On the second-order cone complementarity problem

Jein-Shan Chen

jschen@math.ntnu.edu.tw

Department of Mathematics

National Taiwan Normal University

Outline

- Introduction
- Nonsmooth functions approach
- Merit functions approach
- Algorithms
- Numerical experiments

Introduction

I

Introduction

History

1. **NCP** (1960): To find $x \in \mathbb{R}^n$ satisfying

$$\langle F(x), x \rangle = 0\ ,\ \ F(x) \geq 0\ ,\ \ x \geq 0.$$

2. **SDCP** (1980): To find $X \in \mathbb{R}^{n \times n}$ satisfying

$$\langle F(X), X \rangle = 0\ ,\ \ F(X) \succeq O\ ,\ \ X \succeq O.$$

3. **SOCCP** (1998): To find $x \in \mathbb{R}^n$ satisfying

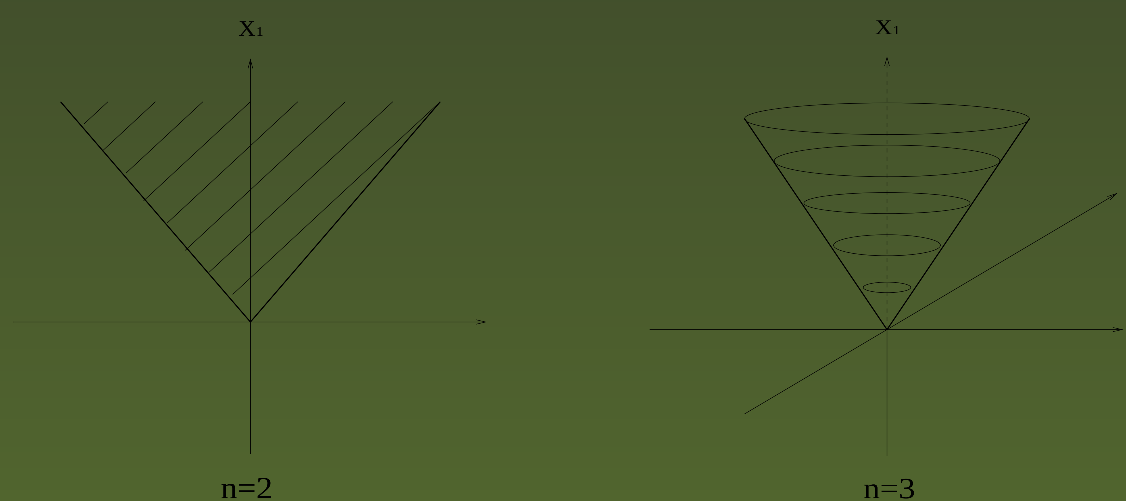
$$\langle F(x), x \rangle = 0\ ,\ \ F(x) \succeq_{\mathcal{K}^n} 0\ ,\ \ x \succeq_{\mathcal{K}^n} 0.$$

What's SOC ?

The second-order cone (SOC) in $\mathbb{R}^n$, also called Lorentz cone, is defined to be

$$\mathcal{K}^n = \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \le x_1\}$$

where $\|\cdot\|$ denotes the Euclidean norm. If $n = 1$, let $\mathcal{K}^n$ be the set of nonnegative reals $\mathbb{R}_+$.



(a) the graph of $\mathcal{K}^2$ (b) the graph of $\mathcal{K}^3$

What's SOCCP ?

The second-order cone complementarity problem (SOCCP) is to find $x, y \in \mathbb{R}^n$ and $\zeta \in \mathbb{R}^l$ such that

$$\begin{cases} \langle x, y \rangle = 0 \\ x \in \mathcal{K} \quad , \quad y \in \mathcal{K} \\ x = F(\zeta) \quad , \quad y = G(\zeta) \end{cases}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product, $F, G : \mathbb{R}^l \to \mathbb{R}^n$ are smooth mappings, $\mathcal{K}$ is the Cartesian product of SOC, that is,

$$\mathcal{K} = \mathcal{K}^{n_1} \times \cdots \times \mathcal{K}^{n_m},$$

where $n_1 + n_2 + \cdots + n_m = n$.

Why study SOCCP ?

- This problem has wide applications, e.g., Robust linear programming, filter design, antenna design, etc.. (Lobo, Vandenberghe, Boyd, Lebret, 1998)
- It includes a large class of quadratically constrained problems and minimization of sum of Euclidean norms as special cases.
- It also includes as a special case the well-known nonlinear complementarity problem (NCP).

Difficulty: $\mathcal{K}$ is closed and convex, but non-polyhedral.

Nonsmooth function

II

Nonsmooth functions approach

The matrix-valued functions

Let $\mathcal{S}^n$ be the space of $n \times n$ real symmetric matrices. For any $X \in \mathcal{S}^n$, its eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ are real and admits a spectral decomposition:

$$X = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} P^T,$$

where P is orthogonal (i.e., $P^T = P^{-1}$). Then, for any function $f : \mathbb{R} \to \mathbb{R}$, we define a corresponding matrix-valued function $f^{\mathrm{mat}} : \mathcal{S}^n \to \mathcal{S}^n$ by

$$f^{\mathrm{mat}}(X) = P \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} P^T.$$

Known results about f^{mat}

(a) f^{mat} is continuous iff f is continuous.

(b) f^{mat} is directionally differentiable iff f is directionally differentiable.

(c) f^{mat} is Fréchet-differentiable iff f is Fréchet-differentiable.

(d) f^{mat} is continuously differentiable iff f is continuously differentiable.

