On the convergence analysis of augmented Lagrangian method for matrix optimization

Strong variational sufficiency for nonlinear SDP

Chao Ding
Institute of Applied Mathematics
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Optimization in the Big Data Era
National University of Singapore
2022.12.15

中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

中国科学院
CHINESE ACADEMY OF SCIENCES

Acknowledgements

Based on the joint work with

- **Shiwei Wang** @ *Institute of Applied Mathematics, AMSS*
- **Yangjing Zhang** @ *Institute of Applied Mathematics, AMSS*
- **Xinyuan Zhao** @ *College of Applied Sciences, BJUT*

S.W. Wang, **D.**, Y.J. Zhang, and X.Y. Zhao, *Strong variational sufficiency for nonlinear semidefinite programming and its implications*, arXiv:2210.04448, 2022.

Outline

Generalized convexity

(Strong) variational sufficiency

Characterization of strong variational sufficiency I

ALM for NLSDP

Characterization of strong variational sufficiency II

Numerical experiments

Generalized convexity

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Definition

A set-valued mapping T is **monotone** if

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T.$$

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Definition

A set-valued mapping T is **monotone** if

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T.$$

$\star$ A proper, lsc function

$$\boxed{f \text{ is convex}} \iff \boxed{\partial f \text{ is (maximal) monotone}}$$

$\hat{\partial} f$: $f(x') \geq f(x) + \langle v, x' - x \rangle + o(\|x' - x\|)$
∂f: $\exists v^k \in \hat{\partial} f(x^k)$ such that $v^k \to v$, $x^k \to x$, $f(x^k) \to f(x)$

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Definition

A set-valued mapping T is **monotone** if

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T.$$

$\star$ A proper, lsc function

$$\boxed{f \text{ is convex}} \iff \boxed{\partial f \text{ is (maximal) monotone}}$$

$\hat{\partial} f$: $f(x') \geq f(x) + \langle v, x' - x \rangle + o(\|x' - x\|)$
∂f: $\exists v^k \in \hat{\partial} f(x^k)$ such that $v^k \to v$, $x^k \to x$, $f(x^k) \to f(x)$

- **Duality theorem**

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Definition

A set-valued mapping T is **monotone** if

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T.$$

$\star$ A proper, lsc function

$$\boxed{f \text{ is convex}} \iff \boxed{\partial f \text{ is (maximal) monotone}}$$

$\hat{\partial} f$: $f(x') \geq f(x) + \langle v, x' - x \rangle + o(\|x' - x\|)$
∂f: $\exists v^k \in \hat{\partial} f(x^k)$ such that $v^k \to v$, $x^k \to x$, $f(x^k) \to f(x)$

- **Duality theorem**
- **PPA** & **ALM**

Convexity

Definition

Let $f : \mathbb{X} \to [-\infty, +\infty]$. We define f to be a **convex** function on $\mathrm{dom} f$ if $\mathrm{epi} f$ is convex.

Definition

A set-valued mapping T is **monotone** if

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T.$$

$\star$ A proper, lsc function

$$\boxed{f \text{ is convex}} \iff \boxed{\partial f \text{ is (maximal) monotone}}$$

$\hat{\partial} f$: $f(x') \geq f(x) + \langle v, x' - x \rangle + o(\|x' - x\|)$
∂f: $\exists v^k \in \hat{\partial} f(x^k)$ such that $v^k \to v$, $x^k \to x$, $f(x^k) \to f(x)$

- **Duality theorem**
- **PPA** & **ALM**

Nonconvexity may lose the above properties

Beyond convexity (I): local convexity

Definition

If f is convex relative to some neighborhood $\mathcal{X}$ of a given $\bar{x}$, then f is a **locally convex** function.

Beyond convexity (I): local convexity

Definition

If f is convex relative to some neighborhood $\mathcal{X}$ of a given $\bar{x}$, then f is a **locally convex** function.

Definition (Pennanen MOR 02', in extension of local PPA)

A mapping T is **locally monotone** if there exists a neighborhood $\mathcal{X} \times \mathcal{V}$ of given $(\bar{x}, \bar{v})$ such that

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T \cap \mathcal{X} \times \mathcal{V}$$

Beyond convexity (I): local convexity

Definition

If f is convex relative to some neighborhood $\mathcal{X}$ of a given $\bar{x}$, then f is a **locally convex** function.

Definition (Pennanen MOR 02', in extension of local PPA)

A mapping T is **locally monotone** if there exists a neighborhood $\mathcal{X} \times \mathcal{V}$ of given $(\bar{x}, \bar{v})$ such that

$$\langle v' - v, x' - x \rangle \geq 0 \text{ for all } (x, v), (x', v') \in \mathrm{gph} T \cap \mathcal{X} \times \mathcal{V}$$

$$\boxed{\textbf{local convexity of } f \textbf{ on } \mathcal{X}} \rightleftarrows \boxed{\textbf{local monotonicity of } \partial f \textbf{ on } \mathcal{X} \times \mathbb{R}^n}$$

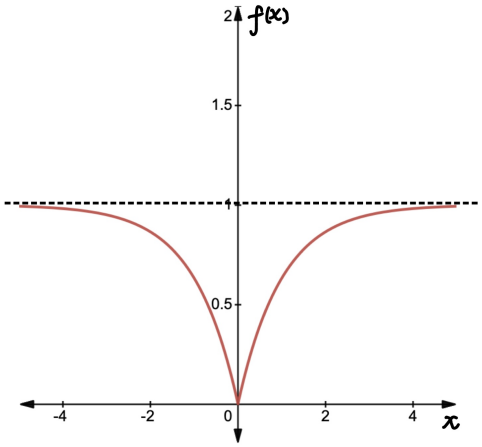
Beyond convexity (I): local convexity (con't)

$\boxed{\textbf{local convexity of } f \textbf{ on } \mathcal{X}} \not\Longleftarrow \boxed{\textbf{local monotonicity of } \partial f \textbf{ on } \mathcal{X} \times \mathbb{R}^n}$

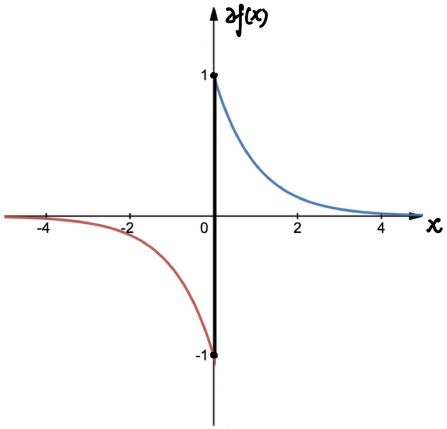
Beyond convexity (I): local convexity (con't)

$\boxed{\text{local convexity of } f \text{ on } \mathcal{X}} \not\Longleftarrow \boxed{\text{local monotonicity of } \partial f \text{ on } \mathcal{X} \times \mathbb{R}^n}$

Example 1: $f(x) = \max\{1 - e^x, 1 - e^{-x}\}$, $\partial f(x) = \begin{cases} \{-e^x\}, & x < 0 \\ [-1, 1], & x = 0 \\ \{e^{-x}\}, & x > 0 \end{cases}$



(a)



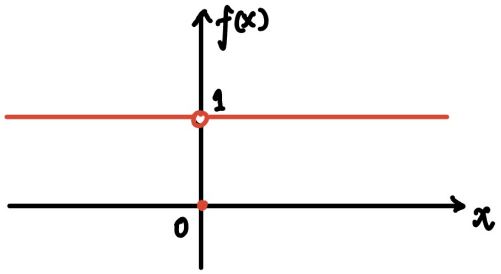
(b)

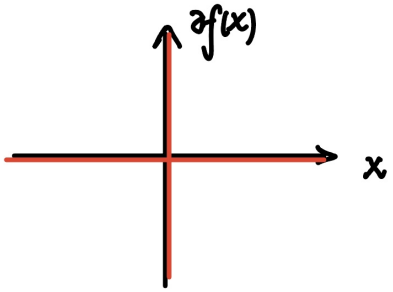
- f is **concave** on $(-\infty, 0]$ and $[0, \infty)$
- $0 \in \partial f(0)$; on a neighborhood of $(0, 0)$, ∂f is **locally maximal monotone**

