symmetric primal-dual IPM $\quad (F_D = F_P^*)$

Primal-dual central path: $(x(t), s(t), y(t)) \in K \times K^* \times H$:

$$\begin{array}{rcl} x(t) & = & \arg\min\limits_{Ax=b} [t\langle c, x\rangle + F_P(x)], \quad t > 0, \\ y(t) & = & \arg\max\limits_{y} [t\langle b, y\rangle - F_D(c - A^T y)]. \end{array}$$

Many important identities: $\quad Ax(t) = b, \quad s(t) + A^* y(t) = c,$

$$s(t) = -\tfrac{1}{t}\nabla F_P(x(t)), \quad x(t) \; = \; -\tfrac{1}{t}\nabla F_D(s(t)),$$

$$\overbrace{\langle c, x(t)\rangle - \langle b, y(t)\rangle}^{unbounded} \; = \; \tfrac{1}{t}\nu_{F_P}, \quad F_P(x(t)) + F_D(s(t)) \; = \; \nu - \nu \ln t.$$

Symmetric primal-dual IPM $\quad (F_D = F_P^*)$

Primal-dual central path: $(x(t), s(t), y(t)) \in K \times K^* \times H$:

$$
\begin{array}{rcl}
x(t) & = & \arg\min\limits_{Ax=b} [t\langle c, x\rangle + F_P(x)], \quad t > 0, \\
y(t) & = & \arg\max\limits_{y} [t\langle b, y\rangle - F_D(c - A^T y)].
\end{array}
$$

Many important identities: $\quad Ax(t) = b, \quad s(t) + A^* y(t) = c,$

$$
s(t) = -\tfrac{1}{t}\nabla F_P(x(t)), \quad x(t) \; = \; -\tfrac{1}{t}\nabla F_D(s(t)),
$$

$$
\overbrace{\langle c, x(t)\rangle - \langle b, y(t)\rangle}^{unbounded} \; = \; \tfrac{1}{t}\nu_{F_P}, \quad F_P(x(t)) + F_D(s(t)) \; = \; \nu - \nu \ln t.
$$

Consequences:

- Good search directions (primal-dual affine-scaling, centering)

Symmetric primal-dual IPM $\quad (F_D = F_P^*)$

Primal-dual central path: $(x(t), s(t), y(t)) \in K \times K^* \times H$:

$$\begin{aligned} x(t) &= \arg\min_{Ax=b} [t\langle c, x\rangle + F_P(x)], \quad t > 0, \\ y(t) &= \arg\max_{y} [t\langle b, y\rangle - F_D(c - A^T y)]. \end{aligned}$$

Many important identities: $\quad Ax(t) = b, \quad s(t) + A^* y(t) = c,$

$$s(t) = -\tfrac{1}{t}\nabla F_P(x(t)), \quad x(t) = -\tfrac{1}{t}\nabla F_D(s(t)),$$

$$\overbrace{\langle c, x(t)\rangle - \langle b, y(t)\rangle}^{unbounded} = \tfrac{1}{t}\nu_{F_P}, \quad F_P(x(t)) + F_D(s(t)) = \nu - \nu \ln t.$$

Consequences:

- Good search directions (primal-dual affine-scaling, centering)
- Long-step IPM, infeasible-start IPM.

Symmetric primal-dual IPM $\quad (F_D = F_P^*)$

Primal-dual central path: $(x(t), s(t), y(t)) \in K \times K^* \times H$:

$$\begin{aligned} x(t) &= \arg\min_{Ax=b} [t\langle c, x\rangle + F_P(x)], \quad t > 0, \\ y(t) &= \arg\max_{y} [t\langle b, y\rangle - F_D(c - A^T y)]. \end{aligned}$$

Many important identities: $\quad Ax(t) = b, \quad s(t) + A^* y(t) = c,$

$$s(t) = -\tfrac{1}{t}\nabla F_P(x(t)), \quad x(t) = -\tfrac{1}{t}\nabla F_D(s(t)),$$

$$\overbrace{\langle c, x(t)\rangle - \langle b, y(t)\rangle}^{unbounded} = \tfrac{1}{t}\nu_{F_P}, \quad F_P(x(t)) + F_D(s(t)) = \nu - \nu \ln t.$$

Consequences:

- Good search directions (primal-dual affine-scaling, centering)
- Long-step IPM, infeasible-start IPM.
- Powerful software (SeDuMi, T3, Mosek, etc.)