(e) f^{mat} is strictly continuous iff f is strictly continuous.

(f) f^{mat} is Lipschitz continuous with constant κ iff f is Lipschitz continuous with constant κ.

(g) f^{mat} is semismooth iff f is semismooth.

Remarks

1. Strictly continuous is also called locally Lipschitz continuous.
2. Let $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is strictly continuous, then F is semismooth at x if F is directionally differentiable at x and $\forall V \in \partial F(x+h)$, we have

$$F(x+h) - F(x) - Vh = \circ(\|h\|)$$

3. Convex functions and piecewise continuously differentiable functions are examples of semismooth functions.

Spectral Factorization

Let $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, then x can be decomposed as

$$x = \lambda_1 u^{(1)} + \lambda_2 u^{(2)} \quad ,$$

where λ_1, λ_2 and $u^{(1)}, u^{(2)}$ are the spectral values and the associated spectral vectors of x are given by

$$\begin{aligned} \lambda_i &= x_1 + (-1)^i \|x_2\| \quad , \\ u^{(i)} &= \begin{cases} \frac{1}{2}\left(1 \ , \ (-1)^i \dfrac{x_2}{\|x_2\|}\right), & \text{if } x_2 \neq 0 \\ \frac{1}{2}\left(1 \ , \ (-1)^i w\right), & \text{if } x_2 = 0 \quad , \end{cases} \end{aligned}$$

for $i = 1, 2$ with w being any vector in $\mathbb{R}^{n-1}$ satisfying $\|w\| = 1$. If $x_2 \neq 0$, the decomposition is unique.

The SOC-functions

For any function $f : \mathbb{R} \rightarrow \mathbb{R}$, we define a corresponding function on $\mathbb{R}^n$ associated with SOC by

$$f^{\mathrm{soc}}(x) = f(\lambda_1)\, u^{(1)} + f(\lambda_2)\, u^{(2)},$$

for all $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, where λ_1, λ_2 and $u^{(1)}, u^{(2)}$ are the spectral values and vectors of x.

If f is defined only on a subset of $\mathbb{R}$, then f^{soc} is defined on the corresponding subset of $\mathbb{R}^n$.

Parallel results about f^{soc}

(a) f^{soc} is continuous iff f is continuous.

(b) f^{soc} is directionally differentiable iff f is directionally differentiable.

(c) f^{soc} is Fréchet-differentiable iff f is Fréchet-differentiable.

(d) f^{soc} is continuously differentiable iff f is continuously differentiable.

(e) f^{soc} is strictly continuous iff f is strictly continuous.

(f) f^{soc} is Lipschitz continuous with constant κ iff f is Lipschitz continuous with constant κ.

(g) f^{soc} is semismooth iff f is semismooth.

A bridge from $f^{\rm mat}$ to $f^{\rm soc}$

For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, let λ_1, λ_2 be its spectral values, then

(a) For any $t \in \mathbb{R}$, the matrix $L_x + tM_{x_2}$ has eigenvalues of λ_1, λ_2 and $x_1 + t$ of multiplicity $n - 2$, where

$$L_x = \begin{bmatrix} x_1 & x_2^T \\ x_2 & x_1 I \end{bmatrix} \text{ and } M_{x_2} = \begin{bmatrix} 0 & 0 \\ 0 & I - \dfrac{x_2 x_2^T}{\|x_2\|^2} \end{bmatrix}.$$

(b) For any $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $t \in \mathbb{R}$, we have

$$f^{\rm soc}(x) = f^{\rm mat}\Big(L_x + tM_{x_2}\Big)e,$$

where $e = (1, 0, \cdots, 0)^T \in \mathbb{R}^n$.

A reformulation of SOCCP

Proposition [Fukushima-Luo-Tseng, 2002]

$$\langle x, y \rangle = 0, \quad x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad F(x, y, \zeta) = 0.$$

$$\Updownarrow$$

$$H(x, y, \zeta) := \begin{pmatrix} x - [x - y]_+ \\ F(x, y, \zeta) \end{pmatrix} = 0,$$

where $[\cdot]_+ : \mathbb{R}^n \to \mathcal{K}$ denotes the nearest-point projection onto $\mathcal{K}$.

In Summary

$$\mathrm{SOCCP} \iff H(x, y, \zeta) = 0.$$

- H is nonsmooth due to the nonsmoothness of the projection operator $[\cdot]_+$.
- $[\cdot]_+$ is semismooth so that H is semismooth.
- This approach can be used to design and analyze nonsmooth Newton-type methods for solving $H(x, y, \zeta) = 0$.

Merit functions

III

Merit functions approach

Merit function approach

We will show that the SOCCP can be reformulated as an unconstrained differentiable minimization problem. The objectve function of such unconstrained minimization problem is called a **merit function**.
For simplicity, we assume $m = 1$, so $\mathcal{K} = \mathcal{K}^n$.

Jordan product

For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ and $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define their *Jordan product* associated with $\mathcal{K}^n$ as

$$x \circ y = (\langle x, y \rangle \ , \ y_1 x_2 + x_1 y_2)$$

The identity element under this product is $e := (1, 0, \cdots, 0)^T \in \mathbb{R}^n$.