Beyond convexity (II): variational convexity

Example 2:

$$f(x) = \begin{cases} 0, & x = 0 \\ 1, & x \neq 0 \end{cases}$$



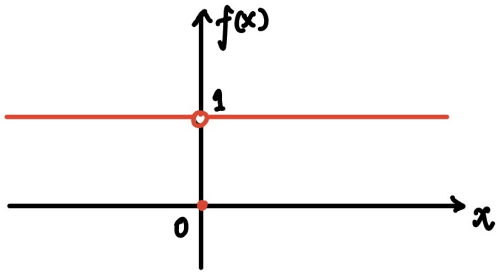


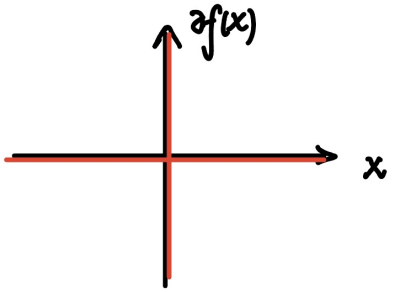
- $\mathrm{gph}\partial f$ is the union of horizontal and vertical axes of $\mathbb{R}^2$.
- no local monotonicity of ∂f around (0,0)

Beyond convexity (II): variational convexity

Example 2:

$$f(x) = \begin{cases} 0, & x = 0 \\ 1, & x \neq 0 \end{cases}$$





- $\mathrm{gph}\partial f$ is the union of horizontal and vertical axes of $\mathbb{R}^2$.
- no local monotonicity of ∂f around (0,0)

An observation: alternating the **neighborhood** $\mathcal{X}$ by

$$\mathcal{X}_\varepsilon := \{x \in \mathcal{X} \mid f(x) < f(0) + \varepsilon\} = \{0\}$$

for some $\varepsilon > 0$.

Then, ∂f is **(maximal) monotone** on $\mathcal{X}_\varepsilon \times \mathcal{V}$ with $\mathcal{V}$ is a neighborhood of 0.

Beyond convexity (II): variational convexity (con't)

Definition (f-local monotonicity of subgradient, Rockafellar VJM 19')

For lsc $f : \mathbb{R}^n \to (-\infty, +\infty]$ the mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is **f-locally monotone** around $(\bar{x}, \bar{v})$ if there is a neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v})$ such that

$$[\mathcal{X}_\varepsilon \times \mathcal{V}] \cap \operatorname{gph} \partial f \text{ is monotone in } \mathcal{X} \times \mathcal{V}$$

where $\mathcal{X}_\varepsilon := \{x \in \mathcal{X} \mid f(x) < f(\bar{x}) + \varepsilon\}$ for some $\varepsilon > 0$.

Beyond convexity (II): variational convexity (con't)

Definition (f-local monotonicity of subgradient, Rockafellar VJM 19')

For lsc $f : \mathbb{R}^n \to (-\infty, +\infty]$ the mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is **f-locally monotone** around $(\bar{x}, \bar{v})$ if there is a neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v})$ such that

$$[\mathcal{X}_\varepsilon \times \mathcal{V}] \cap \operatorname{gph} \partial f \text{ is monotone in } \mathcal{X} \times \mathcal{V}$$

where $\mathcal{X}_\varepsilon := \{x \in \mathcal{X} \mid f(x) < f(\bar{x}) + \varepsilon\}$ for some $\varepsilon > 0$.

Definition (Variational convexity, Rockafellar VJM 19')

Let $f : \mathbb{X} \to (-\infty, \infty]$ be a lsc function. f is **(strongly) variational convex** with respect to $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ if there exists an open convex neighborhood $\mathcal{X}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{v}$ such that there exists a proper lsc (strongly) convex function $h \leq f$ on $\mathcal{X}$ and a number $\varepsilon > 0$ such that

$$[\mathcal{X}_\varepsilon \times \mathcal{V}] \cap \operatorname{gph} \partial f = [\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial h,$$

and, for any (x, v) belonging to this common set, also $h(x) = f(x)$.

Beyond convexity (II): variational convexity (con't)

Definition (f-local monotonicity of subgradient, Rockafellar VJM 19')

For lsc $f : \mathbb{R}^n \to (-\infty, +\infty]$ the mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is **f-locally monotone** around $(\bar{x}, \bar{v})$ if there is a neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v})$ such that

$$[\mathcal{X}_\varepsilon \times \mathcal{V}] \cap \operatorname{gph} \partial f \text{ is monotone in } \mathcal{X} \times \mathcal{V}$$

where $\mathcal{X}_\varepsilon := \{x \in \mathcal{X} \mid f(x) < f(\bar{x}) + \varepsilon\}$ for some $\varepsilon > 0$.

Definition (Variational convexity, Rockafellar VJM 19')

Let $f : \mathbb{X} \to (-\infty, \infty]$ be a lsc function. f is **(strongly) variational convex** with respect to $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ if there exists an open convex neighborhood $\mathcal{X}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{v}$ such that there exists a proper lsc (strongly) convex function $h \leq f$ on $\mathcal{X}$ and a number $\varepsilon > 0$ such that

$$[\mathcal{X}_\varepsilon \times \mathcal{V}] \cap \operatorname{gph} \partial f = [\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial h,$$

and, for any (x, v) belonging to this common set, also $h(x) = f(x)$.

variational convexity of f $\iff$ **f-local monotonicity of ∂f**

Beyond convexity (II): variational convexity (con't)

- motivated by Poliquin **NA** 90'
- requires localization in **function values** and **subgradients**

Variational convexity vs. local convexity

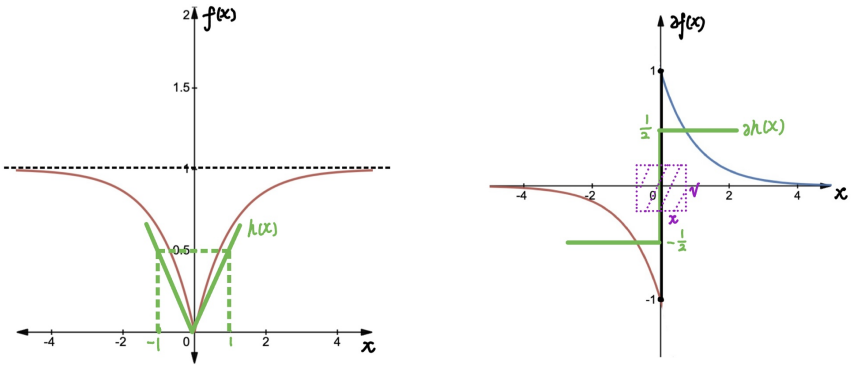
- $\star$ (strong) variational convexity $\Leftarrow$ local (strong) convexity
- $\star$ (strong) variational convexity $\not\Rightarrow$ local (strong) convexity

Variational convexity vs. local convexity

⋆ (strong) variational convexity $\Leftarrow$ local (strong) convexity

⋆ (strong) variational convexity $\not\Rightarrow$ local (strong) convexity

Back to **Example 1**: $f(x) = \max\{1 - e^x, 1 - e^{-x}\}$, $\partial f(x) = \begin{cases} \{-e^x\}, & x < 0 \\ [-1, 1], & x = 0 \\ \{e^{-x}\}, & x > 0 \end{cases}$



(a) (b)

- f is concave on $(-\infty, 0]$ and $[0, \infty)$
- variational convex at $\bar{x} = 0$ for $\bar{v} = 0$

  with $h(x) = \lambda|x|$ and any $\lambda \in (0, 1)$, the neighborhoods $\mathcal{X}$ and $\mathcal{V}$ being intervals adjusted for the size of λ.

