Nonsmooth optimization over Riemannian manifolds

Manifold strong variational sufficiency and Riemannian ALM

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

Workshop on Optimization, Equilibrium and Complementarity
The Hong Kong Polytechnic University
2023.08.18

勤数
笃系
求天
真地
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

Acknowledgements

Based on the joint work with

- **Yuexin Zhou** @ *Institute of Applied Mathematics, AMSS*
- **Yangjing Zhang** @ *Institute of Applied Mathematics, AMSS*

♢ Y.X. Zhou, **D.** and Y.J. Zhang, Strong variational sufficiency of nonsmooth optimization problems on Riemannian manifolds, arXiv:2308.06793, 2023.

Outline

Nonsmooth optimization problems over manifolds

Riemannian Augmented Lagrangian method (RALM)

Strong variational sufficiency for NManOPs

Local convergence rate of RALM and its subproblem

Numerical experiments

Nonsmooth optimization problems over manifolds

NManOP:

$$\begin{aligned} &\min\ f(x) + \theta(g(x)) \\ &\text{s.t. } x \in \mathcal{M}, \end{aligned}$$

- $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$: the finite dimensional Euclidean spaces
- $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{X} \to \mathbb{Y}$: smooth functions
- $\mathcal{M} \subseteq \mathbb{X}$: a **Riemannian manifold**
- $\theta : \mathbb{Y} \to (-\infty, \infty]$: a closed convex function, e.g., $\|\cdot\|_1$; $\|\cdot\|_{(k)}$; $\delta_{\mathbb{R}^n_+}(\cdot)$...

Nonsmooth optimization problems over manifolds (con't)

NManOP:

$$\begin{aligned} \min\ & f(x) + \theta(g(x)) \\ \text{s.t. } & x \in \mathcal{M}, \end{aligned}$$

NManOP:

$$\min \; f(x) + \theta(g(x))$$
$$\text{s.t. } x \in \mathcal{M},$$

Consider the following "**simple**" nonsmooth problem:

- $\theta \equiv \delta_{\mathbb{R}_+^{n\times n}}(\cdot)$
- $\mathcal{M} \equiv \{X \in \mathbb{R}^{n\times n} \mid X^T X = I_n\}$, i.e., the set of all $n \times n$ orthogonal matrices.
- $f \equiv \langle X, AXB + C\rangle$, where A, B and C are given $n \times n$ real symmetric matrices.

NManOP:

$$\begin{aligned} &\min\ f(x) + \theta(g(x)) \\ &\text{s.t. } x \in \mathcal{M}, \end{aligned}$$

Consider the following "**simple**" nonsmooth problem:

- $\theta \equiv \delta_{\mathbb{R}_+^{n\times n}}(\cdot)$
- $\mathcal{M} \equiv \{X \in \mathbb{R}^{n\times n} \mid X^T X = I_n\}$, i.e., the set of all $n \times n$ orthogonal matrices.
- $f \equiv \langle X, AXB + C\rangle$, where A, B and C are given $n \times n$ real symmetric matrices.

The **NManOP** is exactly the quadratic assignment problem (QAP)

$$\mathrm{OPT}_{\mathrm{QAP}} = \min\left\{\langle X, AXB + C\rangle \ \mid \ X^T X = I,\ X \geq 0\right\}.$$

NManOP:

$$\begin{aligned} &\min\ f(x) + \theta(g(x)) \\ &\text{s.t. } x \in \mathcal{M}, \end{aligned}$$

Consider the following "**simple**" nonsmooth problem:

- $\theta \equiv \delta_{\mathbb{R}^{n\times n}_+}(\cdot)$
- $\mathcal{M} \equiv \{X \in \mathbb{R}^{n\times n} \mid X^T X = I_n\}$, i.e., the set of all $n \times n$ orthogonal matrices.
- $f \equiv \langle X, AXB + C\rangle$, where A, B and C are given $n \times n$ real symmetric matrices.

The **NManOP** is exactly the quadratic assignment problem (QAP)

$$\mathrm{OPT}_{\mathrm{QAP}} = \min\left\{\langle X, AXB + C\rangle \ \mid\ X^T X = I,\ X \geq 0\right\}.$$

However, the nonsmooth term "θ" makes the problem very difficult to solve. In fact, the Riemannian **Robinson CQ** does not hold at any feasible point.