Basic Property

Property

(a) $e \circ x = x \ , \quad \forall x \in \mathbb{R}^n \ .$

(b) $x \circ y = y \circ x \ , \quad \forall x, y \in \mathbb{R}^n \ .$

(c) $(x + y) \circ z = x \circ z + y \circ z \ , \quad \forall x, y, z \in \mathbb{R}^n \ .$

Remarks

1. The Jordan product is **not** associative.
2. $\mathcal{K}^n$ is **not** closed under Jordan product.

Jordan Product associated with SOC

We write x^2 to mean $x \circ x$ and write $x + y$ to mean the usual componentwise addition of vectors. Then we have:

- If $x \in \mathcal{K}^n$, then there exists a unique vector in $\mathcal{K}^n$, denoted by $x^{1/2}$ such that

$$(x^{1/2})^2 = x^{1/2} \circ x^{1/2} = x.$$

- For any $x \in \mathbb{R}^n$, we have $x^2 \in \mathcal{K}^n$. Hence, there exists a unique vector $(x^2)^{1/2} \in \mathcal{K}^n$ denoted by $|x|$.
- For any $x \in \mathbb{R}^n$, we have $x^2 = |x|^2$.

A partial order associated with $\mathcal{K}^n$

For any x, y in $\mathbb{R}^n$, we denote

$$x \succeq_{\mathcal{K}^n} y \Longleftrightarrow x - y \in \mathcal{K}^n,$$

and

$$x \succ_{\mathcal{K}^n} y \Longleftrightarrow x - y \in \mathrm{int}(\mathcal{K}^n).$$

Then we have:

- Any $x \in \mathbb{R}^n$ satisfies $|x| \succeq_{\mathcal{K}^n} x$.
- For any $x, y \succeq_{\mathcal{K}^n} 0$, if $x \succeq_{\mathcal{K}^n} y$, then $x^{1/2} \succeq_{\mathcal{K}^n} y^{1/2}$.
- For any $x, y \in \mathbb{R}^n$, if $x^2 \succeq_{\mathcal{K}^n} y^2$, then $|x| \succeq_{\mathcal{K}^n} |y|$.

Fischer-Burmeister function

We define $\phi : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$, by

$$\phi(x, y) = (x^2 + y^2)^{1/2} - (x + y).$$

ϕ is the Fischer-Burmeister function and is well-defined here.

A merit function

We define $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, by

$$\psi(x, y) = \|\phi(x, y)\|^2.$$

It is proved by Fukushima-Luo-Tseng in 2002 that

$$\psi(x, y) = 0 \Longleftrightarrow x, y \in \mathcal{K}^n \,, \; \langle x, y \rangle = 0$$

A smooth reformulation of SOCCP

Hence, we can reformulate SOCCP as an equivalent unconstrained minimization problem :

$$\text{(SOCCP)} \qquad \begin{cases} \text{Find } x, y \in \mathbb{R}^n, \zeta \in \mathbb{R}^l \text{ satisfying} \\ \langle x, y \rangle = 0 \ , \ \ x, y \in \mathcal{K}^n \\ x = F(\zeta) \ , \ \ y = G(\zeta) \end{cases}$$

$$\Updownarrow$$

$$\min_{\zeta \in \mathbb{R}^l} \left\{ f(\zeta) = \psi(F(\zeta), G(\zeta)) \right\} .$$

Thus f is a merit function for SOCCP.

Proposition

The function ψ is smooth everywhere. Moreover,
(i) If $x^2 + y^2 \in \operatorname{int}(\mathcal{K}^n)$, then

$$\begin{aligned} \nabla_x \psi(x,y) &= \left(L_x L^{-1}_{(x^2+y^2)^{1/2}} - I \right) 2\phi(x,y) \ , \\ \nabla_y \psi(x,y) &= \left(L_y L^{-1}_{(x^2+y^2)^{1/2}} - I \right) 2\phi(x,y) \ . \end{aligned}$$

where $L_x = \begin{bmatrix} x_1 & {x_2}^T \\ x_2 & x_1 I \end{bmatrix}$.

(ii) If $x^2 + y^2 \notin \operatorname{int}(\mathcal{K}^n)$, i.e., $\|x\|^2 + \|y\|^2 = 2\|x_1 x_2 + y_1 y_2\|$, then

$$\begin{aligned} \nabla_x \psi(x,y) &= \left(L_{\bar{x}} - I \right) 2\phi(x,y) \\ \nabla_y \psi(x,y) &= \left(L_{\bar{y}} - I \right) 2\phi(x,y) \ , \end{aligned}$$

where $\bar{x} = \left(\frac{\sqrt{2}x_1}{\sqrt{\|x\|^2+\|y\|^2}} \ , 0 \right)$ and $\bar{y} = \left(\frac{\sqrt{2}y_1}{\sqrt{\|x\|^2+\|y\|^2}} \ , 0 \right)$.

Two Important Lemmas

- For any $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with $x^2 + y^2 \notin \mathrm{int}(\mathcal{K}^n)$, we have

$$
\begin{aligned}
x_1^2 &= \|x_2\|^2, \\
y_1^2 &= \|y_2\|^2, \\
x_1 y_1 &= x_2^T y_2, \\
x_1 y_2 &= y_1 x_2.
\end{aligned}
$$

- For any $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with $x_1 x_2 + y_1 y_2 \neq 0$, we have

$$
\begin{aligned}
\left(x_1 - \frac{(x_1x_2+y_1y_2)^T x_2}{\|x_1x_2+y_1y_2\|}\right)^2 &\leq \left\|x_2 - x_1 \frac{x_1x_2+y_1y_2}{\|x_1x_2+y_1y_2\|}\right\|^2 \\
&\leq \|x\|^2 + \|y\|^2 - 2\|x_1x_2 + y_1y_2\|.
\end{aligned}
$$