(Strong) variational sufficiency

(Strong) variational sufficiency

Consider the general composite optimization problem:

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

- $f : \mathbb{X} \to \mathbb{R}$, $G : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable
- $\theta : \mathbb{Y} \to (-\infty, \infty]$ is a closed proper convex function

(Strong) variational sufficiency

Consider the general composite optimization problem:

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

- $f : \mathbb{X} \to \mathbb{R}$, $G : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable
- $\theta : \mathbb{Y} \to (-\infty, \infty]$ is a closed proper convex function

Define

- $\phi(x, u) = f(x) + \theta(G(x) + u)$ be the **perturbed objective function** with the parameter u
- For $\rho > 0$, the **augmented objective function** $\phi_\rho(x, u) := \phi(x, u) + \frac{\rho}{2}\|u\|^2$

(Strong) variational sufficiency

Consider the general composite optimization problem:

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

- $f : \mathbb{X} \to \mathbb{R}$, $G : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable
- $\theta : \mathbb{Y} \to (-\infty, \infty]$ is a closed proper convex function

Define

- $\phi(x, u) = f(x) + \theta(G(x) + u)$ be the **perturbed objective function** with the parameter u
- For $\rho > 0$, the **augmented objective function** $\phi_\rho(x, u) := \phi(x, u) + \frac{\rho}{2}\|u\|^2$

Definition

The **(strong) variational sufficient condition** for local optimality holds with respect to $\bar{x}$ and $\overline{Y}$ satisfying the first order condition if there exists $\bar{\rho} > 0$ such that **augmented objective function** $\phi_{\bar{\rho}}(x, u)$ is **(strong) variational convex** with respect to the pair $\big((\bar{x}, 0), (0, \overline{Y})\big)$ in $\operatorname{gph} \partial\phi_{\bar{\rho}}$, where $\phi_{\bar{\rho}}(x, u) := \phi(x, u) + \frac{\bar{\rho}}{2}\|u\|^2$.

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Augmented Lagrangian function

$$\mathcal{L}_\rho(x, Y) := f(x) + e_{1/\rho}\theta(G(x) + \frac{Y}{\rho}) - \frac{\|Y\|^2}{2\rho},$$

where $e_{1/\rho}\theta(Z) = \inf\{\theta(W) + \frac{\rho}{2}\|W - Z\|^2\}$

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Augmented Lagrangian function

$$\mathcal{L}_\rho(x, Y) := f(x) + e_{1/\rho}\theta(G(x) + \frac{Y}{\rho}) - \frac{\|Y\|^2}{2\rho},$$

where $e_{1/\rho}\theta(Z) = \inf\{\theta(W) + \frac{\rho}{2}\|W - Z\|^2\}$

Lagrangian dual

$$(D) \quad \max_Y h(Y) := \inf_x \mathcal{L}(x, Y)$$

$$(P) \quad \min_x \phi(x) := \sup_Y \mathcal{L}(x, Y)$$

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Augmented Lagrangian function

$$\mathcal{L}_\rho(x, Y) := f(x) + e_{1/\rho}\theta(G(x) + \frac{Y}{\rho}) - \frac{\|Y\|^2}{2\rho},$$

where $e_{1/\rho}\theta(Z) = \inf\{\theta(W) + \frac{\rho}{2}\|W - Z\|^2\}$

Lagrangian dual

$$(D) \quad \max_Y h(Y) := \inf_x \mathcal{L}(x, Y)$$

$$(P) \quad \min_x \phi(x) := \sup_Y \mathcal{L}(x, Y)$$

Duality theorem (convex)

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Augmented Lagrangian function

$$\mathcal{L}_\rho(x, Y) := f(x) + e_{1/\rho}\theta(G(x) + \frac{Y}{\rho}) - \frac{\|Y\|^2}{2\rho},$$

where $e_{1/\rho}\theta(Z) = \inf\{\theta(W) + \frac{\rho}{2}\|W - Z\|^2\}$

Lagrangian dual

$$(D) \quad \max_Y h(Y) := \inf_x \mathcal{L}(x, Y)$$

$$(P) \quad \min_x \phi(x) := \sup_Y \mathcal{L}(x, Y)$$

Augmented Lagrangian dual

$$(\widehat{D}) \quad \max_{Y \in \mathcal{Y}} \widehat{h}(Y) := \inf_{x \in \mathcal{X}} \mathcal{L}_{\bar{\rho}}(x, Y)$$

$$(\widehat{P}) \quad \min_{x \in \mathcal{X}} \widehat{\phi}(x) := \sup_{Y \in \mathcal{Y}} \mathcal{L}_{\bar{\rho}}(x, Y)$$

Duality theorem (convex)

Local Augmented Lagrangian Duality

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x))$$

Generalized Lagrangian function

$$\mathcal{L}(x, Y) := L(x, Y) - \theta^*(Y) = f(x) + \langle G(x), Y \rangle - \theta^*(Y)$$

Augmented Lagrangian function

$$\mathcal{L}_\rho(x, Y) := f(x) + e_{1/\rho}\theta(G(x) + \frac{Y}{\rho}) - \frac{\|Y\|^2}{2\rho},$$

where $e_{1/\rho}\theta(Z) = \inf\{\theta(W) + \frac{\rho}{2}\|W - Z\|^2\}$

Lagrangian dual

$$(D) \quad \max_Y h(Y) := \inf_x \mathcal{L}(x, Y)$$

$$(P) \quad \min_x \phi(x) := \sup_Y \mathcal{L}(x, Y)$$

Duality theorem (convex)

Augmented Lagrangian dual

$$(\widehat{D}) \quad \max_{Y \in \mathcal{Y}} \widehat{h}(Y) := \inf_{x \in \mathcal{X}} \mathcal{L}_{\bar{\rho}}(x, Y)$$

$$(\widehat{P}) \quad \min_{x \in \mathcal{X}} \widehat{\phi}(x) := \sup_{Y \in \mathcal{Y}} \mathcal{L}_{\bar{\rho}}(x, Y)$$

?

Local Augmented Lagrangian Duality (con't)

Theorem (Rockafellar MP 22')

Suppose **variational sufficient condition** *with respect to* $(\bar{x}, \overline{Y})$ *for local optimality is satisfied. Then, the problems* $(\widehat{P})$ *and* $(\widehat{D})$ *have optimal solutions with the same optimal value, and*

$$\bar{x} \text{ sloves } (\widehat{P}) \quad \Longleftrightarrow \quad \bar{x} \text{ minimizes in (P) relative to } \mathcal{X}.$$

Moreover, the following conditions on $(\bar{x}, \overline{Y}) \in \operatorname{int} \mathcal{X} \times \operatorname{int} \mathcal{Y}$ *are equivalent and guarantee that* $\bar{x}$ *is locally optimal relative to* $\mathcal{X}$ *in (P) with the objective value agreeing with the common optimal value in* $(\widehat{P})$ *and* $(\widehat{D})$ *as well as with* $\mathcal{L}_{\bar{\rho}}(\bar{x}, \overline{Y})$.

(a) $(\bar{x}, \overline{Y}) \in S_{KKT}(0, 0)$,

(b) $\bar{x}$ *minimize in* $(\widehat{P})$ *and* $\overline{Y}$ *maximize in* $(\widehat{D})$,

(c) $(\bar{x}, \overline{Y})$ *is a saddle point of* $\mathcal{L}_{\bar{\rho}}$ *on* $\mathcal{X} \times \mathcal{Y}$,

(d) $(\bar{x}, \overline{Y})$ *is a saddle point of* $\mathcal{L}_{\rho}$ *on* $\mathcal{X} \times \mathcal{Y}$ *for every* $\rho \geq \bar{\rho}$,

Local Augmented Lagrangian Duality (con't)

Theorem (Rockafellar MP 22')

Suppose **variational sufficient condition** *with respect to* $(\bar{x}, \overline{Y})$ *for local optimality is satisfied. Then, the problems* $(\widehat{P})$ *and* $(\widehat{D})$ *have optimal solutions with the same optimal value, and*

$$\bar{x} \text{ sloves } (\widehat{P}) \quad \Longleftrightarrow \quad \bar{x} \text{ minimizes in (P) relative to } \mathcal{X}.$$

Moreover, the following conditions on $(\bar{x}, \overline{Y}) \in \operatorname{int} \mathcal{X} \times \operatorname{int} \mathcal{Y}$ *are equivalent and guarantee that* $\bar{x}$ *is locally optimal relative to* $\mathcal{X}$ *in (P) with the objective value agreeing with the common optimal value in* $(\widehat{P})$ *and* $(\widehat{D})$ *as well as with* $\mathcal{L}_{\bar{\rho}}(\bar{x}, \overline{Y})$.

