Second Order Optimization Algorithms III: Primal-Dual Potential-Reduction and Infeasibility/Unboundedness Detection

Yinyu Ye

MS&E and ICME, Stanford University

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

Chapter 5.4-7, 6.6, Chapter 15

$$\begin{array}{l|l|l} \min\ c^Tx & \max\ b^Ty & x^Ts=0 \\ \text{s.t.}\ \ Ax=b & \text{s.t.}\ \ c-A^Ty-s=0 & \\ \qquad x\ge 0 & \qquad\qquad s\ge 0 & \end{array}$$

$\frac{1}{n}\sum_{j=1}^{n} x_j s_j \geq \left(\prod_{j=1}^{n}(x_j s_j)\right)^{1/n}$

Primal-Dual Potential Function for LP

Karmarkar:

For $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}$, the joint primal-dual potential function is defined by

$(\log(x_j) + \log(s_j))$

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := \boxed{(n+\rho)}\log(\mathbf{x}^T\mathbf{s}) - \sum_{j=1}^{n}\log(x_j s_j), \quad \text{for some } \rho > 0.$$

$-\infty$

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) = \boxed{\rho\log(\mathbf{x}^T\mathbf{s})} + \boxed{\psi_n(\mathbf{x}, \mathbf{s})} \geq \rho\log(\mathbf{x}^T\mathbf{s}) + n\log n,$$

then, for $\rho > 0$, $\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \to -\infty$ implies that $\mathbf{x}^T\mathbf{s} \to 0$. More precisely, we have

$$\mathbf{x}^T\mathbf{s} \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) - n\log n}{\rho}\right).$$

0.25 1/4

Given a pair $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \operatorname{int} \mathcal{F}$, compute direction vectors $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from the Newton iteration:

$$\left\{\begin{array}{rcl} S^k\mathbf{d}_x + X^k\mathbf{d}_s &=& \frac{(\mathbf{x}^k)^T\mathbf{s}^k}{n+\rho}\mathbf{e} - X^k\mathbf{s}^k, \\ A\mathbf{d}_x &=& \mathbf{0}, \\ A^T\mathbf{d}_y + \mathbf{d}_s &=& \mathbf{0}. \end{array}\right. \tag{1}$$

How to solve the equation system efficiently using the block structures?

Block Structure in the KKT System

$$
\begin{aligned}
S^k \mathbf{d}_x + X^k \mathbf{d}_s &= \mathbf{r}^k, \\
A\mathbf{d}_x &= \mathbf{0}, \\
A^T \mathbf{d}_y + \mathbf{d}_s &= \mathbf{0}.
\end{aligned}
$$

Scale the first block to: $\mathbf{d}_x + (S^k)^{-1} X^k \mathbf{d}_s = (S^k)^{-1}\mathbf{r}^k$.

Multiplying A to both sides and using the second block equations: $A(S^k)^{-1} X^k \mathbf{d}_s = A(S^k)^{-1}\mathbf{r}^k$.

Applying the third block equations: $-A(S^k)^{-1} X^k A^T \mathbf{d}_y = A(S^k)^{-1}\mathbf{r}^k$.

This is an $m \times m$ positive definite system, and solve it for $\mathbf{d}_y$; then $\mathbf{d}_s$ from the third block; then $\mathbf{d}_x$ from the first block.

Positive Definite System Equation Solver: $Q\mathbf{d} = \mathbf{r}$ where Q is a PD matrix.

Matrix Factorization:

- Cholesky: $R^T R = Q$, where R is a Right-Triangle matrix
- $LDL^T = Q$, where L is a Left-Triangle matrix.

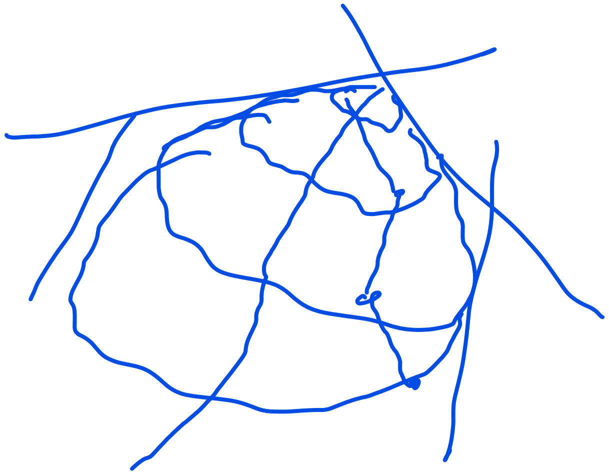
Description of Algorithm for LP

Given $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0) \in \text{int}\,\mathcal{F}$. Set $\rho \geq \sqrt{n}$ and $k := 0$.