Another applications

1. Flows in fluid networks

$$\begin{aligned} &\min_f \sum_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\ \text{s.t.} \quad & E \cdot f = d, \quad (\text{node balance}) \\ & f_i \geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

Another applications

1. Flows in fluid networks

$$
\begin{aligned}
& \min_f \sum_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\
\text{s.t.} \quad & E \cdot f = d, \quad (\text{node balance}) \\
& f_i \geq 0, \quad i \in \mathcal{I}.
\end{aligned}
$$

Electricity: $\gamma = 2$.

Another applications

1. Flows in fluid networks

$$\begin{array}{rl} \min\limits_{f} & \sum\limits_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\ \text{s.t.} & E \cdot f = d, \quad (\text{node balance}) \\ & f_i \geq 0, \quad i \in \mathcal{I}. \end{array}$$

Electricity: $\gamma = 2$. Gas networks: $\gamma = 3$.

Another applications

1. Flows in fluid networks

$$\begin{aligned} & \min_f \sum_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\ \text{s.t.} \quad & E \cdot f = d, \quad (\text{node balance}) \\ & f_i \geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

Electricity: $\gamma = 2$. Gas networks: $\gamma = 3$. Water: $\gamma = 2.75$.

Another applications

1. Flows in fluid networks

$$\begin{aligned} & \min_f \sum_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\ \text{s.t.} \quad & E \cdot f = d, \quad (\text{node balance}) \\ & f_i \geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

Electricity: $\gamma = 2$. Gas networks: $\gamma = 3$. Water: $\gamma = 2.75$.
Usually, the dimension is not too big!

Another applications

1. Flows in fluid networks

$$
\begin{aligned}
& \min_f \sum_{\alpha \in \mathcal{A}} c_\alpha |f_\alpha|^\gamma \quad (= \text{Energy}) \\
\text{s.t.} \quad & E \cdot f = d, \quad (\text{node balance}) \\
& f_i \geq 0, \quad i \in \mathcal{I}.
\end{aligned}
$$

Electricity: $\gamma = 2$. Gas networks: $\gamma = 3$. Water: $\gamma = 2.75$.
Usually, the dimension is not too big!

2. Random scheduling: minimize in x the objective

$$
\langle T, x \rangle + \sum_{i=1}^{n} \sum_{j=i+1}^{n} d^{(i)} d^{(j)} \ln \left(e^{x^{(i)}/d^{(i)}} + e^{x^{(j)}/d^{(j)}} \right) - D \langle \bar{e}, x \rangle.
$$

Asymmetric barriers:

1. Power cone

$$K_\alpha \;=\; \left\{ x \in R^2_+ \times R:\; \left(x^{(1)}\right)^\alpha \left(x^{(2)}\right)^{1-\alpha} \geq \left| x^{(3)} \right| \right\}, \quad \alpha \in (0,1),$$

$$F_{P,\alpha}(x) \;=\; -\ln\left[\left(x^{(1)}\right)^{2\alpha} \left(x^{(2)}\right)^{2(1-\alpha)} - \left(x^{(3)}\right)^2 \right] - \ln x^{(1)} - \ln x^{(2)},$$

$$K^*_\alpha \;=\; \left\{ s \in R^2_+ \times R:\; \left(\frac{s^{(1)}}{\alpha}\right)^\alpha \left(\frac{s^{(2)}}{1-\alpha}\right)^{1-\alpha} \geq \left| s^{(3)} \right| \right\},$$

$$F_{D,\alpha}(s) \;=\; -\ln\left[\left(\frac{s^{(1)}}{\alpha}\right)^{2\alpha} \left(\frac{s^{(2)}}{1-\alpha}\right)^{2(1-\alpha)} - \left(s^{(3)}\right)^2 \right] - \ln s^{(1)} - \ln s^{(2)},$$

with $\nu_{F_{P,\alpha}} = \nu_{F_{D,\alpha}} = 4$. $K^*_\alpha = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 1-\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} K_\alpha$, but $F_{D,\alpha} \neq F^*_{P,\alpha}$.