(a) $(\bar{x}, \overline{Y}) \in S_{KKT}(0, 0)$,

(b) $\bar{x}$ *minimize in* $(\widehat{P})$ *and* $\overline{Y}$ *maximize in* $(\widehat{D})$,

(c) $(\bar{x}, \overline{Y})$ *is a saddle point of* $\mathcal{L}_{\bar{\rho}}$ *on* $\mathcal{X} \times \mathcal{Y}$,

(d) $(\bar{x}, \overline{Y})$ *is a saddle point of* $\mathcal{L}_{\rho}$ *on* $\mathcal{X} \times \mathcal{Y}$ *for every* $\rho \geq \bar{\rho}$,

can be extended to Riemannian manifold optimization: Zhou and **D.**, 2022

PPA & ALM

- **PPA for dual**

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\rho^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + \rho^k \left[\nabla_Y \mathcal{L}_{\rho^k}(x, Y^k)\right] \end{cases}$$

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\rho^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + \rho^k \left[\nabla_Y \mathcal{L}_{\rho^k}(x, Y^k)\right] \end{cases}$$

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + c^k \left[\nabla_Y \mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\right] \end{cases}$$

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\rho^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + \rho^k \left[\nabla_Y \mathcal{L}_{\rho^k}(x, Y^k)\right] \end{cases}$$

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + c^k \left[\nabla_Y \mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\right] \end{cases}$$

- **PPA $\Leftrightarrow$ ALM**

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\rho^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + \rho^k \left[\nabla_Y \mathcal{L}_{\rho^k}(x, Y^k)\right] \end{cases}$$

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + c^k \left[\nabla_Y \mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\right] \end{cases}$$

- **PPA $\Leftrightarrow$ ALM**

$$\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$
$$= \inf_x \mathcal{L}_{\rho^k}(x, Y^k)$$

PPA & ALM

- **PPA for dual**

Convex

$$Y^{k+1} \approx P_h(Y^k)$$
$$= \arg\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$

Nonconvex with variational sufficiency

$$Y^{k+1} \approx P_{\widehat{h}}(Y^k)$$
$$= \arg\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$

- **ALM for primal**

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\rho^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + \rho^k \left[\nabla_Y \mathcal{L}_{\rho^k}(x, Y^k)\right] \end{cases}$$

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\} \\ Y^{k+1} = Y^k + c^k \left[\nabla_Y \mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)\right] \end{cases}$$

- **PPA $\Leftrightarrow$ ALM**

$$\max_Y \{h(Y) - \frac{1}{2\rho^k}\|Y - Y^k\|^2\}$$
$$= \inf_x \mathcal{L}_{\rho^k}(x, Y^k)$$

$$\max_{Y \in \mathcal{Y}} \{\widehat{h}(Y) - \frac{1}{2c^k}\|Y - Y^k\|^2\}$$
$$= \inf_{x \in \mathcal{X}} \mathcal{L}_{\bar{\rho}+c^k}(x, Y^k)$$

Characterization of strong variational sufficiency I

Strong variational sufficiency for polyhedral

For polyhedral (Rockafellar **MP** 22'):

$$\text{strong variational sufficiency} \iff \text{strong SOSC}$$

$\star$ In particular, this case include NLP and NLSOC with $G(x) \neq 0$.

For general OP (Rockafellar **MP** 22'):

Proposition

For general optimization problem

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x)),$$

the augmented Lagrangian $\mathcal{L}_\rho$ *is* C^{1+}. *Let* $\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$ *denote the* **partial Hessian bundle**. *Then the* **strong variational sufficient condition** *for local optimality with respect to* $(\bar{x}, \overline{Y})$ *is equivalent to*

every $H \in \pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$ *is* **positive definite.**

Strong variational sufficiency for general OP (con't)

Proposition (Rockafellar MP 22')

For general optimization problem

$$\min_{x \in \mathbb{X}} f(x) + \theta(G(x)),$$

strong variational sufficient condition *for local optimality with respect to* $(\bar{x}, \overline{Y})$ *is equivalent to that every* $q \in \operatorname{quad} \theta(G(\bar{x}) \mid \overline{Y})$ *has*

$$\frac{1}{2}\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle + q(G'(\bar{x})d) > 0 \quad \text{when } d \neq 0,$$

where

$$\operatorname{quad} \theta(G(\bar{x}) \mid \overline{Y}) = \begin{cases} \text{the collection of generalized quadratic forms } q \text{ for which} \\ \exists (X^k, Y^k) \to (G(\bar{x}), \overline{Y}) \text{ with } \theta \text{ generalized twice differentiable} \\ \text{at } X^k \text{ for } Y^k \text{ and such that the generalized quadratic} \\ \text{forms } q_k = \frac{1}{2} d^2\theta(X^k \mid Y^k) \text{ converge epigraphically to } q. \end{cases}$$

is the **quadratic bundle** *of* θ.

Our result: toy example

Consider the following NLSOC:

$$\min f(x) + \delta_{\mathcal{K}}(g(x)),$$

where $\mathcal{K} = \{(u_1, \ldots, u_m) \in \mathbb{R}^m \mid u_1 \geq \|(u_2, \ldots, u_m)\|\} = \{u \mid h(u) \leq 0\}$.

$$\mathcal{M}_{\text{soc}}(\bar{x}) := \left\{ y \in \mathbb{R}^m \,\middle|\, \begin{array}{l} L'_x(\bar{x}, y) = 0, \\ \mathcal{K} \ni g(\bar{x}) \perp y \in \mathcal{K}^\circ \end{array} \right\}.$$

Proposition

Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSOC and $\bar{y} \in \mathcal{M}_{SOC}(\bar{x})$. The **strong variational sufficient condition** *with respect to $(\bar{x}, \bar{y})$ holds if and only if the* **strong second order sufficient condition** *holds at $(\bar{x}, \bar{y})$, i.e.,*

$$\langle L''_{xx}(\bar{x}, \bar{y})d, d\rangle - \Upsilon_{g(\bar{x})}\big(\bar{y}, g'(\bar{x})d\big) > 0 \quad \forall\, 0 \neq g'(\bar{x})d \in \operatorname{aff} \mathcal{C}_{\mathcal{K}}(g(\bar{x}), \bar{y}),$$

where $\Upsilon_{g(\bar{x})}\big(\bar{y}, g'(\bar{x})d\big)$ is the so-called "σ-term".

Our result: toy example (con't)

Main idea: calculating the quadratic bundle for NLSOC:

- when $g(\bar{x}) \neq 0 \Longrightarrow$ polyhedral case
- when $g(\bar{x}) = 0 \Longrightarrow$ the $\sigma-$term happens to be 0, which makes the calculation much simpler

Our result: toy example (con't)

Main idea: calculating the quadratic bundle for NLSOC:

- when $g(\bar{x}) \neq 0 \Longrightarrow$ polyhedral case
- when $g(\bar{x}) = 0 \Longrightarrow$ the $\sigma-$term happens to be 0, which makes the calculation much simpler

$\star$ **main tool** (Mohammadi, Mordukhovich and Sarabi **TAMS** 20'): Suppose $v \in \mathcal{N}_{\mathcal{K}}(x)$,

$$d^2\delta_{\mathcal{K}}(x \mid v)(w) = -\Upsilon_x\big(v, w\big) + \delta_{\mathcal{C}_{\mathcal{K}}(x,v)}(w)$$

Our result: toy example (con't)

Main idea: calculating the quadratic bundle for NLSOC:

- when $g(\bar{x}) \neq 0 \Longrightarrow$ polyhedral case
- when $g(\bar{x}) = 0 \Longrightarrow$ the $\sigma-$term happens to be 0, which makes the calculation much simpler

$\star$ **main tool** (Mohammadi, Mordukhovich and Sarabi **TAMS** 20'): Suppose $v \in \mathcal{N}_{\mathcal{K}}(x)$,

$$d^2\delta_{\mathcal{K}}(x \mid v)(w) = -\Upsilon_x\big(v, w\big) + \delta_{\mathcal{C}_{\mathcal{K}}(x,v)}(w)$$

Extension of Rockafellar **MP** 22' in which only consider the case $g(\bar{x}) \neq 0$

Our result: toy example (con't)

Main idea: calculating the quadratic bundle for NLSOC:

- when $g(\bar{x}) \neq 0 \Longrightarrow$ polyhedral case
- when $g(\bar{x}) = 0 \Longrightarrow$ the $\sigma-$term happens to be 0, which makes the calculation much simpler

$\star$ **main tool** (Mohammadi, Mordukhovich and Sarabi **TAMS** 20'): Suppose $v \in \mathcal{N}_{\mathcal{K}}(x)$,

$$d^2\delta_{\mathcal{K}}(x \mid v)(w) = -\Upsilon_x\big(v, w\big) + \delta_{\mathcal{C}_{\mathcal{K}}(x,v)}(w)$$

Extension of Rockafellar **MP** 22' in which only consider the case $g(\bar{x}) \neq 0$

However,

what about **NLSDP? general MOP?**

Our results: for NLSDP

Consider the following NLSDP:

$$\min f(x) + \delta_{\mathbb{S}^n_+}(G(x))$$

Our results: for NLSDP

Consider the following NLSDP:

$$\min f(x) + \delta_{\mathbb{S}^n_+}(G(x))$$

Theorem

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP and $\overline{Y}$ be a corresponding multiple. The following conditions are **equivalent**.