While $(\mathbf{x}^k)^T \mathbf{s}^k \geq \epsilon$ **do**

1. Set $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$ and compute $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from (4).
2. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}_x$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k \mathbf{d}_y$, and $\mathbf{s}^{k+1} = \mathbf{s}^k + \alpha^k \mathbf{d}_s$ where

$$\alpha^k = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(\mathbf{x}^k + \alpha \mathbf{d}_x, \mathbf{s}^k + \alpha \mathbf{d}_s).$$



3. Let $k := k + 1$ and return to Step 1.

Theorem 1 *Let $\rho \geq \sqrt{n}$. Then, the potential reduction algorithm generates the (interior) feasible solution sequence $\{\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k\}$ such that*

$$\psi_{n+\rho}(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) - \psi_{n+\rho}(\mathbf{x}^k, \mathbf{s}^k) \leq -0.15.$$

Thus, if $\psi_{n+\rho}(\mathbf{x}^0, \mathbf{s}^0) \leq \rho \log((\mathbf{x}^0)^T\mathbf{s}^0) + n \log n$, the algorithm terminates in at most $O(\rho \log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon))$ iterations with $(\mathbf{x}^k)^T\mathbf{s}^k = \mathbf{c}^T\mathbf{x}^k - \mathbf{b}^T\mathbf{y}^k \leq \epsilon$.

The proof used a key fact: $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T \mathbf{d}_y = 0$ for the directions. Also

$$\begin{aligned}
(\mathbf{x}^k)^T\mathbf{s}^k &\leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^k,\mathbf{s}^k) - n\log n}{\rho}\right) \\
&\leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^0,\mathbf{s}^0) - n\log n - \rho\log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon)}{\rho}\right) \\
&\leq \exp\left(\frac{\rho\log(\mathbf{x}^0,\mathbf{s}^0) - \rho\log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon)}{\rho}\right) \\
&= \exp(\log(\epsilon)) = \epsilon.
\end{aligned}$$

The role of ρ? And more aggressive step size?

(LPpdpath...m and LPpdpotential.m of Chapter 5)

Proof Sketch of the Reduction Theorem

We first have the following lemma:

Lemma 1 *Let the direction vector* $\mathbf{d} = (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ *be computed by (4), and let* $\theta = \frac{\alpha\sqrt{\min(XS\mathbf{e})}}{\|(XS)^{-1/2}\mathbf{r}\|}$ *where* α *is a positive constant less than* 1. *Let*

$$\mathbf{x}^+ = \mathbf{x} + \theta\mathbf{d}_x, \quad \mathbf{y}^+ = \mathbf{y} + \theta\mathbf{d}_y, \quad \textit{and} \quad \mathbf{s}^+ = \mathbf{s} + \theta\mathbf{d}_s.$$

Then, we have $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \mathrm{int}\,\mathcal{F}$ *and*

$$\begin{array}{l} \psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \\ \le -\alpha\sqrt{\min(XS\mathbf{e})}\|(XS)^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{x}^T\mathbf{s}}X\mathbf{s})\| + \frac{\alpha^2}{2(1-\alpha)} \end{array}.$$