Asymmetric barriers:

2. Conic hull of the epigraph of exponent

$$K_{P,e} \;=\; \left\{ x \in R \times R^2_+ : \; x^{(1)} \geq x^{(2)} \ln \tfrac{x^{(2)}}{x^{(3)}} \right\},$$

$$F_{P,e}(x) \;=\; -\ln \left(x^{(1)} - x^{(2)} \ln \tfrac{x^{(2)}}{x^{(3)}} \right) - \ln x^{(2)} - \ln x^{(3)},$$

$$K_{D,e} \;=\; \left\{ s \in R_+ \times R \times R_+ : \; s^{(1)} + s^{(2)} \geq s^{(1)} \ln \tfrac{s^{(1)}}{s^{(3)}} \right\},$$

$$F_{D,e}(x) \;=\; -\ln \left(s^{(1)} + s^{(2)} - s^{(1)} \ln \tfrac{s^{(1)}}{s^{(3)}} \right) - \ln s^{(1)} - \ln s^{(3)},$$

with $\nu_{F_{P,e}} = \nu_{F_{D,e}} = 3$. $K_{D,e} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} K_{P,e}$, and $F_{D,e} \neq F^*_{P,e}$.

Asymmetric barrier for primal-dual cone

Denote

$$\begin{aligned} \Psi(z) &\equiv \Psi(x,s) = F_P(x) + F_D(s), \quad \nu_\Psi = \nu_{F_P} + \nu_{F_D}, \\ z &= (x,s) \in \operatorname{int} K \times \operatorname{int} K^* \stackrel{\mathrm{def}}{=} \operatorname{int} \bar{K}. \end{aligned}$$

Define

$$\kappa(z_0) = \Psi(z_0) - \min_{x,s} \left\{ \Psi(z) : \langle s_0, x \rangle + \langle s, x_0 \rangle = 2\langle s_0, x_0 \rangle \right\},$$

where $z_0 = (x_0, s_0)$ is the point from A1. Note that $\kappa(z_0) \geq 0$.

Lemma: For any $z \in \bar{K}$ we have

$$\Psi(z) \geq \Psi(z_0) - \kappa(z_0) - \nu_\Psi \ln \frac{\langle s_0,x\rangle + \langle s,x_0\rangle}{2\langle s_0,x_0\rangle}.$$

Feasible-start potential reduction IPM

Primal-dual problem:

$$w \;=\; (x,s,y,\tau) \in K \times K^* \times H \times R_+ :$$
$$Ax = \tau b, \;\; s + A^* y = \tau c,$$
$$\text{Normalization constraint: } \tau = 1.$$
$$\text{Optimal value of the objective: } \langle c, x\rangle - \langle b, y\rangle = 0.$$

Karmarkar setting: For cone $\mathcal{C}$ with ν-normal barrier $F(x)$, consider

$$\min_{w \in C} \{ \langle d, w\rangle : \; Bw = 0, \; \langle e, w\rangle = 1 \} \;=\; 0.$$

Can be solved by minimizing homogeneous potential

$$\nu \ln \langle d, w\rangle + F(w).$$

Condition: the set $\{ w \in \mathcal{C} : Bw = 0, \; \langle e, w\rangle = 1 \}$ is *bounded*.