Our results: for NLSDP

Consider the following NLSDP:

$$\min f(x) + \delta_{\mathbb{S}^n_+}(G(x))$$

Theorem

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP and $\overline{Y}$ be a corresponding multiple. The following conditions are **equivalent**.

(i) *The* **strong variational sufficient condition** *with respect to $(\bar{x}, \overline{Y})$ holds.*

Our results: for NLSDP

Consider the following NLSDP:

$$\min f(x) + \delta_{\mathbb{S}^n_+}(G(x))$$

Theorem

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP and $\overline{Y}$ be a corresponding multiple. The following conditions are **equivalent**.

(i) *The* **strong variational sufficient condition** *with respect to $(\bar{x}, \overline{Y})$ holds.*

(ii) *The* **strongly second order sufficient condition** *holds at $(\bar{x}, \overline{Y})$,*

Our results: for NLSDP

Consider the following NLSDP:

$$\min f(x) + \delta_{\mathbb{S}^n_+}(G(x))$$

Theorem

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP and $\overline{Y}$ be a corresponding multiple. The following conditions are **equivalent**.

(i) *The* **strong variational sufficient condition** *with respect to $(\bar{x}, \overline{Y})$ holds.*

(ii) *The* **strongly second order sufficient condition** *holds at $(\bar{x}, \overline{Y})$, i.e.,*

$$\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big) > 0 \quad \forall\, 0 \neq G'(\bar{x})d \in \operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y}),$$

*where $\Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big) := 2\langle \overline{Y}, (G'(\bar{x})d)G(\bar{x})^\dagger(G'(\bar{x})d)\rangle$ is the so-called "σ-**term**" and $G(\bar{x})^\dagger$ is the generalized inverse matrix of $G(\bar{x})$.*

Comparison with Khanh, Mordukhovich and Phat's approach

Recently, Khanh, Mordukhovich and Phat 22' provides a sufficient condition for strong variational sufficiency under the following second order qualification:

$$\ker G'(\bar{x})^* \cap \partial^2\theta(G(\bar{x}), \overline{Y})(0) = \{0\},$$

which implies

strong SOSC $\implies$ strong variational sufficiency

Comparison with Khanh, Mordukhovich and Phat's approach

Recently, Khanh, Mordukhovich and Phat 22' provides a sufficient condition for strong variational sufficiency under the following second order qualification:

$$\ker G'(\bar{x})^* \cap \partial^2\theta(G(\bar{x}), \overline{Y})(0) = \{0\},$$

which implies

$$\text{strong SOSC} \quad \Longrightarrow \quad \text{strong variational sufficiency}$$

Proposition

Let $\bar{x} \in \mathbb{X}$ be a stationary point to NLSDP and $\overline{Y}$ be a corresponding multiplier. Then, the second order qualification condition at $(G(\bar{x}), \overline{Y})$ is equivalent to the nondegeneracy condition (LICQ) at $(G(\bar{x}), \overline{Y})$, i.e.,

$$\ker G'(\bar{x})^* \cap \operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^\perp = \{0\}.$$

Comparison with Khanh, Mordukhovich and Phat's approach

Recently, Khanh, Mordukhovich and Phat 22' provides a sufficient condition for strong variational sufficiency under the following second order qualification:

$$\ker G'(\bar{x})^* \cap \partial^2\theta(G(\bar{x}), \overline{Y})(0) = \{0\},$$

which implies

$$\text{strong SOSC} \quad \Longrightarrow \quad \text{strong variational sufficiency}$$

Proposition

Let $\bar{x} \in \mathbb{X}$ be a stationary point to NLSDP and $\overline{Y}$ be a corresponding multiplier. Then, the second order qualification condition at $(G(\bar{x}), \overline{Y})$ is equivalent to the nondegeneracy condition (LICQ) at $(G(\bar{x}), \overline{Y})$, i.e.,

$$\ker G'(\bar{x})^* \cap \operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^\perp = \{0\}.$$

Proof sketch: We only need to verify $\operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^\perp = \partial^2\theta(G(\bar{x}), \overline{Y})(0)$, which can be obtained directly from Ding, Sun and Ye **MP** 14'.

Comparison with Khanh, Mordukhovich and Phat's approach (con't)

Khanh, Mordukhovich and Phat 22':

$$\text{strong SOSC} + \text{LICQ} \quad \Longrightarrow \quad \text{strong variational sufficiency}$$

Comparison with Khanh, Mordukhovich and Phat's approach (con't)

Khanh, Mordukhovich and Phat 22':

$$\text{strong SOSC} + \text{LICQ} \quad \Longrightarrow \quad \text{strong variational sufficiency}$$

This work: without any constraint qualification or the uniqueness of multiplier

$$\text{strong SOSC} \quad \Longleftrightarrow \quad \text{strong variational sufficiency}$$

ALM for NLSDP

Inexact ALM for non-convex NLSDP

Algorithm 1 Inexact ALM for non-convex NLSDP

Input: Let $(x^0, Y^0) \in \mathbb{X} \times \mathbb{S}^n$, $\rho^0, \rho_0^0 > 0$. Set $k := 0$. $\{\epsilon_k\}_{k \geq 0}$ with $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$ and set $k := 0$.

Output: x^{k+1} and Y^{k+1}

1: If (x^k, Y^k) satisfies a suitable termination criterion: STOP.

2: Compute x^{k+1} such that

$$x^{k+1} \approx \bar{x}^{k+1} = \operatorname{argmin}_{x \in \mathcal{X}} \mathcal{L}_{\rho^k}(x, Y^k),$$

3: Update the vector of multipliers to

$$Y^{k+1} := Y^k + \rho_0^k (\mathcal{L}_{\rho^k})'_Y (x^{k+1}, Y^k),$$

where $\rho_0^k = \rho^k - \bar{\rho}$ and $\bar{\rho} > 0$.

4: Update nondecreasing positive sequence ρ^{k+1} according to certain rules.

5: Set $k \leftarrow k + 1$ and go to **Step 1**.

Inexact ALM for non-convex NLSDP

Algorithm 2 Inexact ALM for non-convex NLSDP

Input: Let $(x^0, Y^0) \in \mathbb{X} \times \mathbb{S}^n$, $\rho^0, \rho_0^0 > 0$. Set $k := 0$. $\{\epsilon_k\}_{k\geq 0}$ with $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$ and set $k := 0$.

Output: x^{k+1} and Y^{k+1}

1: If (x^k, Y^k) satisfies a suitable termination criterion: STOP.

2: Compute x^{k+1} such that

$$x^{k+1} \approx \bar{x}^{k+1} = \operatorname{argmin}_{x\in\mathcal{X}} \mathcal{L}_{\rho^k}(x, Y^k),$$

3: Update the vector of multipliers to

$$Y^{k+1} := Y^k + \rho_0^k (\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k),$$

where $\rho_0^k = \rho^k - \bar{\rho}$ and $\bar{\rho} > 0$.

4: Update nondecreasing positive sequence ρ^{k+1} according to certain rules.