$$
\begin{aligned}
& \psi(\mathbf{x}^+, \mathbf{s}^+) - \psi(\mathbf{x}, \mathbf{s}) \\
= \quad & (n+\rho)\log\left(1 + \frac{\theta \mathbf{d}_s^T \mathbf{x} + \theta \mathbf{d}_x^T \mathbf{s}}{\mathbf{x}^T \mathbf{s}}\right) - \textstyle\sum_{j=1}^n \left(\log(1 + \frac{\theta d_{s_j}}{s_j}) + \log(1 + \frac{\theta d_{x_j}}{x_j})\right) \\
\leq \quad & (n+\rho)\left(\frac{\theta \mathbf{d}_s^T \mathbf{x} + \theta \mathbf{d}_x^T \mathbf{s}}{\mathbf{x}^T \mathbf{s}}\right) - \textstyle\sum_{j=1}^n \left(\log(1 + \frac{\theta d_{s_j}}{s_j}) + \log(1 + \frac{\theta d_{x_j}}{x_j})\right) \\
\leq \quad & (n+\rho)\left(\frac{\theta \mathbf{d}_s^T \mathbf{x} + \theta \mathbf{d}_x^T \mathbf{s}}{\mathbf{x}^T \mathbf{s}}\right) - \theta \mathbf{e}^T (S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \frac{\|\theta S^{-1}\mathbf{d}_s\|^2 + \|\theta X^{-1}\mathbf{d}_x\|^2}{2(1-\alpha)} \\
\leq \quad & \tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} \theta(\mathbf{d}_s^T \mathbf{x} + \mathbf{d}_x^T \mathbf{s}) - \theta \mathbf{e}^T (S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & \theta\left(\tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} \mathbf{e}^T (X\mathbf{d}_s + S\mathbf{d}_x) - \mathbf{e}^T (S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x)\right) + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & \theta\left(\tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} \mathbf{e}^T (X\mathbf{d}_s + S\mathbf{d}_x) - \mathbf{e}^T (XS)^{-1}(X\mathbf{d}_s + S\mathbf{d}_x)\right) + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & \theta\left(\tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} XS\mathbf{e} - \mathbf{e}\right)^T (XS)^{-1} (X\mathbf{d}_s + S\mathbf{d}_x) + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & \theta\left(\tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} XS\mathbf{e} - \mathbf{e}\right)^T (XS)^{-1} \left(\tfrac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - XS\mathbf{e}\right) + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & -\theta \cdot \tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}} \cdot \|(XS)^{-1/2}\mathbf{r}\|^2 + \tfrac{\alpha^2}{2(1-\alpha)} \\
= \quad & -\alpha \underbrace{\sqrt{\min(XS\mathbf{e})} \cdot \|\tfrac{n+\rho}{\mathbf{x}^T\mathbf{s}}(XS)^{-1/2}\mathbf{r}\|}_{\geq \frac{\sqrt{3}}{2}} + \tfrac{\alpha^2}{2(1-\alpha)}.
\end{aligned}
$$

$\alpha = \frac{1}{\cancel{4}\,2}$

< 0

$\leq \frac{-\alpha\sqrt{3}}{2} + \frac{\alpha^2}{2(1-\alpha)}$

$-\frac{\sqrt{3}}{4} + \frac{1}{4} \leq \frac{-\sqrt{3}+1}{4}$

$\leq \frac{-.7}{4}$

Let $\mathbf{v} = XS\mathbf{e}$. Then, we can prove the following technical lemma:

Lemma 2 *Let $\mathbf{v} \in \mathcal{R}^n$ be a positive vector and $\rho \geq \sqrt{n}$. Then,*

$$\sqrt{\min(\mathbf{v})}\|V^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{e}^T\mathbf{v}}\mathbf{v})\| \geq \sqrt{3/4}\,.$$

Combining these Lemmas 2 and 3 we have

$$\psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s})$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)} = -\delta$$

for a constant δ.

Path-follow
Potenti.-reduct

x^0
$\begin{pmatrix} y^0 \\ s^0 \end{pmatrix}$

Initialization

$\min \; c^T x + M \|v\|_1$
$\text{st} \; Ax \cancel{= b} + v = b$
$x \geq 0$

- Combining the primal and dual into a single linear feasibility problem, then applying LP algorithms to find a feasible point of the problem. Theoretically, this approach can retain the currently best complexity result.

- The big M method, i.e., add one or more artificial column(s) and/or row(s) and a huge penalty parameter M to force solutions to become feasible during the algorithm.

- Phase I-then-Phase II method, i.e., first try to find a feasible point (and possibly one for the dual problem), and then start to look for an optimal solution if the problem is feasible and bounded.