Other applications

Compressed modes (CM) problem:

$$\min_{X \in \mathrm{St}(n,r)} \ \mathrm{trace}(X^T H X) + \mu \|X\|_1$$

Sparse principal component analysis (SPCA):

$$\min_{X \in \mathrm{St}(n,r)} \ -\mathrm{trace}(X^T A^T A X) + \mu \|X\|_1$$

Constrained SPCA:

$$\begin{aligned} \min_{X \in \mathrm{St}(n,r)} \quad & -\mathrm{trace}(X^T A^T A X) + \mu \|X\|_1 \\ \text{s.t.} \quad & |X_i A^T A X_j| \le \Delta_{ij} \quad \forall\, i \ne j \end{aligned}$$

And many others: l_1-PCA; orthogonal dictionary learning; robust subspace recovery; ONMF; ...

Matrix manifolds

Ambient space $\mathbb{R}^{m\times n}$:

- **Embedded manifolds**: **(orthogonal/compact) Stiefel manifold**; **fixed rank manifold**;

- **Quotient manifolds**: Grassmann manifold

Existing methods

- **Subgradient methods**: Ferreira and Oliveria, (1998); Dirr, et al., (2006); Borckmans, et al., (2014); Grohs and Hosseini, (2016); Hosseini, (2017); Hosseini, et al. (2018); ...
- **ADMs/ADMMs on manifold**: **SOC**: Lai and Osher, (2014); **MADMM**: Kovnatsky, et al., (2016); **EPALMAL**: Zhu, et al. (2017); **PAMAL**: Chen, et al., (2016)
- **Proximal gradient method**: **ManPG**: Chen, et al, (2020); **AManPG**: Huang and Wei, (2019); **ARPG**: Huang and Wei, (2020)
- **Penalty approach**: **PenCPG**: Xiao, et al. (2020); **SLPG**: Xiao, et al. (2021)

Riemannian Augmented Lagrangian method (RALM)

Augmented Lagrangian method[1]

Consider

$$\begin{array}{lll} \min\limits_{x \in \mathbb{X}} & \Phi(x) & \\ \text{s.t.} & h(x) = 0 & \leftarrow \quad y \end{array}$$

[1] a.k.a. the method of multipliers

Augmented Lagrangian method[1]

Consider

$$\begin{array}{lll} \min_{x \in \mathbb{X}} & \Phi(x) & \\ \text{s.t.} & h(x) = 0 & \leftarrow \quad y \end{array}$$

ALM (Hestenes, 69'; Powell, 69'):

$$\begin{cases} x^{k+1} \approx \operatorname{argmin} \left\{ L_\rho(x; y^k) \right\} \\ y^{k+1} = y^k + \rho h(x^{k+1}) \end{cases}$$

[1] a.k.a. the method of multipliers

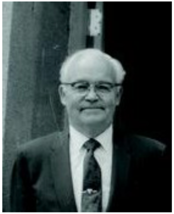
Augmented Lagrangian method[1]

Consider

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & \Phi(x) \\ \text{s.t.} \quad & h(x) = 0 \quad \leftarrow \quad y \end{aligned}$$

ALM (**Hestenes, 69'**; **Powell, 69'**):

$$\begin{cases} x^{k+1} \approx \operatorname{argmin} \left\{ L_\rho(x; y^k) \right\} \\ y^{k+1} = y^k + \rho h(x^{k+1}) \end{cases}$$



Magnus Rudolph Hestenes
(February 13 1906 – May 31 1991)

Michael James David Powell
(29 July 1936 – 19 April 2015)

[1] a.k.a. the method of multipliers

Riemannian ALM (RALM) for NManOPs

By adding a perturbation parameter u, consider the **perturbed NManOP**:

$$\begin{array}{ll} \min & \varphi(x,u) := f(x) + \theta(g(x)+u) \\ \text{s.t.} & x \in \mathcal{M} \end{array}$$

- **Lagrangian function**:

$$l(x,y) = \inf_u \{\varphi(x,u) - \langle y,u\rangle\} = L(x,y) - \theta^*(y),$$

  where $L(x,y) = f(x) + \langle y, g(x)\rangle$ and θ^* is the conjugate function

- **Augmented Lagrangian function**:

$$l^\rho(x,y) = \inf_u \left\{\varphi(x,u) - \langle y,u\rangle + \frac{\rho}{2}\|u\|^2\right\}$$

- **The inexact RALM iteration**:

$$\begin{cases} x^{k+1} \approx \underset{x\in\mathcal{U}}{\operatorname{argmin}}\, l^{\rho_k}\left(x, y^k\right), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), \end{cases}$$

  where $\rho_k,\ \tilde{\rho}_k > 0$ and $\mathcal{U}$ is a subset of $\mathcal{M}$.