5: Set $k \leftarrow k+1$ and go to **Step 1**.

Under strong variational sufficiency, an **implementable** stopping criterion for the updating of x^{k+1}:

$$\sqrt{\rho_0^k}\|(\mathcal{L}_{\rho^k})'_x(x^{k+1}, Y^k)\| \leq \begin{cases} (a)\ \epsilon'_k, \\ (b)\ \epsilon'_k \min\{1, \|\rho_0^k(\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|\}, \\ (c)\ \epsilon'_k \min\{1, \|\rho_0^k(\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|^2\}, \end{cases}$$

where $\epsilon'_k = \epsilon_k\sqrt{s}$ for some $s > 0$.

Bounded linear regularity

Definition

Let $D_1, D_2, \ldots, D_m \subseteq \mathbb{X}$ be closed convex sets for some positive integer m. Suppose that $D := D_1 \cap D_2 \cap \ldots \cap D_m$ is non-empty. The collection $\{D_1, D_2, \ldots, D_m\}$ is said to be boundedly linearly regular if for every bounded set $\mathcal{B} \subseteq \mathbb{X}$, there exists a constant $\kappa > 0$ such that

$$\operatorname{dist}(x, D) \leqslant \kappa \max\left\{\operatorname{dist}\left(x, D_1\right), \ldots, \operatorname{dist}\left(x, D_m\right)\right\} \quad \forall x \in \mathcal{B}.$$

$$\mathcal{G}_1(\bar{x}) = \{Y \in \mathbb{S}^n \mid f'(\bar{x}) + G'(\bar{x})^* Y = 0\}$$

$$\mathcal{G}_2(\bar{x}) = \{Y \in \mathbb{S}^n \mid Y \in \mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x}))\}$$

New local convergence results of ALM for NLSDP

Theorem (Local convergence of ALM for NLSDP)

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP with a multiplier $\overline{Y} \in \mathcal{M}(\bar{x})$. Suppose the **strongly SOSC** *holds at $(\bar{x}, \overline{Y})$. Let Y^0 be an initial point sufficiently close to $\mathcal{M}(\bar{x})$. Suppose the set $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x, Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ is nonempty and bounded when $Y \in \operatorname{int} \mathcal{Y}$.*

New local convergence results of ALM for NLSDP

Theorem (Local convergence of ALM for NLSDP)

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP with a multiplier $\overline{Y} \in \mathcal{M}(\bar{x})$. Suppose the **strongly SOSC** *holds at $(\bar{x}, \overline{Y})$. Let Y^0 be an initial point sufficiently close to $\mathcal{M}(\bar{x})$. Suppose the set $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x, Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ is nonempty and bounded when $Y \in \operatorname{int} \mathcal{Y}$.*

(i) *Under stopping criterion (a): the sequence $\{Y^k\}$ converge to a particular solution $\widehat{Y} \in \mathcal{M}(\bar{x})$ and $x^k \to \bar{x}$.*

New local convergence results of ALM for NLSDP

Theorem (Local convergence of ALM for NLSDP)

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP with a multiplier $\overline{Y} \in \mathcal{M}(\bar{x})$. Suppose the **strongly SOSC** *holds at $(\bar{x}, \overline{Y})$. Let Y^0 be an initial point sufficiently close to $\mathcal{M}(\bar{x})$. Suppose the set $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x, Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ is nonempty and bounded when $Y \in \operatorname{int} \mathcal{Y}$.*

(i) *Under stopping criterion (a): the sequence $\{Y^k\}$ converge to a particular solution $\widehat{Y} \in \mathcal{M}(\bar{x})$ and $x^k \to \bar{x}$.*

(ii) *Under stopping criterion (b): suppose $\{\mathcal{G}_1(\bar{x}), \mathcal{G}_2(\bar{x})\}$ is also* **boundedly linearly regular**. *Then, we have $\operatorname{dist}(Y^k, \mathcal{M}(\bar{x})) \to 0$ with*

$$\operatorname{dist}(Y^{k+1}, \mathcal{M}(\bar{x})) \leq \frac{1}{\sqrt{1 + b^2(\rho^\infty)^2}} \operatorname{dist}(Y^k, \mathcal{M}(\bar{x}))$$

and $\bar{x}^k \to \bar{x}$ with

$$\|\bar{x}^k - \bar{x}\| \leq \frac{1}{s} \operatorname{dist}(Y^k, \mathcal{M}(\bar{x})),$$

where $\bar{x}^k$ **is the exact solution of ALM subproblems**.

New local convergence results of ALM for NLSDP (con't)

(iii) Under stopping criterion (c): the sequence $\{Y^k\}$ converge to a particular solution $\widehat{Y} \in \mathcal{M}(\bar{x})$ with

$$\|Y^{k+1} - \widehat{Y}\| \leq \frac{1}{\sqrt{1 + b^2(\rho^\infty)^2}} \|Y^k - \widehat{Y}\|.$$

Moreover, if the stopping criterion is further supplemented by

$$\|(\mathcal{L}_{\rho^k})'_x(x^{k+1}, Y^k)\| \leq c\|Y^{k+1} - Y^k\| \text{ for some fixed } c,$$

we have $x^k \to \bar{x}$ with

$$\|x^k - \bar{x}\| \leq p\|Y^k - \widehat{Y}\|$$

for some $p > 0$.

New local convergence results of ALM for NLSDP (con't)

Comments:

- The bounded linear regularity of the collection $\{\mathcal{G}_1(\bar{x}), \mathcal{G}_2(\bar{x})\}$ holds under one of the following two conditions:

  (i) $\mathcal{G}_2(\bar{x})$ is a polyhedron, i.e., $|\gamma| \geq n-1$, where γ is the negative eigenvalue index set of matrix $G(\bar{x}) + \overline{Y}$;

  (ii) there exists a strict complementarity KKT pair $(\bar{x}, \widetilde{Y})$ ($\widetilde{Y}$ does not have to be $\overline{Y}$), i.e., $\mathrm{rank}(G(\bar{x})) + \mathrm{rank}(\widetilde{Y}) = n$.

- $\bar{x}^k$ is usually unknown.

R-linear convergence of the KKT residual

Inspired by Cui, Sun and Toh **MP** 19', we obtain the following R-linear convergence of the KKT residual.

R-linear convergence of the KKT residual

Inspired by Cui, Sun and Toh **MP** 19', we obtain the following R-linear convergence of the KKT residual.

Theorem

Define the following residual function

$$R(x, Y) := \|L'_x(x, Y)\| + \|G(x) - \Pi_{\mathbb{S}^n_+}(G(x) + Y)\|.$$

Then, under stopping criterion (b), for k sufficiently large, if $\sqrt{\rho_0^k}\epsilon_k < 1$, we have there exists $a > 0$ such that

$$R(x^{k+1}, Y^{k+1}) \leq c^k \mathrm{dist}(Y^k, \mathcal{M}(\bar{x}))$$

with $c^k = (\epsilon'_k\sqrt{\rho_0^k} + (1 + \bar{\rho} + a\bar{\rho})(\rho_0^k)^{-1})(1 - \sqrt{\rho_0^k}\epsilon_k)^{-1}$.