- Combined Phase I-Phase II method, i.e., approach feasibility and optimality simultaneously. To our knowledge, the "best" complexity of this approach is $O(n \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon))$.

Homogeneous and Self-Dual Algorithm

- It solves the linear programming problem without any regularity assumption concerning the existence of optimal, feasible, or interior feasible solutions, while it retains the currently best complexity result
- It can start at any positive primal-dual pair, feasible or infeasible, near the central ray of the positive orthant (cone), and it does not use any big M penalty parameter or lower bound.
- Each iteration solves a system of linear equations whose dimension is almost the same as that solved in the standard (primal-dual) interior-point algorithms.
- If the LP problem has a solution, the algorithm generates a sequence that approaches feasibility and optimality simultaneously; if the problem is infeasible or unbounded, the algorithm will produce an infeasibility certificate for at least one of the primal and dual problems.

Primal-Dual Alternative Systems

Recall that a pair of LP has two alternatives

$$
\begin{array}{lrl}
\text{(Solvable)} & A\mathbf{x} - \mathbf{b} & = \mathbf{0} \\
& -A^T\mathbf{y} + \mathbf{c} & \geq \mathbf{0}, \\
& \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & = 0, \\
& \mathbf{y} \text{ free}, \ \mathbf{x} & \geq \mathbf{0}
\end{array}
\quad \text{or} \quad
\begin{array}{lrl}
\text{(Infeasible)} & A\mathbf{x} & = \mathbf{0} \\
& -A^T\mathbf{y} & \geq \mathbf{0}, \\
& \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & > 0, \\
& \mathbf{y} \text{ free}, \ \mathbf{x} & \geq \mathbf{0}
\end{array}
$$

$$
\begin{array}{lrl}
(HP) & A\mathbf{x} - \mathbf{b}\tau & = \mathbf{0} \\
& -A^T\mathbf{y} + \mathbf{c}\tau & = \mathbf{s} \geq \mathbf{0}, \\
& \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & = \kappa \geq 0, \\
& \mathbf{y} \text{ free}, \ (\mathbf{x}; \tau) & \geq \mathbf{0}
\end{array}
$$

$x + s > 0$

$\kappa + \tau > 0$

where the two alternatives are:

$$\text{(Solvable)} : \ (\tau > 0, \kappa = 0) \quad \text{or} \quad \text{(Infeasible)} : \ (\tau = 0, \kappa > 0)$$

Let's Find a Feasible Solution of (HP)

Given $\mathbf{x}^0 = \mathbf{e} > \mathbf{0}$, $\mathbf{s}^0 = \mathbf{e} > \mathbf{0}$, and $\mathbf{y}^0 = \mathbf{0}$, we formulate a self-dual LP problem:

$$
\begin{array}{llcccccl}
(HS-DP) & \min & & & & (n+1)\theta & & \\
& \text{s.t.} & & A\mathbf{x} & -\mathbf{b}\tau & +\bar{\mathbf{b}}\theta & = \mathbf{0}, \\
& & -A^T\mathbf{y} & & +\mathbf{c}\tau & -\bar{\mathbf{c}}\theta & \geq \mathbf{0}, \\
& & \mathbf{b}^T\mathbf{y} & -\mathbf{c}^T\mathbf{x} & & +\bar{z}\theta & \geq 0, \\
& & -\bar{\mathbf{b}}^T\mathbf{y} & +\bar{\mathbf{c}}^T\mathbf{x} & -\bar{z}\tau & & = -(n+1), \\
& & \mathbf{y} \text{ free}, & \mathbf{x} \geq \mathbf{0}, & \tau \geq 0, & \theta \text{ free}. &
\end{array}
$$

Note that $(\mathbf{y} = \mathbf{0},\ \mathbf{x} = \mathbf{e},\ \tau = 1,\ \theta = 1)$ is a strictly feasible point for (HSDP). Moreover, one can show that the constraints imply

$$\mathbf{e}^T x + \mathbf{e}^T s + \tau + \kappa - (n+1)\theta = (n+1),$$

which serves as a normalizing constraint for (HSDP) to prevent the all-zero solution.