Feasible-start potential reduction IPM

Primal-dual problem:

$$w \;=\; (x,s,y,\tau) \in K \times K^* \times H \times R_+ :$$
$$Ax = \tau b,\;\; s + A^* y = \tau c,$$
$$\text{Normalization constraint: } \tau = 1.$$
$$\text{Optimal value of the objective: } \langle c, x\rangle - \langle b, y\rangle = 0.$$

Karmarkar setting: For cone $\mathcal{C}$ with ν-normal barrier $F(x)$, consider

$$\min_{w \in C} \{ \langle d, w\rangle :\; Bw = 0,\; \langle e, w\rangle = 1 \} \;=\; 0.$$

Can be solved by minimizing homogeneous potential

$$\nu \ln\langle d, w\rangle + F(w).$$

Condition: the set $\{ w \in \mathcal{C} : Bw = 0,\; \langle e, w\rangle = 1\}$ is *bounded*.

NOTE: the set $\{(x,s,y) \in K \times K^* \times H : Ax = b, s + A^* y = c\}$ is never bounded

Feasible-start potential reduction IPM

$$
\begin{array}{rcl}
\text{Denote } \mathcal{F} & = & \{w = (x, s, y, \tau) \in K \times K^* \times H \times R_+ : \\
& & Ax = \tau b, \; s + A^* y = \tau c\}, \\
F(w) & = & F_P(x) + F_D(s) - \ln \tau, \quad \nu_F = \nu_P + \nu_D + 1, \\
\phi(w) & = & \nu_F \ln[\langle c, x\rangle - \langle b, y\rangle] + F(w), \quad w \in \mathrm{int}\, \mathcal{F}.
\end{array}
$$

Main identity: for $w \in \mathcal{F}$ we have

$$\langle s_0, x\rangle + \langle s, x_0\rangle \; = \; \langle c, x\rangle - \langle b, y\rangle + \tau\langle s_0, x_0\rangle.$$

Theorem: Let $w_0 \equiv (x_0, s_0, y_0, 1)$. If we form $w_k = (x_k, s_k, y_k, \tau_k)$ by minimizing $\phi(w)$ by the Newton method, then for the point

$$\tilde{x}_k = \frac{x_k}{\tau_k}, \quad \tilde{s}_k = \frac{s_k}{\tau_k}, \quad \tilde{y}_k = \frac{y_k}{\tau_k}$$

we have

$$\frac{\langle s_0, x_0\rangle}{\langle c, \tilde{x}_k\rangle - \langle b, \tilde{y}_k\rangle} \; \geq \; 2 \exp\left\{ \frac{\omega \cdot k - \kappa(z_0) - \ln 2}{\nu_F} \right\} - 1.$$

Feasible-start potential reduction IPM

Advantages:

- Polynomial-time $O(\nu_F \ln \frac{1}{\epsilon})$ complexity bound.
- Asymmetric barrier $F(w)$.

Feasible-start potential reduction IPM

Advantages:

- Polynomial-time $O(\nu_F \ln \frac{1}{\epsilon})$ complexity bound.
- Asymmetric barrier $F(w)$.

Disadvantages:

- The worst-case complexity bound is not the best one.
- Feasible point is needed.
- Unavoidable expensive exact matrix operations.

Full-dimensional potential reduction IPM

Denote $\mathcal{C} \equiv \{w = (x, s, y, \tau) \in K \times K^* \times H \times R_+\}$.
By two positive-definite operators $B_H : \ H \to H^*$, $B_E : \ E \to E^*$, define the convex quadratic function

$$\begin{aligned}\langle Qw, w\rangle \;&=\; \|Ax - \tau b\|^2_{B_H^{-1}} + \|s + A^* y - \tau c\|^2_{B_E^{-1}} \\ &\quad + [\langle c, x\rangle - \langle b, y\rangle]^2 \;\stackrel{\mathrm{def}}{=}\; \|w\|^2_Q,\end{aligned}$$

Optimal set: $W^* = \{w \in \mathcal{C} : \ Qw = 0\}$.