Summary

- For **non-convex** and **non-polyhedral** problems (**NLSDP**, **NLSOC**)

	optimality	CQs
Sun, Sun and Zhang **MP** 08'	Strong SOSC	LICQ
Kanzow and Steck **MP** 19'	SOSC	SRCQ
Wang and **D.** 21'	SOSC	partially free+
This work	Strong SOSC	free+

Characterization of strong variational sufficiency II

Solving ALM subproblem: semi-smooth Newton CG method

ALM subproblem:

$$\min_{x \in \mathcal{X}} \ \mathcal{L}_\rho(x, Y)$$

with

$$\mathcal{L}_\rho(x, Y) := f(x) + \frac{\rho}{2} \|\Pi_{\mathbb{S}_-}(G(x) + \frac{Y}{\rho})\|^2 - \frac{\|Y\|^2}{2\rho}$$

Solving ALM subproblem: semi-smooth Newton CG method

ALM subproblem:

$$\min_{x \in \mathcal{X}} \ \mathcal{L}_\rho(x, Y)$$

with

$$\mathcal{L}_\rho(x, Y) := f(x) + \frac{\rho}{2} \|\Pi_{\mathbb{S}_-}(G(x) + \frac{Y}{\rho})\|^2 - \frac{\|Y\|^2}{2\rho}$$

$\mathcal{L}_\rho(x, Y)$ is continuously differentiable with

$$(\mathcal{L}_\rho)'_x(x, Y) = f'(x) + \rho G'(x)^* \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1} Y)$$

Solving ALM subproblem: semi-smooth Newton CG method

ALM subproblem:

$$\min_{x \in \mathcal{X}} \ \mathcal{L}_\rho(x, Y)$$

with

$$\mathcal{L}_\rho(x, Y) := f(x) + \frac{\rho}{2} \|\Pi_{\mathbb{S}_-}(G(x) + \frac{Y}{\rho})\|^2 - \frac{\|Y\|^2}{2\rho}$$

$\mathcal{L}_\rho(x, Y)$ is continuously differentiable with

$$(\mathcal{L}_\rho)'_x(x, Y) = f'(x) + \rho G'(x)^* \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1} Y)$$

B(ouligand)-subdifferential of $(\mathcal{L}_\rho)'_x$:

$$\partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y) := \{\lim_{k \to \infty} (\mathcal{L}_\rho)''_{xx}(x^k, Y^k) \mid (x^k, Y^k) \in \mathcal{U},\ (x^k, Y^k) \to (x, Y)\},$$

where $\mathcal{U}$ is the set of F-differentiable points of $(\mathcal{L}_\rho)'_x$

Semi-smooth Newton CG method

- Sun, Sun, Zhang **MP** 08':

$$\pi_x \partial_B \big((\mathcal{L}_\rho)'_x \big)(x, Y)(\Delta x)$$
$$= L''_{xx}(x, \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1} Y))(\Delta x) + \rho G'(x)^* \partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1} Y) G'(x)(\Delta x)$$

with

  - $\pi_x(\cdot)$: the projection onto the space $\mathbb{X}$
  - $\partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1} Y)$: the B-subdifferential of $\Pi_{\mathbb{S}^n_-}(\cdot)$ at $G(x) + \rho^{-1} Y$

Semi-smooth Newton CG method

- Sun, Sun, Zhang **MP** 08':

$$\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)(\Delta x)$$
$$= L''_{xx}(x, \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y))(\Delta x) + \rho G'(x)^* \partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y) G'(x)(\Delta x)$$

with

  - $\pi_x(\cdot)$: the projection onto the space $\mathbb{X}$
  - $\partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)$: the B-subdifferential of $\Pi_{\mathbb{S}^n_-}(\cdot)$ at $G(x) + \rho^{-1}Y$

- For any $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)$, define the following matrix

$$\mathcal{A}_\rho(Y, W) := L''_{xx}(x, \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)) + \rho G'(x)^* W G'(x) \in \pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$$

Semi-smooth Newton CG method

- Sun, Sun, Zhang **MP** 08':

$$\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x,Y)(\Delta x)$$
$$= L''_{xx}(x, \Pi_{\mathbb{S}^n_-}(G(x)+\rho^{-1}Y))(\Delta x) + \rho G'(x)^* \partial_B \Pi_{\mathbb{S}^n_-}(G(x)+\rho^{-1}Y) G'(x)(\Delta x)$$

with

  - $\pi_x(\cdot)$: the projection onto the space $\mathbb{X}$
  - $\partial_B \Pi_{\mathbb{S}^n_-}(G(x)+\rho^{-1}Y)$: the B-subdifferential of $\Pi_{\mathbb{S}^n_-}(\cdot)$ at $G(x)+\rho^{-1}Y$

- For any $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(x)+\rho^{-1}Y)$, define the following matrix

$$\mathcal{A}_\rho(Y,W) := L''_{xx}(x, \Pi_{\mathbb{S}^n_-}(G(x)+\rho^{-1}Y)) + \rho G'(x)^* W G'(x) \in \pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x,Y)$$

Newton equation:

$$\mathcal{A}_\rho(Y,W)d = -(\mathcal{L}_\rho)'_x(x,Y)$$

The non-singularity of $\mathcal{A}_\rho(Y, W)$

Rockefeller **MP** 22':

- $\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$: the (partial) **Hessian bundle** for $\mathcal{L}_\rho$

The non-singularity of $\mathcal{A}_\rho(Y, W)$

Rockefeller **MP** 22':

- $\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$: the (partial) **Hessian bundle** for $\mathcal{L}_\rho$
- **Hessian bundle criterion**:

$$\textbf{Strong variational sufficiency} \quad \Longleftrightarrow \quad \pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y) \succ 0$$

The non-singularity of $\mathcal{A}_\rho(Y, W)$

Rockefeller **MP** 22':

- $\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$: the (partial) **Hessian bundle** for $\mathcal{L}_\rho$
- **Hessian bundle criterion**:

$$\textbf{Strong variational sufficiency} \quad \Longleftrightarrow \quad \pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y) \succ 0$$

Theorem (Characterization of strong variational sufficiency for NLSDP)

Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP and $\overline{Y}$ be a corresponding multiple. Then the following conditions are equivalent.

(i) *The* **strong variational sufficient condition** *with respect to $(\bar{x}, \overline{Y})$ holds.*

(ii) *The* **strong second order sufficient condition** *holds at $(\bar{x}, \overline{Y})$.*

(iii) *there exist $\rho_0 > 0$ and $\underline{\eta} > 0$ such that for any $\rho \geq \rho_0$ and any $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(\bar{x}) + \rho^{-1}\overline{Y})$,*

$$\langle d, \mathcal{A}_\rho(\overline{Y}, W) d \rangle \geq \underline{\eta} \|d\|^2 \quad \forall d \in \mathbb{X},$$

The non-singularity of $\mathcal{A}_\rho(Y, W)$ (con't)

Proof sketch:

(i)$\Longrightarrow$(ii): Done.

Proposition

Let $\bar{x} \in \mathbb{X}$ be a stationary point to NLSDP and $\overline{Y} \in \mathcal{M}(\bar{x})$. Then, there exists $q \in \operatorname{quad} \delta_{\mathbb{S}^n_+}(G(\bar{x}) \mid \overline{Y})$ such that for all $H \in \mathbb{S}^n$,

$$\begin{aligned} q(H) &= -\frac{1}{2}\Upsilon_{G(\bar{x})}\big(\overline{Y}, H\big) + \delta_{\operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\overline{Y})}(H) \\ &= \sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(A)}(\widetilde{H}_{ij})^2 + \delta_{\operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\overline{Y})}(H), \end{aligned}$$

where $\widetilde{H} = P^T H P$.

(ii)$\Longrightarrow$(iii): Similar to Sun, Sun and Zhang **MP** 08', Proposition 4

by applying variational properties of SDP cone

(iii)$\Longrightarrow$(i): Directly by Rockafellar **MP** 22'

Augmented tilt stability

Definition

In context of the local (strong) convexity-concavity of $\mathcal{L}_\rho(x, Y)$, the **augmented tilt stability** will be said to hold if there is a neighborhood $\mathcal{V}$ of 0 such that the mapping

$$(v, Y) \mapsto \arg\min_{x \in \mathcal{X}} \{\mathcal{L}_\rho(x, Y) - \langle v, x \rangle\} \quad \text{for} \quad (v, y) \in \mathcal{V} \times \mathcal{Y}$$

is single-valued and Lipschitz continuous.