Proposition

Let $f(\zeta) = \psi(F(\zeta), G(\zeta))$. Then f is smooth and, for every $\zeta \in \mathbb{R}^n$ with $\nabla F(\zeta), -\nabla G(\zeta)$ column monotone, we have either (i) $f(\zeta) = 0$ or (ii) $\nabla f(\zeta) \neq 0$.
In case (ii), if $\nabla G(\zeta)$ is invertible, then $\langle d(\zeta), \nabla f(\zeta) \rangle < 0$, where

$$d(\zeta) := -(\nabla G(\zeta)^{-1})^T \nabla_x \psi(F(\zeta), G(\zeta)).$$

Note: We say $M, N \in \mathbb{R}^{n \times n}$ are column monotone if, for any $u, v \in \mathbb{R}^n$,

$$Mu + Nv = 0 \Rightarrow u^T v \geq 0.$$

Advantages

- Every stationary point ζ of f is a global minimum if $\nabla F(\zeta), -\nabla G(\zeta)$ are column monotone.
- For every non-stationary point ζ of f, we provide a descent direction $d(\zeta)$ without computing the Jacobian of $F(\zeta)$.
- The assumption of $\nabla F(\zeta), -\nabla G(\zeta)$ being column monotone is reasonable and holds for second-order cone program (SOCP).

Another merit function

Another merit function motivated by Yamashita and Fukushima is defined as

$$\psi_{\mathrm{YF}}(x, y) = \psi_0(\langle x, y \rangle) + \psi(x, y),$$

where $\psi_0 : \mathbb{R} \to [0, \infty)$ is any smooth function satisfying

$$\psi_0(t) = 0 \ \ \forall t \le 0 \qquad \text{and} \qquad \psi_0'(t) > 0 \ \ \forall t > 0.$$

Remarks

- An example of ψ_0 is $\psi_0(t) = \frac{1}{4}(\max\{0, t\})^4$.
- Let $f_{\mathrm{YF}}(\zeta) = \psi_{\mathrm{YF}}(F(\zeta), G(\zeta))$. It follows from properties of ψ and ψ_0 that f_{YF} is also a smooth merit function.
- f_{YF} has similar properties as f. In addition, f_{YF} has properties of bounded level sets and provides error bounds when F, G are jointly strongly monotone, whereas f does not.

Proposition

Let $f_{\mathrm{YF}}(\zeta) := \psi_{\mathrm{YF}}(F(\zeta), G(\zeta))$. Then f_{YF} is smooth and, for every $\zeta \in \mathbb{R}^n$ with $\nabla F(\zeta)$, $-\nabla G(\zeta)$ are column monotone, either (i) $f_{\mathrm{YF}}(\zeta) = 0$ or (ii) $\nabla f_{\mathrm{YF}}(\zeta) \neq 0$.

In case (ii), if $\nabla G(\zeta)$ is invertible, then $\langle d_{\mathrm{YF}}(\zeta), \nabla f_{\mathrm{YF}}(\zeta)\rangle < 0$, where

$$
\begin{aligned}
& d_{\mathrm{YF}}(\zeta) := \\
& -(\nabla G(\zeta)^{-1})^T \Big(\psi_0'(\langle F(\zeta), G(\zeta)\rangle) G(\zeta) \\
& +\nabla_x \psi(F(\zeta), G(\zeta)) \Big).
\end{aligned}
$$

Error Bounds for f_{YF}

Suppose that F and G are jointly strongly monotone mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$. Also, suppose that SOCCP has a solution ζ^*. Then there exists a scalar $\tau > 0$ such that

$$
\begin{aligned}
&\tau\|\zeta - \zeta^*\|^2 \\
&\leq \max\{0, \langle F(\zeta), G(\zeta)\rangle\} + \|(-F(\zeta))_+\| \\
&+\|(-G(\zeta))_+\|, \quad \forall \zeta \in \mathbb{R}^n.
\end{aligned}
$$

Moreover,

$$\tau\|\zeta - \zeta^*\|^2 \leq \psi_0^{-1}\left(f_{\mathrm{YF}}(\zeta)\right) + 2\sqrt{2} f_{\mathrm{YF}}(\zeta)^{1/2},$$

for all $\zeta \in \mathbb{R}^n$.

Remark

F and G are jointly strongly monotone if there exist $\rho > 0$ such that

$$\langle F(\zeta) - F(\xi)\ ,\ G(\zeta) - G(\xi)\rangle \geq \rho\|\zeta - \xi\|^2,$$

for all $\zeta, \xi \in \mathbb{R}^n$.

Bounded level sets for f_{YF}

Suppose that F and G are differentiable, jointly strongly monotone mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$. Then the level set

$$\mathcal{L}(\gamma) := \{\zeta \in \mathbb{R}^n \mid f_{\mathrm{YF}}(\zeta) \leq \gamma\}$$

is nonempty and bounded for all $\gamma \geq 0$.

Remark:
The merit function f lacks these properties due to the absence of the term $\psi_0(\langle F(\zeta), G(\zeta)\rangle)$.

Algorithms

IV

Algorithms

Solve SOCP via merit function

The second-order cone program (SOCP) is

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b\ ,\ x \in \mathcal{K}, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$ has full rank and $m < n$. The KKT optimality conditions can be reformulated as

$$\min_{\zeta \in \mathbb{R}^n} \{ f(\zeta) = \psi(F(\zeta), G(\zeta)) \}\ ,$$

where

$$F(\zeta) = d + \left(I - A^T(AA^T)^{-1}A\right)\zeta$$
$$G(\zeta) = c - A^T(AA^T)^{-1}A\zeta,$$

with d satisfying $Ad = b$.