Main Result

Theorem 2 *The interior-point algorithm solves (HS-DP) in* $O(\sqrt{n}\log\frac{n}{\epsilon})$ *steps and each step solves a system of linear equations as the same size as in feasible algorithms, and it always produces an optimal solution* $(\mathbf{y}^*, \mathbf{x}^*, \tau^*, \mathbf{s}^*, \kappa^*, \theta^* = 0)$ *where* $\tau^* + \kappa^* > 0$. *If* $\tau^* > 0$ *then it produces an optimal solution pair for the original LP problem; if* $\kappa^* > 0$, *then it produces a certificate to prove (at least) one of the pair is infeasible.*

$ee^T \cdot X = 0 \Leftrightarrow e^T X e = 0$

$\Updownarrow$

$(\sum x_j)^2 = 0$

n^2 $\quad A_i = \boxed{a_i a_i^T}$ $\quad a_i \in R^n$

min $C \cdot X$

s.t. $X_{ii} = 1$

$X \succeq 0$

$x_{11} = -1$

$e_i e_i^T \cdot X = 1$

$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$\begin{array}{ll} \min & C \cdot X \\ \text{s.t.} & A_1 \cdot X = b_1 \\ & \cdots \\ X \succeq 0 & A_m \cdot X = b_m \end{array} \qquad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & \sum_{i=1}^{m} A_i y_i - C + S = 0 \end{array}$$

Extensions to Solving SDP: Potential Function

For any $X \in \operatorname{int} \mathcal{F}_p$ and $(\mathbf{y}, S) \in \operatorname{int} \mathcal{F}_d$, let parameter $\rho > 0$ and

$$\psi_{n+\rho}(X, S) := (n + \rho) \log(X \bullet S) - \log(\det(X) \cdot \det(S)),$$

$$\psi_{n+\rho}(X, S) = \rho \log(X \bullet S) + \psi_n(X, S) \geq \rho \log(X \bullet S) + n \log n.$$

Then, $\psi_{n+\rho}(X, S) \to -\infty$ implies that $X \bullet S \to 0$. More precisely, we have

$$X \bullet S \leq \exp\left(\frac{\psi_{n+\rho}(X, S) - n \log n}{\rho}\right).$$

Primal-Dual SDP Alternative Systems

A pair of SDP has two alternatives under mild conditions

$$\text{(Solvable)}\quad \begin{array}{rl} \mathcal{A}X - \mathbf{b} & = \mathbf{0} \\ -\mathcal{A}^T\mathbf{y} + C & \succeq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - C \bullet X & = 0, \\ \mathbf{y} \text{ free},\ X & \succeq \mathbf{0} \end{array} \quad \text{or} \quad \text{(Infeasible)}\quad \begin{array}{rl} \mathcal{A}X & = \mathbf{0} \\ -\mathcal{A}^T\mathbf{y} & \succeq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - C \bullet X & > 0, \\ \mathbf{y} \text{ free},\ X & \succeq \mathbf{0} \end{array}$$

An Integrated Homogeneous System

The two alternative systems can be homogenized as one:

$$
\begin{array}{lrl}
(HSDP) & \mathcal{A}X - \mathbf{b}\tau & = \mathbf{0} \\
& -\mathcal{A}^T\mathbf{y} + C\tau & = \mathbf{s} \geq \mathbf{0}, \\
& \mathbf{b}^T\mathbf{y} - C \bullet X & = \kappa \geq 0, \\
& \mathbf{y} \text{ free},\; X \succeq \mathbf{0}, & \tau \geq 0,
\end{array}
$$

where the three alternatives are

$$
\begin{array}{rl}
\text{(Solvable)}: & (\tau > 0, \kappa = 0) \quad : \\
\text{(Infeasible)}: & (\tau = 0, \kappa > 0) \\
\text{(All others)}: & (\tau = \kappa = 0).
\end{array}
$$

Primal-Dual Interior-Point Algorithms for More General Convex Optimization I

We now present an algorithm for solving more general convex optimization problems:

$$\min \; f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \geq \mathbf{0}$$

where, with the notation $X = \text{diag}(\mathbf{x})$, we look for a constrained root of

$$X\mathbf{g}(\mathbf{x}) = \mathbf{0}, \; \mathbf{x} \geq \mathbf{0}, \; \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \quad \text{where} \quad \mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}).$$

Vector function $\mathbf{g}(\mathbf{x})$ would be a monotone mapping and the solution is also called the monotone complemtarity point.