Sun, Sun and Zhang **MP** 08':

SSOSC+LICQ $\Rightarrow$ positive definiteness of $\mathcal{A}_\rho(\overline{Y}, W)$ + Assumption B2 (directional derivative estimation of $x(Y)$)

$\Rightarrow$ **augmented tilt stability** with $v = 0$ for NLSDP

Rockafellar **MP** 22':

strong variational sufficiency $\Rightarrow$ **augmented tilt stability** for general OP

Numerical experiments

A toy example

Consider the following **non-convex** NLSDP:

$$\begin{array}{ll} \min & \frac{1}{2}x^3 \\ \text{s.t} & -x^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{S}^3_+ \quad \Leftarrow \quad \Gamma \end{array}$$

A toy example

Consider the following **non-convex** NLSDP:

$$\begin{array}{ll} \min & \frac{1}{2}x^3 \\ \text{s.t} & -x^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{S}^3_+ \quad \Leftarrow \quad \Gamma \end{array}$$

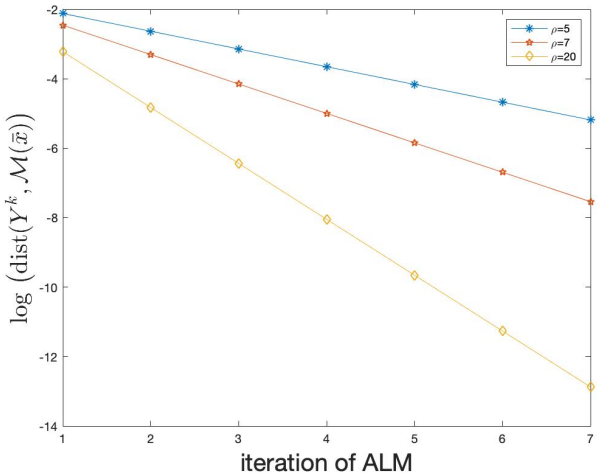
- $\bar{x} = 0$ is the optimal solution
- The corresponding multiplier set $\mathcal{M}(\bar{x}) := \{\Gamma \mid \Gamma \in \mathcal{S}^3_-\}$
- Robinson's CQ (MFCQ) does not hold
- Let $\overline{\Gamma} = \mathrm{Diag}(0, -1, -2)$. It is easy to verify $L''_{xx}(\bar{x}, \overline{\Gamma}) = 4 > 0$, which implies strongly second order sufficient condition holds at $(\bar{x}, \overline{\Gamma})$.

A toy example

Consider the following **non-convex** NLSDP:

$$\begin{array}{ll} \min & \frac{1}{2}x^3 \\ \text{s.t} & -x^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{S}^3_+ \quad \Leftarrow \quad \Gamma \end{array}$$

- $\bar{x} = 0$ is the optimal solution
- The corresponding multiplier set $\mathcal{M}(\bar{x}) := \{\Gamma \mid \Gamma \in \mathcal{S}^3_-\}$
- Robinson's CQ (MFCQ) does not hold
- Let $\overline{\Gamma} = \mathrm{Diag}(0, -1, -2)$. It is easy to verify $L''_{xx}(\bar{x}, \overline{\Gamma}) = 4 > 0$, which implies strongly second order sufficient condition holds at $(\bar{x}, \overline{\Gamma})$.



A slightly complicated example

$$\begin{array}{ll} \min\limits_{X \in \mathbb{S}^n} & \frac{1}{2}\langle X, Q \circ X \rangle \\ \text{s.t.} & B \circ X = 0 \\ & X \in \mathbb{S}^n_+, \end{array}$$

where "$\circ$" denotes the Hadamard product of matrices,

$$B = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} q & \cdots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & q & 0 \\ 0 & \cdots & 0 & -1 \end{bmatrix}$$

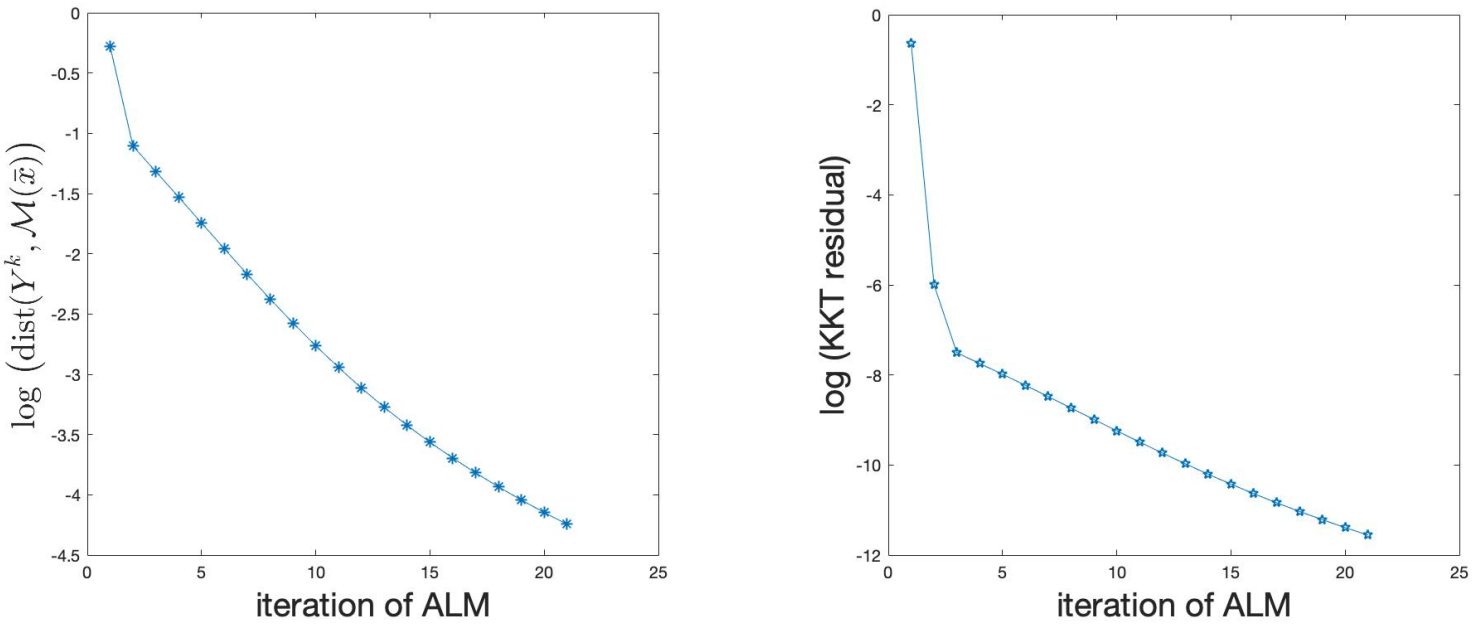
with a given $q > n - 1$.

- $\overline{X} = 0$ is a local optimal solution
- the corresponding multiplier set is given by

$$\mathcal{M}(\overline{X}) = \{(Y, Z) \in \mathbb{S}^n \times \mathbb{S}^n \mid Y + B \circ Z = 0, Y \in \mathbb{S}^n_-\}.$$

- Robinson's CQ does not hold
- strongly SOSC holds and $\exists\, \widehat{Y} \in \mathcal{M}(\bar{x})$ s.t. $(\bar{x}, \widehat{Y})$ is a strict complementarity KKT pair

A slightly complicated example (con't)



(a) (b)

Figure 2: Semismooth Newton-CG based ALM for problem with $n = 1000$ and $q = 1500$

The non-singularity of Hessian V: e.g.,

- $\lambda_{\min}(V) \approx 1.01$ if $n = 3$ and $q = 2$
- $\lambda_{\min}(V) \approx 1.26$ if $n = 100$ and $q = 200$

A slightly complicated example (con't)

n	q	iteration	KKT residual	cpu(s)
3	2	8	8.27e-06	0.22
100	200	11	8.98e-06	3.11
1000	1500	21	9.57e-06	1083.34

Table 1: Numerical results of semismooth Newton-CG based ALM

Conclusions

In this talk:

- The explicit characterization of strong variational sufficiency without any CQ
- Convergence rate of ALM for **NLSDP**
- Simple numerical experiments are conducted to verify our theory

Similar results can be obtained for the general matrix optimization.

Conclusions

In this talk:

- The explicit characterization of strong variational sufficiency without any CQ
- Convergence rate of ALM for **NLSDP**
- Simple numerical experiments are conducted to verify our theory

Similar results can be obtained for the general matrix optimization.

ALM for non-smooth optimization on Riemannian manifolds :

- Y.H. Zhou, C.L. Bao, **D.**, and J. Zhu, A semismooth Newton based augmented Lagrangian method for nonsmooth optimization on matrix manifolds, **MP**, 1-61 (2022).
- Y.X. Zhou, C.L. Bao and **D.**, On the robust isolated calmness of a class of nonsmooth optimizations on Riemannian manifolds and its applications, 2022, arXiv: 2208.07518.

Thank you!