Riemannian ALM (RALM) for NManOPs

By adding a perturbation parameter u, consider the **perturbed NManOP**:

$$\begin{array}{ll} \min & \varphi(x,u) := f(x) + \theta(g(x)+u) \\ \text{s.t.} & x \in \mathcal{M} \end{array}$$

- **Lagrangian function**:

$$l(x,y) = \inf_u \{\varphi(x,u) - \langle y, u\rangle\} = L(x,y) - \theta^*(y),$$

  where $L(x,y) = f(x) + \langle y, g(x)\rangle$ and θ^* is the conjugate function

- **Augmented Lagrangian function**:

$$l^\rho(x,y) = \inf_u \left\{\varphi(x,u) - \langle y,u\rangle + \frac{\rho}{2}\|u\|^2\right\}$$

- **The inexact RALM iteration**:

$$\begin{cases} x^{k+1} \approx \underset{x\in\mathcal{U}}{\operatorname{argmin}}\ l^{\rho_k}\left(x, y^k\right), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), \end{cases}$$

  where $\rho_k,\ \tilde{\rho}_k > 0$ and $\mathcal{U}$ is a subset of $\mathcal{M}$.

Zhou, Bao, **D.** and Zhu, **MP** (2023): a semismooth Newton based RALM.

$$\begin{array}{ll} \min & x_2^2 + |x_1 - x_2| \\ \text{s.t.} & 2x_1 + x_2 \geq 0 \\ & x_1^2 + x_2^2 = 1 \end{array}$$

The unique optimal solution is $(x_1^*, x_2^*) = (\sqrt{2}/2, \sqrt{2}/2)$ with the corresponding multipliers $y^* = \sqrt{2}/2$ and $z^* = 0$.

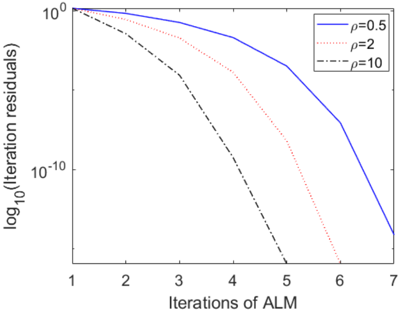


Figure 1: the residuals generated by exact ALM with different ρ

Consider the CM problem to solve the Schrödinger equation of 1D free-electron model with periodic boundary condition

$$\min_{X \in \mathrm{St}(n,r)} \operatorname{tr}(X^T H X) + \mu \|X\|_1$$

where H is the discretization of the Hamilton operator.

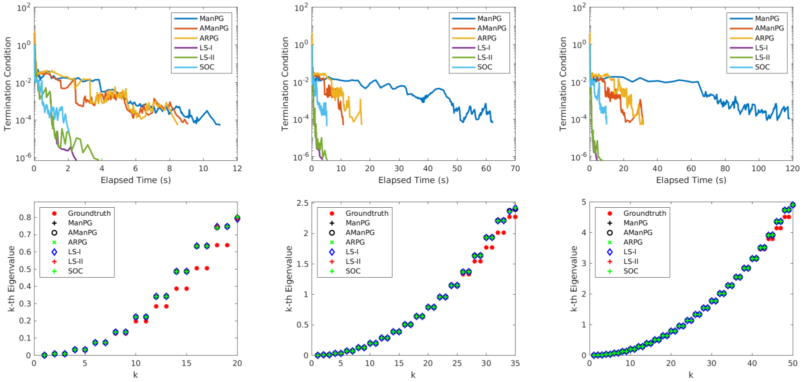


Figure 2: Zhou, Bao, **D.** and Zhu, **MP** (2023)

Local convergence analysis of ALM: the Euclidean case

For the polyhedral case:

- **NLP with equality constraints**: cf. Powell, 69'
- **Convex OPs**: Rockafellar, 76'
- **NLP (non-convex)**: Bertsekas, 82'; Conn, Gould and Toint, 91'; Contesse-Becker 93'; Ito & Kunisch 91'; Fernández and Solodov, 12'; Nocedal and Wright, 06', ...

Local convergence analysis of ALM: the Euclidean case

For the polyhedral case:

- **NLP with equality constraints**: cf. Powell, 69'
- **Convex OPs**: Rockafellar, 76'
- **NLP (non-convex)**: Bertsekas, 82'; Conn, Gould and Toint, 91'; Contesse-Becker 93'; Ito & Kunisch 91'; Fernández and Solodov, 12'; Nocedal and Wright, 06', ...