We assume that f meets a Scaled Lipschitz condition: for any point $\mathbf{x} > \mathbf{0}$

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

Examples of Scaled Lipschitz Functions

- $f(x) = -\log(x)$, $g(x) = \frac{-1}{x}$ and $g'(x) = \frac{1}{x^2}$: Not Lipschitz but Scaled Lipschitz

$$\frac{-1}{x+d} - \frac{-1}{x} - \frac{d}{x^2} = \frac{1}{x}\left(\sum_{p=2}^{\infty}(\frac{-d}{x})^p\right) \leq \frac{d^2}{x^3}\frac{1}{1-\alpha} \Rightarrow \beta_\alpha = \frac{1}{1-\alpha}.$$

- $f(x) = x\log(x)$, $g(x) = 1 + \log(x)$ and $g'(x) = \frac{1}{x}$: Not Lipschitz but Scaled Lipschitz.
- $f(x) = e^x$, $g(x) = e^x$ and e^x: Both Lipschitz and Scaled Lipschitz at Bounded x.

Interior-Point Algorithms for More General Convex Optimization II

We start from a solution $\mathbf{x}^k > \mathbf{0}$ and $\mathbf{g}(\mathbf{x}^k) > \mathbf{0}$ and they approximately satisfy the equations

$$X\mathbf{s} = \mu^k \mathbf{e}, \quad \mathbf{s} = \mathbf{g}(\mathbf{x}), \quad \text{for some} \quad \mu^k > 0. \tag{3}$$

Such a solution exists because it is the (unique) optimal solution for the problem with logarithmic barrier

$$\min \ f(\mathbf{x}) - \mu^k \sum_j \log(x_j).$$

We replace μ^k by $\mu^{k+1} = (1 - \frac{\eta}{\sqrt{n}})\mu^k$ and aim to find a solution $\mathbf{x} > \mathbf{0}$ such that $\mathbf{g}(\mathbf{x}) > \mathbf{0}$

$$X\mathbf{s} = \mu^{k+1}\mathbf{e}, \quad \mathbf{s} = \mathbf{g}(\mathbf{x}).$$

Starting from $(\mathbf{x}^k, \mathbf{s}^k)$, we apply the Newton iteration using the auxiliary variables $\mathbf{s} = \mathbf{g}(\mathbf{x})$ ($\mathbf{s}^k = \mathbf{g}(\mathbf{x}^k)$):

$$\begin{aligned} X^k\mathbf{d}_s + S^k\mathbf{d}_x &= (1 - \tfrac{\eta}{\sqrt{n}})\mu^k\mathbf{e} - X^k\mathbf{s}^k, \\ \mathbf{d}_s &= \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x. \end{aligned} \tag{4}$$

Now we analyze the quality of the new iterate $\mathbf{x}^+ = \mathbf{x}^k + \mathbf{d}_x$ together with $\mathbf{s}^+ = \mathbf{g}(\mathbf{x}^+)$.

Multiplying $(X^k S^k)^{-1/2}$ to both sides of the first equation of (4), we have

$$D^{-1}\mathbf{d}_x + D\mathbf{d}_s = (X^k S^k)^{-1/2}((1 - \frac{\eta}{\sqrt{n}})\mu^k \mathbf{e} - X^k \mathbf{s}^k),$$

where $D = (X^k)^{1/2}(S^k)^{-1/2}$. Note that $\mathbf{d}_x^T \mathbf{d}_s = \mathbf{d}_x^T \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}_x \geq 0$

$$\begin{aligned} \|D^{-1}\mathbf{d}_x\|^2 + \|D\mathbf{d}_s\|^2 &\leq \|D^{-1}\mathbf{d}_x + D\mathbf{d}_s\|2 \\ &= \|(X^k S^k)^{-1/2}((1 - \tfrac{\eta}{\sqrt{n}})\mu^k \mathbf{e} - X^k \mathbf{s}^k)\|^2 \\ &= \eta^2 \mu^k. \end{aligned}$$