Main inequality:

For any $w \in \mathcal{C}$ we have

$$\begin{aligned}\langle s_0, x\rangle + \langle s, x_0\rangle \;&\leq\; \overbrace{[1 + \langle B_H y_0, y_0\rangle + \langle B_E x_0, x_0\rangle]^{1/2}}^{\stackrel{\mathrm{def}}{=}\Omega} \cdot \|w\|_Q \\ &\quad + \tau \cdot \langle s_0, x_0\rangle,\end{aligned}$$

Quadratic potential function

$$\text{Denote} \quad F(w) = \Psi(z) - \ln\tau = F_P(x) + F_D(s) - \ln\tau,$$

$$\Phi(w) = \tfrac{1}{2}\|w\|_Q^2 + F(w), \quad w \in \mathrm{int}\,\mathcal{C}.$$

Main properties

- $\Phi(w)$ is a s.-c. function with positive definite Hessians.
- It is unbounded from below. Hence, the local norms of the gradients are all ≥ 1.
- For $w \in \mathrm{int}\,\mathcal{C}$, define *homogeneous* quadratic potential:

$$\widetilde{\Phi}(w) = \min_{\lambda > 0} \Phi(\lambda w) \quad \leq \Phi(w).$$

The minimum is attained at $\lambda = \lambda(w) \stackrel{\text{def}}{=} \frac{\nu_F^{1/2}}{\|w\|_Q}$.

Properties of homogeneous potential

- It admits a closed-form representation:

$$\widetilde{\Phi}(w) \quad = \quad \nu_F \ln \|w\|_Q + F(w) + \tfrac{\nu_F}{2}[1 - \ln \nu_F].$$

- This is a quasi-convex function with zero degree of homogeneity.
- For any point $w \in \operatorname{int} \mathcal{C}$, we have

$$\widetilde{\Phi}(w) \quad \geq \quad \Psi(z_0) - \kappa(z_0) + \nu_\Psi \ln 2 + \tfrac{\nu_F}{2}[1 - \ln \nu_F]$$

$$-\nu_F \ln \left(\tfrac{\Omega}{\langle s_0, x_0 \rangle} + \tfrac{\tau}{\|w\|_Q} \right).$$

Possible strategies

A. Minimization of quadratic potential

1. Compute point $\bar{w}_k$ by a Newton step from w_k.
2. Define $w_{k+1} = \lambda(w_k) \cdot w_k$.

B. Minimization of homogeneous potential

At each iteration, apply Newton step to an upper convex approximation of $\widetilde{\Phi}$.

In both cases, $\widetilde{\Phi}$ is decreased by an absolute constant (at least).

Possible strategies

A. Minimization of quadratic potential

1. Compute point $\bar{w}_k$ by a Newton step from w_k.
2. Define $w_{k+1} = \lambda(w_k) \cdot w_k$.

B. Minimization of homogeneous potential

At each iteration, apply Newton step to an upper convex approximation of $\widetilde{\Phi}$.

In both cases, $\widetilde{\Phi}$ is decreased by an absolute constant (at least).

New even for LP!

Full-dimensional potential reduction IPM

Advantages:

- Polynomial-time $O(\nu_F \ln \frac{1}{\epsilon})$ complexity bound.
- Asymmetric barrier $F(w)$.
- Can start from infeasible points*.
- Search directions may be computed by gradient schemes with reasonable accuracy*.

Full-dimensional potential reduction IPM

Advantages:

- Polynomial-time $O(\nu_F \ln \frac{1}{\epsilon})$ complexity bound.
- Asymmetric barrier $F(w)$.
- Can start from infeasible points*.
- Search directions may be computed by gradient schemes with reasonable accuracy*.

Disadvantages:

- The worst-case complexity bound is not the best one.
  However, the long steps can accelerate the convergence.

Central path for infeasible-start IPM

Denote $v = (x, s, \tau) \in \widehat{\mathcal{C}} \stackrel{\mathrm{def}}{=} K \times K^* \times R_+$, and

$$\begin{array}{rcl} \langle \hat{Q}v, v\rangle & \stackrel{\mathrm{def}}{=} & \min\limits_{y}\{\langle Qw, w\rangle :\ w = (x, s, y, \tau)\}, \\ y(v) & \stackrel{\mathrm{def}}{=} & \arg\min\limits_{y}\{\langle Qw, w\rangle :\ w = (x, s, y, \tau)\}, \\ \hat{F}(v) & = & F_P(x) + F_D(s) - \ln\tau, \quad v \in \widehat{\mathcal{C}}, \quad \nu_{\hat{F}} = \nu_F, \\ \widehat{\Phi}(v) & = & \frac{1}{2}\|v\|^2_{\hat{Q}} + \hat{F}(v), \quad v \in \mathrm{int}\,\widehat{\mathcal{C}}. \end{array}$$