For non-polyhedral & non-convex:

- **NLSDP** (Sun, Sun and Zhang, **MP** 08'):
  strong SOSC $+$ **LICQ** $\Longrightarrow$ primal-dual linear
- C^2**-cone reducible conic problems** (Kanzow and Steck, **MP** 18'):
  SOSC $+$ **SRCQ** $\Longrightarrow$ primal-dual linear

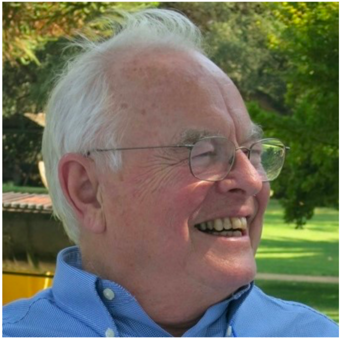
For the polyhedral case:

- **NLP with equality constraints**: cf. Powell, 69'
- **Convex OPs**: Rockafellar, 76'
- **NLP (non-convex)**: Bertsekas, 82'; Conn, Gould and Toint, 91'; Contesse-Becker 93'; Ito & Kunisch 91'; Fernández and Solodov, 12'; Nocedal and Wright, 06', ...

For non-polyhedral & non-convex:

- **NLSDP** (Sun, Sun and Zhang, **MP** 08'):
  strong SOSC + **LICQ** $\Longrightarrow$ primal-dual linear
- C^2**-cone reducible conic problems** (Kanzow and Steck, **MP** 18'):
  SOSC + **SRCQ** $\Longrightarrow$ primal-dual linear
- Both under the **uniqueness** of the KKT solution.

Rockafellar's recent work: (strong) variational sufficiency



Rockafellar **MP** 22' shows that under so-called **strong variational sufficiency**, ALM has **the Q-linear convergence of multiplier** and **R-linear of the primal variable** even for non-convex problems.

Convexity and monotonicity

A proper, lsc function

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

A proper, lsc function

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

A proper, lsc function

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

- **Duality theorem**

Convexity and monotonicity

A proper, lsc function

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

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

A proper, lsc function

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

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

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

Beyond convexity : variational convexity

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 : variational convexity

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 $h(x) = f(x)$.

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**

(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 $\operatorname{gph} \partial \phi_{\bar{\rho}}$.

(Strong) variational sufficiency and weak convexity

Recall the general composite optimization problem:

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

(Strong) variational sufficiency and weak convexity

Recall the general composite optimization problem:

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

When $G(x) = \mathcal{A}x + b$, we are able to show **locally**

(strong) variational sufficiency $\quad \Longrightarrow \quad$ ψ is **weakly convex**

i.e., there exists $m > 0$ such that $\psi(x) + \frac{m}{2}\|x\|^2$ is convex.

(Strong) variational sufficiency and weak convexity

Recall the general composite optimization problem:

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

When $G(x) = \mathcal{A}x + b$, we are able to show **locally**

$$\text{(strong) variational sufficiency} \quad \Longrightarrow \quad \psi \text{ is } \textbf{weakly convex}$$

i.e., there exists $m > 0$ such that $\psi(x) + \frac{m}{2}\|x\|^2$ is convex.

However,

$$\text{(strong) variational sufficiency} \quad \not\Longleftarrow \quad \psi \text{ is } \textbf{weakly convex}$$

Consider the following problem:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}} & |x| - x^2 \\ \text{s.t.} & -0.5 \le x \le 0.5, \end{array}$$

since the corresponding augmented Lagrangian function is not convex.

(Strong) variational sufficiency: for non-polyhedral case

For **NLSDP**,

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

(Strong) variational sufficiency: for non-polyhedral case

For **NLSDP**,

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

Recently, Wang, **D.**, Zhang and Zhao, **SIOPT** 23' shows that

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 $(\bar{x}, \overline{Y})$.*

(Strong) variational sufficiency: for non-polyhedral case

For **NLSDP**,

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

Recently, Wang, **D.**, Zhang and Zhao, **SIOPT** 23' shows that

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 $(\bar{x}, \overline{Y})$.*

Under the Euclidean setting, the local convergence rate of ALM even for **non-convex** and **non-polyhedral** problems, e.g., NLSDP and NLSOC:

	optimality	**CQs**
Sun, Sun and Zhang **MP** 07'	Strong SOSC	LICQ
Kanzow and Steck **MP** 19'	SOSC	SRCQ
Wang and **D. COAP** 23'	SOSC	partially free+
Wang, **D.**, Zhang and Zhao **SIOPT** 23'	Strong SOSC	free+

Strong variational sufficiency for NManOPs

Geodesic convexity?