From now on, we set $0 < \eta \leq 1/2$. Thus,

$$\|(X^k)^{-1}\mathbf{d}_x\| = \|(X^k S^k)^{-1/2} D^{-1}\mathbf{d}_x\| \leq \|(X^k S^k)^{-1/2}\| \|D^{-1}\mathbf{d}_x\| \leq \eta \leq 1/2.$$

Consequently, $1/2 \leq x_j^+/x_j^k \leq 3/2$ for all j. Moreover,

$$\|D_x \mathbf{d}_s\| = \|D^{-1} D_x D \mathbf{d}_s\| = \|D^{-1}\mathbf{d}_x\| \|D\mathbf{d}_s\| \leq (\|D^{-1}\mathbf{d}_x\|^2 + \|Dd_s\|^2)/2 \leq \eta^2 \mu^k/2,$$

and

$$\mathbf{d}_x^T \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}_x = \mathbf{d}_x^T \mathbf{d}_s = \mathbf{d}_x^T D^{-1} D \mathbf{d}_s \le \|D^{-1}\mathbf{d}_x\| \|D\mathbf{d}_s\| \le \eta^2 \mu^k / 2.$$

Consider

$$\begin{aligned} & X^+ s^+ - (1-\gamma)\mu^k \mathbf{e} \\ = \quad & X^+(\mathbf{s}^k + \mathbf{d}_s + \mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}_x) - (1-\gamma)\mu^k \mathbf{e} \\ = \quad & (X^k + D_x)(\mathbf{s}^k + \mathbf{d}_s) - (1-\gamma)\mu^k \mathbf{e} + X^+(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}_x) \\ = \quad & D_x \mathbf{d}_s + X^+(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla \mathbf{g}(\mathbf{x}^k)\mathbf{d}_x). \end{aligned}$$

From $\mathbf{g}$ being scaled Lipschitz with $\beta_{1/2}$ and the above inequalities we obtain

$$\begin{aligned}
\|X^+\mathbf{s}^+ - (1-\tfrac{\eta}{\sqrt{n}})\mu^k\mathbf{e}\| &= \|D_x\mathbf{d}_s + X^+(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x)| \\
&= \|D_x\mathbf{d}_s + (X^k)^{-1}X^+X^k(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x)\| \\
&\le \|D_x\mathbf{d}_s\| + \|(X^k)^{-1}X^+\|\|X^k(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x)\| \\
&\le \|D_x\mathbf{d}_s\| + (3/2)\|X^k(\mathbf{g}(\mathbf{x}^+) - \mathbf{g}(\mathbf{x}^k) - \nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x)\| \\
&\le \|D_x\mathbf{d}_s\| + (3/2)\beta_{1/2}\mathbf{d}_x^T\nabla\mathbf{g}(\mathbf{x}^k)\mathbf{d}_x \\
&\le (1/2 + (3/4)\beta_{1/2})\eta^2\mu^k.
\end{aligned}$$

comparing with the initial error is $\|X^k\mathbf{s}^k - (1-\frac{\eta}{\sqrt{n}})\mu^k\mathbf{e}\| = \eta\mu^k$.

There is also a Homogeneous and Self-Dual Algorithm for solving the monotone complementarity problem, which is a basic solver of MOSEK. The algorithm produces a certificate if no complementarity solution exists.

(HOmcp.m, mcpfun.m and mcpJacobian.m of Chapter 15)

$\to f(x)$

$\boxed{g(x)} = \nabla f(x)$

$H(x) = \nabla^2 f(x)$

Software Implementation

Cplex-Barrier IBM, GUROBI, COPT ← all-round

SEDUMI: `http://sedumi.mcmaster.ca/` SDP

MOSEK: `http://www.mosek.com/products_mosek.html` all-round

SDDPT3: `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html` SDP

DSDP (Dual Semidefinite Programming Algorithm): HDSDP

`http://www.stanford.edu/~yyye/Col.html`

CVX/ECOS: `http://www.stanford.edu/~boyd/cvx`

hsdLPsolver and more: `http://www.stanford.edu/~yyye/matlab.html`