$\widehat{\mathcal{C}}$ is normal $\quad \Rightarrow \quad \nabla^2 \hat{F}(\cdot) \succ 0.$

Central path: fix $v_1 \in \mathrm{int}\,\widehat{\mathcal{C}}$ with $\lambda(v_1) = 1$.

Define the *central path* $v(\mu)$ as $\mu \to 0$:

$$\nabla\widehat{\Phi}(v(\mu)) \equiv \hat{Q}v(\mu) + \nabla\hat{F}(v(\mu)) \quad = \quad \mu[\hat{Q}v_1 + \nabla\hat{F}(v_1)] \stackrel{\mathrm{def}}{=} \mu g_1.$$

Clearly, $v(1) = v_1$.

Properties of the central path

- $\|v(\mu)\|_{\hat{Q}} \leq \nu_F^{1/2}, \quad \mu \in (0,1]$.
- Denote $V^* = \{v \in \widehat{\mathcal{C}} : \hat{Q}v = 0\}$. For any $v^* \in V^*$ we have

$$\begin{aligned} \mu \cdot \langle -\nabla \hat{F}(v_1), v^* \rangle &= \langle -\nabla \hat{F}(v(\mu)), v^* \rangle \\ &\geq \langle \nabla^2 \hat{F}(v(\mu)) v^*, v^* \rangle^{1/2}. \end{aligned}$$

- If $v^* = (x^*, s^*, 1) \in V^*$, then

$$\frac{\langle s_0, x_0 \rangle}{\mu \cdot \langle -\nabla \hat{F}(v_1), v^* \rangle} \leq \langle s_0, x(\mu) \rangle + \langle s(\mu), x_0 \rangle.$$

Since $\langle s_0, x \rangle + \langle s, x_0 \rangle \leq \Omega \cdot \|v\|_{\hat{Q}} + \tau \cdot \langle s_0, x_0 \rangle$, we get

Lemma: For any $\mu \in (0,1]$

$$\frac{\tau(\mu)}{\|v(\mu)\|_{\hat{Q}}} \geq \frac{1}{\nu_F^{1/2} \langle -\nabla \hat{F}(v_1), v^* \rangle \cdot \mu} - \frac{\Omega}{\langle s_0, x_0 \rangle}.$$

Augmented self-concordant barriers (N. & Vial 2004)

Denote $\psi_\mu(v) = \widehat{\Phi}(v) - \mu\langle g_1, w\rangle$. For $v \in \operatorname{int}\widehat{\mathcal{C}}$, consider two local metrics:

$$\begin{aligned}\sigma_v^*(g) &= \langle \nabla^2 \hat{F}(v)[\nabla^2\widehat{\Phi}(v)]^{-1}g, [\nabla^2\widehat{\Phi}(v)]^{-1}g\rangle, \\ \theta_v^*(g) &= \langle g, [\nabla^2\hat{F}(v)]^{-1}g\rangle^{1/2}. \quad \text{(Simple to compute!)}\end{aligned}$$

Properties:

- $\sigma_v^*(g) \le \theta_v^*(g)$.
- For all $v \in \operatorname{int}\widehat{\mathcal{C}}$, $\sigma_v^*(\nabla\widehat{\Phi}(v)) \le \nu_F^{1/2}$.
- For the Newton iterate $v_+ = v - [\nabla^2\psi_\mu(v)]^{-1}\nabla\psi_\mu(v)$, we have
$$\theta_{v_+}^*(\nabla\psi_\mu(v_+))) \le \left(\frac{\sigma_v^*(\nabla\psi_\mu(v))}{1-\sigma_v^*(\nabla\psi_\mu(v))}\right)^2.$$