Algorithm (1)

- FR-CG method (Fletcher-Reeves):

$$
\begin{aligned}
\zeta^{k+1} &= \zeta^k + \alpha^k d^k \ , \\
d^k &= -\nabla f(\zeta^k) + \beta_{\mathrm{FR}}^k d^{k-1} \ ,
\end{aligned}
$$

where β_{FR}^k is given by

$$
\beta_{\mathrm{FR}}^k = \begin{cases} 0, & k = 1 \\ \dfrac{\nabla f(\zeta^k)^T \nabla f(\zeta^k)}{\nabla f(\zeta^{k-1})^T \nabla f(\zeta^{k-1})}, & k \geq 2. \end{cases}
$$

Algorithm (2)

- PR-CG method (Polak-Ribier):

$$\begin{aligned} \zeta^{k+1} &= \zeta^k + \alpha^k d^k \ , \\ d^k &= -\nabla f(\zeta^k) + \beta^k_{\mathrm{PR}} d^{k-1} \ , \end{aligned}$$

where β^k_{PR} is given by

$$\beta^k_{\mathrm{PR}} = \frac{\nabla f(\zeta^k)^T \left(\nabla f(\zeta^k) - \nabla f(\zeta^{k-1})\right)}{\nabla f(\zeta^{k-1})^T \nabla f(\zeta^{k-1})}.$$

Remark: Powell, in 1986, suggested to modify the PR-CG method by setting

$$\beta^k_+ = \max\{\beta^k_{\mathrm{PR}}, 0\}.$$

Algorithm (3)

- BFGS method (Broyden-Fletcher-Goldfarb-Shanno):

$$\begin{aligned} \zeta^{k+1} &= \zeta^k + \alpha^k d^k \ , \\ d^k &= -D^k \nabla f(\zeta^k) \ , \end{aligned}$$

where

$$\begin{aligned} D^{k+1} &= D^k + \left[1 + \frac{(q^k)^T D^k q^k}{(p^k)^T q^k} \right] \frac{p^k (p^k)^T}{(p^k)^T q^k} \\ &- \frac{D^k q^k (p^k)^T + p^k (p^k)^T D^k}{(p^k)^T q^k}, \end{aligned}$$

and $\left\{ \begin{aligned} p^k &= \zeta^{k+1} - \zeta^k, \\ q^k &= \nabla f(\zeta^{k+1}) - \nabla f(\zeta^k). \end{aligned} \right.$

Remark: Need big storage when dimension is large.

Algorithm (4)

- L-BFGS method: Rewrite BFGS update as

$$D^{k+1} = (V^k)^T D^k V^k + \rho^k p^k (p^k)^T,$$

where

$$\rho^k = \frac{1}{(q^k)^T p^k} \; , \quad V^k = I - \rho^k q^k (p^k)^T,$$

and

$$\begin{aligned}
D^k &= (V_{k-1}^T \cdots V_{k-m}^T) D_0^k (V_{k-m} \cdots V_{k-1}) \\
&+ \rho_{k-m} (V_{k-1}^T \cdots V_{k-m+1}^T) \\
&\quad p_{k-m} p_{k-m}^T (V_{k-m+1} \cdots V_{k-1}) \\
&+ \rho_{k-m+1} (V_{k-1}^T \cdots V_{k-m+2}^T) \\
&\quad p_{k-m+1} p_{k-m+1}^T (V_{k-m+2} \cdots V_{k-1}) \\
&+ \cdots + \\
&+ \rho_{k-1} p_{k-1} p_{k-1}^T .
\end{aligned}$$

Remarks

- D^{k+1} is obtained by updating D^k using the pair $\{p^k, q^k\}$.
- We store a modified version of D^k implicitly by storing the most recently computed m pairs $\{p^k, q^k\}$.
- When the $m+1$ pair is computed, the oldest pair is discarded and its location in memory taken by the new pair.
- The L-BFGS is suitable for large problems because it has been observed in practice that small values of m $(3 \leq m \leq 20)$ gives satisfactory results.

Initial scalings of D_0^k

Liu and Nocedal studied four different types of initial scalings of D_0^k for L-BFGS.

Scaling 1 $\quad D_0^k = D_0 \quad$ (no scaling).

Scaling 2 $\quad \gamma_0 D_0 \quad$ (only initial scaling).

Scaling 3 $\quad D_0^k = \gamma^k D_0, \text{where} \gamma^k = (q^k)^T p^k / \|q^k\|^2.$

Scaling 4 $\quad$ Same as Scaling 3 during the first m iterations. For $k > m$, $D_0^k = \mathrm{diag}(\omega_k^i)$,

$$\omega_k^i = \frac{p_{k-1}^i q_{k-1}^i + \cdots + p_{k-m}^i q_{k-m}^i}{(q_{k-1}^i)^2 + \cdots + (q_{k-m}^i)^2},$$

$i = 1, \cdots, n.$

Remark: Scaling 3 is the most efficient initial scaling.

Step-size

Armijo rule : Fix $\beta, \sigma \in (0,1)$ and $s > 0$, choose α^k that is the largest $\alpha \in \{s, s\beta, s\beta^2, \cdots\}$ satisfying

$$\frac{f(\zeta^k + \alpha d^k) - f(\zeta^k)}{\alpha} \leq \sigma \nabla f(\zeta^k)^T d^k,$$

i.e.,

$$f(\zeta^k + \alpha d^k) \leq f(\zeta^k) + \sigma \alpha \nabla f(\zeta^k)^T d^k.$$

Numerical Experiments

V

Numerical Experiments

Test Problems

Table 1: Set of test problems

Problem Names	n	m	# of nonzero elts of matrix A	structure of SOCs
nb	2383	123	192439	$[4 \times 1; 793 \times 3]$
nql30	6302	3680	26819	$[3602 \times 1; 900 \times 3]$
qssp30	7568	3691	36851	$[2 \times 1; 1891 \times 4]$
nb-L2-bessel	2641	123	209924	$[4 \times 1; 1 \times 123; 838 \times 3]$

In the structure of SOCs, for example,
$[\ 4 \times 1; 1 \times 123; 838 \times 3\ ]$
means that $\mathcal{K}$ consists of the product of 4 $\mathcal{K}^1$, one $\mathcal{K}^{123}$, and 838 $\mathcal{K}^3$.

FR-CG Convergence

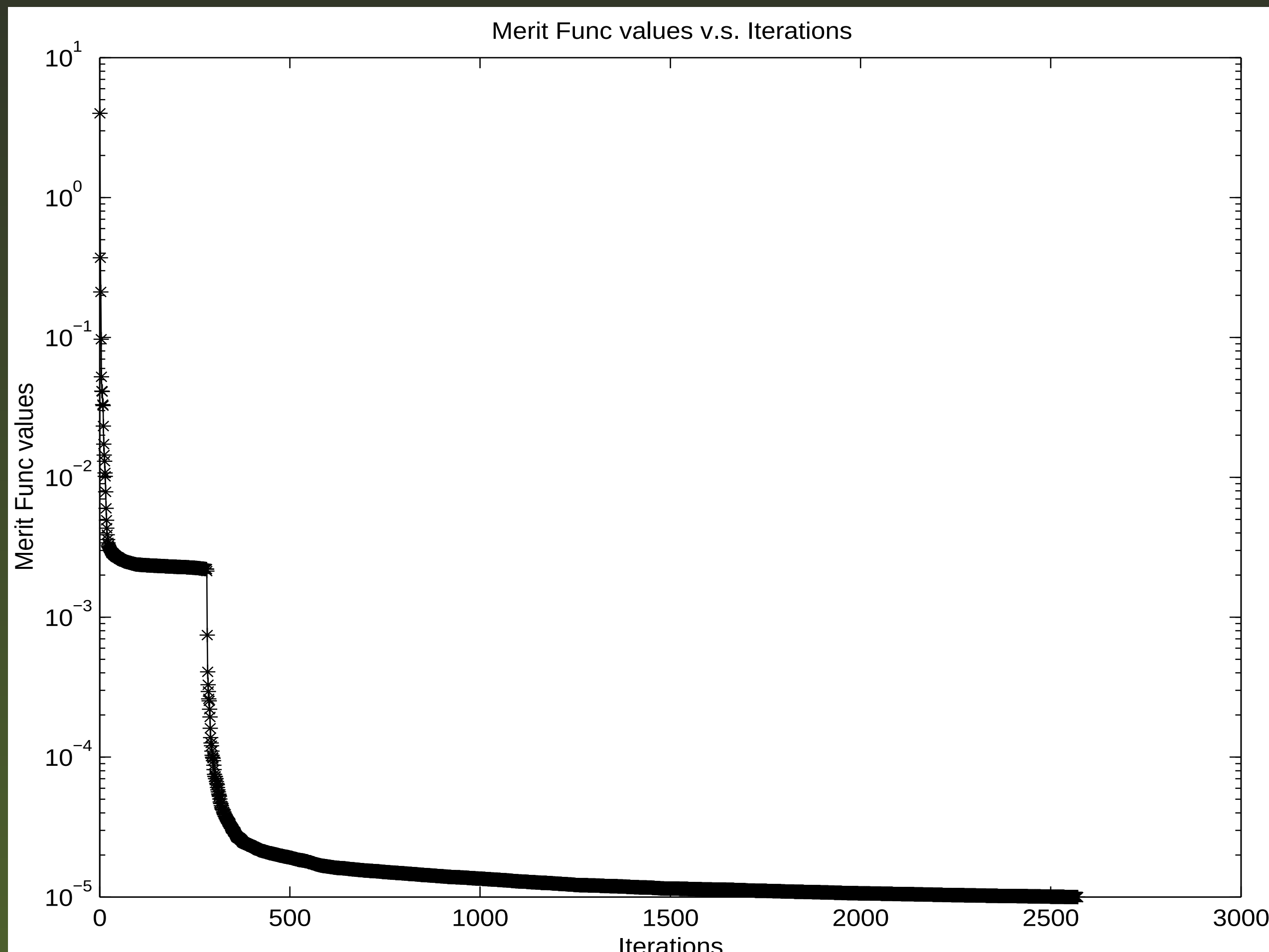


Figure 1: The convergence by FR-CG for 'nb'.