Predictor-corrector technique

Neighborhood of the Central Path:

$$\mathcal{N}_\beta(\mu) \;=\; \{v:\; \theta^*_v(\nabla\psi_\mu(v)) \le \beta\}, \quad \mu \in (0,1],\; \beta \in \left[0, \tfrac{3-\sqrt{5}}{2}\right).$$

Predictor step

$$T_v(\alpha) \;=\; \underbrace{v - [\nabla^2\widehat{\Phi}(v)]^{-1}\nabla\psi_\mu(v)}_{\approx v(\mu)} - \alpha\,\underbrace{[\nabla^2\widehat{\Phi}(v)]^{-1}g_1}_{\approx v'(\mu)}, \quad \alpha \ge 0.$$

Goal: for $v \in \mathcal{N}_{\beta_0}(\mu)$ with $\beta_0 \le \beta_1$, choose the maximal α:

$$T_v(\alpha) \in \mathcal{N}_{\beta_1}(\mu - \alpha).$$

Predictor-corrector technique

Neighborhood of the Central Path:

$$\mathcal{N}_\beta(\mu) \;=\; \{v :\; \theta_v^*(\nabla\psi_\mu(v)) \le \beta\}, \quad \mu \in (0,1],\; \beta \in \left[0, \tfrac{3-\sqrt{5}}{2}\right).$$

Predictor step

$$T_v(\alpha) \;=\; \underbrace{v - [\nabla^2\widehat{\Phi}(v)]^{-1}\nabla\psi_\mu(v)}_{\approx\, v(\mu)} - \alpha\underbrace{[\nabla^2\widehat{\Phi}(v)]^{-1}g_1}_{\approx\, v'(\mu)}, \quad \alpha \ge 0.$$

Goal: for $v \in \mathcal{N}_{\beta_0}(\mu)$ with $\beta_0 \le \beta_1$, choose the maximal α:

$$T_v(\alpha) \in \mathcal{N}_{\beta_1}(\mu - \alpha).$$

Theorem: for $v \in \mathcal{N}_{\beta_0}(\mu)$, and $\alpha \ge 0$ satisfying

$$\beta_0 + \tfrac{\alpha}{\mu}(\beta_0 + \sigma_v^*(\nabla\widehat{\Phi}(v))) \;\le\; \tfrac{\sqrt{\beta_1}}{1+\sqrt{\beta_1}},$$

we have $T_v(\alpha) \in \mathcal{N}_{\beta_1}(\mu - \alpha)$.

Possible path-following strategies

1. **Lazy scheme.** $\beta_1 = \beta_0$. Efficiency estimate :

$$\alpha_k \quad \geq \quad \frac{5\ \mu_k}{4 + 36\nu_F^{1/2}} \quad (\beta_0 = \tfrac{1}{9}).$$

Possible path-following strategies

1. **Lazy scheme.** $\beta_1 = \beta_0$. Efficiency estimate :

$$\alpha_k \geq \frac{5\,\mu_k}{4 + 36\nu_F^{1/2}} \quad (\beta_0 = \tfrac{1}{9}).$$

2. **Real predictor-corrector.**
   Note that

$$\theta^*_{T_{v_k}(\alpha_k)}(\nabla \psi_{\mu_k - \alpha_k}(T_{v_k}(\alpha_k)))$$

$$\leq \frac{\beta_0^2}{(1-\beta_0)^2} + \frac{2\beta_0}{(1-\beta_0)^3} \cdot \left[\beta_0 + \sigma^*_{v_k}(\nabla \widehat{\Phi}(v_k))\right] \cdot \frac{\alpha_k}{\mu_k} + O\left(\frac{\alpha_k^2}{\mu_k^2}\right).$$

   Thus, it is better to keep β_0 small with respect to β_1.