PR-CG Convergence

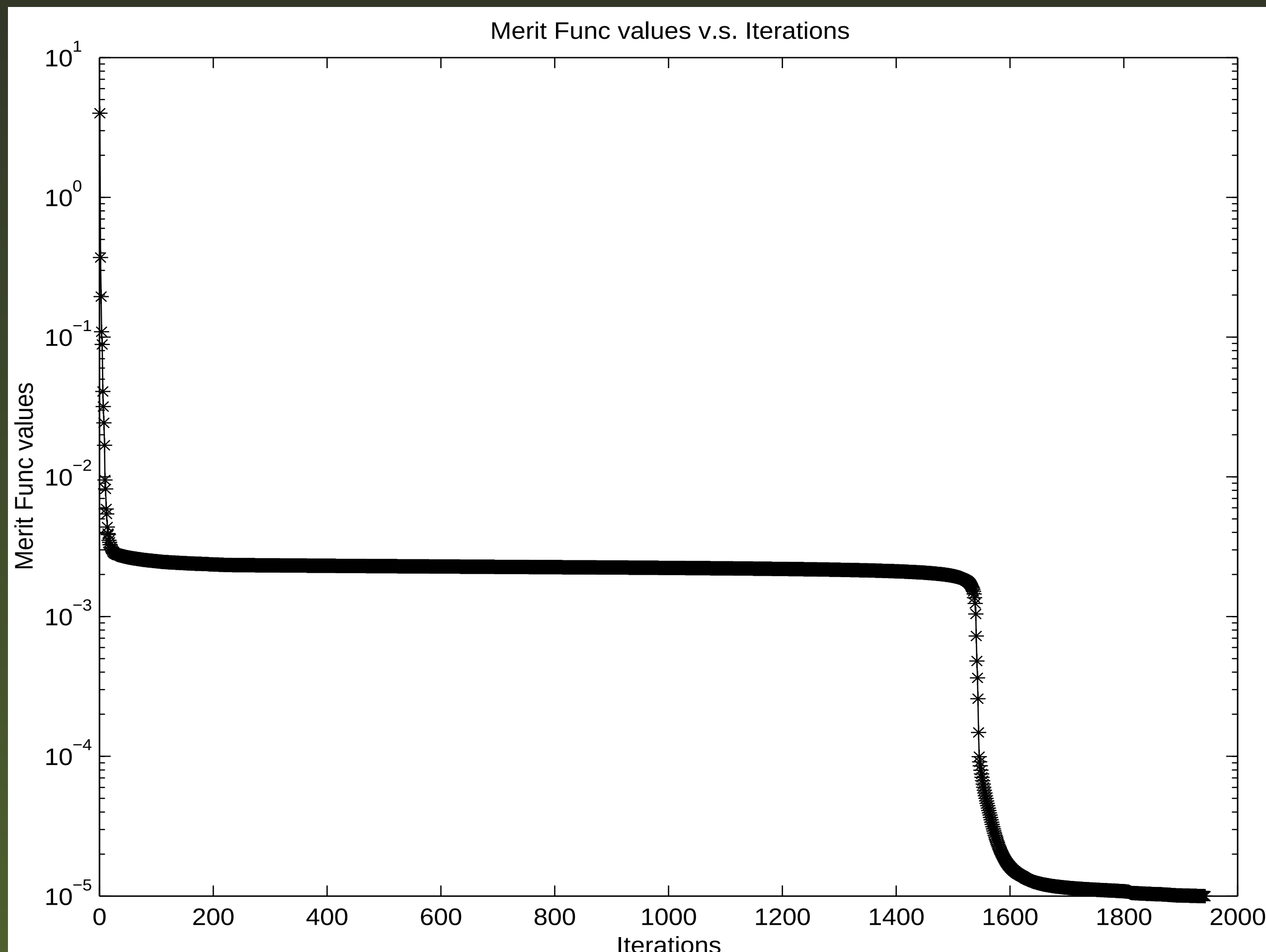


Figure 2: The convergence by PR-CG for 'nb'.

L-BFGS Convergence 1

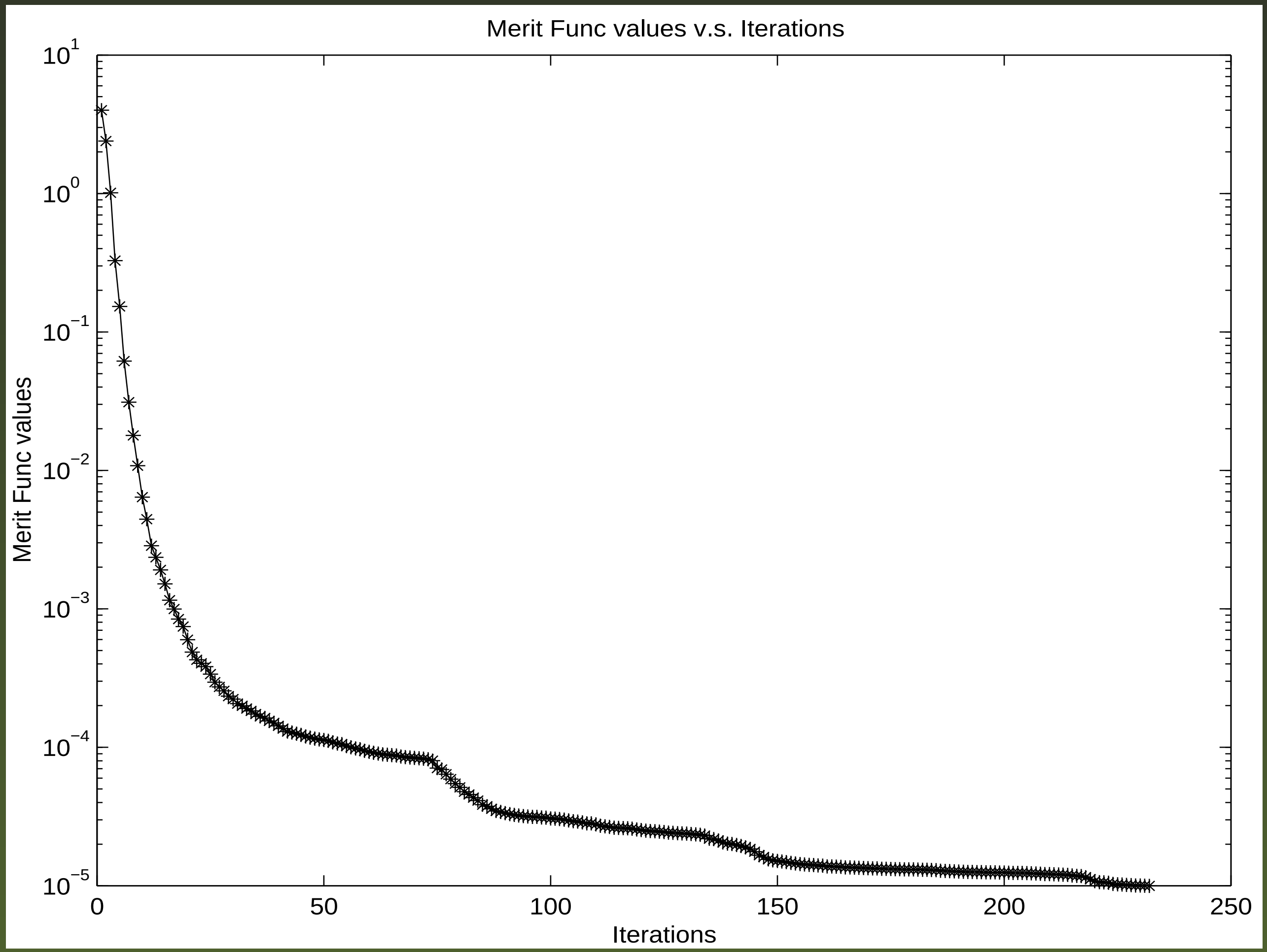


Figure 3: The convergence by L-BFGS (Chol) for 'nb'.

L-BFGS Convergence 2

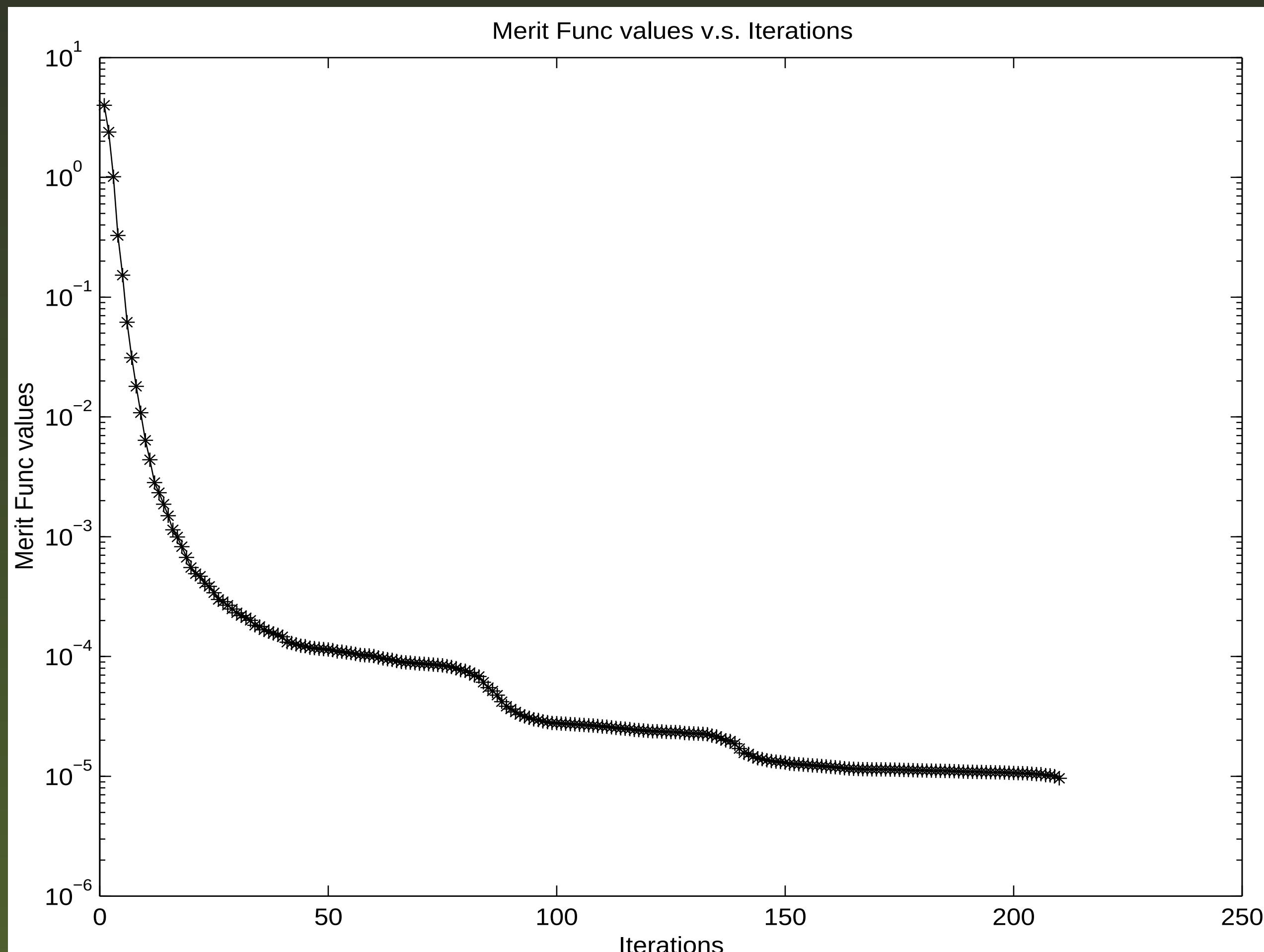


Figure 4: The convergence by L-BFGS (CG) for ’nb’.