Computational aspects

1. Path-following scheme.
The measure

$$\theta_v^*(\nabla\psi_\mu(v)) \;=\; \langle \nabla\psi_\mu(v), [\nabla^2 \hat{F}(v)]^{-1} \nabla\psi_\mu(v)\rangle^{1/2}$$

is easy to compute.
Examples. a) LP. For positive orthant

$$\theta_v^*(g) \;=\; \left[\sum_{i=1}^n \left(g^{(i)} v^{(i)} \right)^2 \right]^{1/2}.$$

b) Separable cones. Direct product of many cones of small dimension.

Computational aspects

2. Potential-reduction scheme. For homogeneous quadratic potential

$$\tilde{\Phi}(w) \quad = \quad \nu_F \ln \|w\|_Q + F(w) + \tfrac{\nu_F}{2}[1 - \ln \nu_F],$$

we have an upper bound

$$\tilde{\Phi}(w+h) - \tilde{\Phi}(w) \quad \leq \quad \nu_F \tfrac{2\langle Qw,h\rangle + \langle Qh,h\rangle}{2\langle Qw,w\rangle} + F(w+h) - F(w).$$

Hence, for the Newton step we need to minimize

$$\nu_F \tfrac{2\langle Qw,h\rangle + \langle Qh,h\rangle}{2\langle Qw,w\rangle} + \langle \nabla F(w), h\rangle + \langle \nabla^2 F(w) h, h\rangle.$$

Note: We need to achieve a constant drop (by CG?).

Computational aspects

Main competitors: Fast gradient schemes for solving the problem

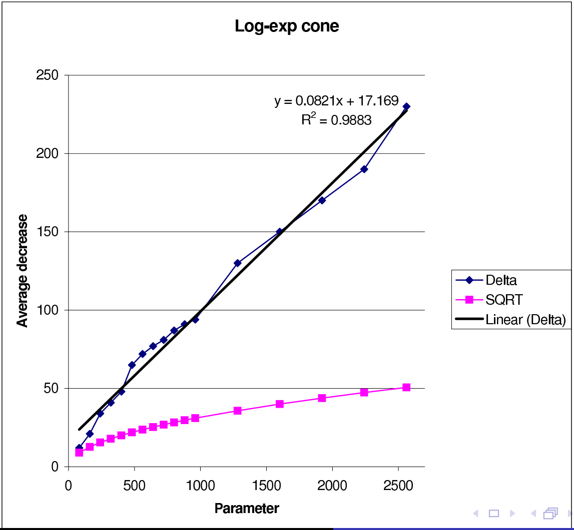
$$\min_{w} \{ \langle Qw, w \rangle : \ \tau = 1, \ w \in \mathcal{F} \}.$$

Note:

- The memory requirements are the same.
- One iteration of CG for potential-reduction methods has the same complexity as one iteration of GM.

n	m	LP	Lorentz	$K_{\frac{1}{2}}$	$K_{\frac{1}{3}}$	K_{exp}
60	20	62	55	74	76	71
120	40	70	62	79	79	68
180	60	76	65	73	73	66
240	80	85	73	84	83	73
300	100	86	78	93	92	76
360	120	86	74	98	98	80
480	160	84	80	95	96	95
600	200	98	84	101	100	85
720	240	102	89	109	110	93
840	280	99	96	111	109	95
960	320	98	90	107	107	102

Table: Performance of long-step PF-method


Log-exp cone
y = 0.0821x + 17.169
R2 = 0.9883
Average decrease
Parameter
250
200
150
100
50
0
0
500
1000
1500
2000
2500
Delta
SQRT
Linear (Delta)

Potential-reduction scheme for LP

n	m	Iterations	Average $\Delta_{\tilde{\Phi}}$
60	20	12	$7.2 \cdot 10^1$
120	40	13	$1.4 \cdot 10^2$
180	60	14	$2.1 \cdot 10^2$
240	80	15	$2.7 \cdot 10^2$
300	100	16	$3.1 \cdot 10^2$
360	120	16	$3.8 \cdot 10^2$
480	160	16	$5.0 \cdot 10^2$
600	200	17	$6.2 \cdot 10^2$
720	240	17	$7.6 \cdot 10^2$
840	280	17	$8.8 \cdot 10^2$
960	320	18	$9.6 \cdot 10^2$