Geodesically convex:

for each geodesic $\gamma : \mathbb{R} \to \mathcal{M}$, $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is convex function: i.e., for any $\lambda \in [0,1]$ and $a, b \in \mathbb{R}$,

$$f \circ \gamma((1-\lambda)a + \lambda b) \leq (1-\lambda) f \circ \gamma(a) + \lambda f \circ \gamma(b).$$

Geodesic convexity?

Geodesically convex:

for each geodesic $\gamma : \mathbb{R} \to \mathcal{M}$, $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is convex function: i.e., for any $\lambda \in [0, 1]$ and $a, b \in \mathbb{R}$,

$$f \circ \gamma((1-\lambda)a + \lambda b) \leq (1-\lambda) f \circ \gamma(a) + \lambda f \circ \gamma(b).$$

- $f : \mathcal{M} \to \mathbb{R}$ is **geodesically convex** over a compact manifold, then f is **constant**

Geodesic convexity?

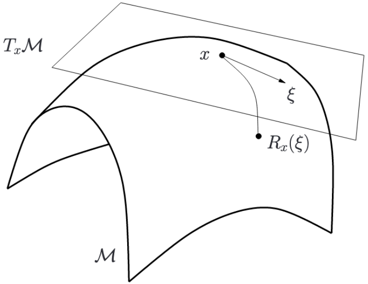
Geodesically convex:

for each geodesic $\gamma : \mathbb{R} \to \mathcal{M}$, $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is convex function: i.e., for any $\lambda \in [0, 1]$ and $a, b \in \mathbb{R}$,

$$f \circ \gamma((1-\lambda)a + \lambda b) \leq (1-\lambda) f \circ \gamma(a) + \lambda f \circ \gamma(b).$$

- $f : \mathcal{M} \to \mathbb{R}$ is **geodesically convex** over a compact manifold, then f is **constant**
- $f(x) = \|x\|_1$ is locally **geodesically concave** around the north pole $N = (0, \cdots, 0, 1)$ of n-sphere $S^n := \{x \in \mathbb{R}^{n+1} : \|x\|_2 = 1\}$

Retraction: a smooth mapping R from the tangent bundle $T\mathcal{M}$ onto $\mathcal{M}$ satisfying $R_x(0_x) = x$ and $DR_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$.



The localization of NManOPs

- At a given point x, by the **inverse function theorem**, we know that any retraction $R_x : T_x\mathcal{M} \to \mathcal{M}$ is a **diffeomorphism** within a neighborhood of 0_x in the tangent space $T_x\mathcal{M}$ for a general Riemannian manifold.

- At a given point x, by the **inverse function theorem**, we know that any retraction $R_x : T_x\mathcal{M} \to \mathcal{M}$ is a **diffeomorphism** within a neighborhood of 0_x in the tangent space $T_x\mathcal{M}$ for a general Riemannian manifold.
- For a given function $F : \mathcal{M} \to \mathbb{R}$, we define $F_{R_x} : T_x\mathcal{M} \to \overline{\mathbb{R}}$ by

$$F_{R_x}(\xi) = \begin{cases} F(R_x\xi), & \xi \in B_x(r_{R_x}), \\ +\infty, & \xi \notin B_x(r_{R_x}), \end{cases}$$

  where r_{R_x} is called the **injectivity radius** of a Riemannian manifold $\mathcal{M}$ at a point x with respect to retraction R_x.

Thus, NManOPs can be locally transformed into the following **equivalent** problem on the tangent space $T_x\mathcal{M}$, i.e.,

$$\begin{array}{ll} \min & f_{R_x}(\xi) + \theta(g_{R_x}(\xi)) \\ \text{s.t.} & \xi \in T_x\mathcal{M}. \end{array}$$

The **perturbed problem** for the localized NManOP:

$$\begin{array}{ll} \min & \varphi_{R_x}(\xi, u) := f_{R_x}(\xi) + \theta(g_{R_x}(\xi) + u) \\ \text{s.t.} & \xi \in T_x\mathcal{M}, \end{array}$$

The **Lagrangian function**:

$$l_{R_x}(\xi, y) = \inf_u \{\varphi_{R_x}(\xi, u) - \langle y, u \rangle\} = L_{R_x}(\xi, y) - \theta^*(y)$$

The **augmented Lagrangian function**:

$$l^\rho_{R_x}(\xi, y) = \inf_u \left\{\varphi_{R_x}(\xi, u) - \langle y, u \rangle + \frac{\rho}{2}\|u\|^2\right\}.$$

Moreover, the **(augmented) objective functions** satisfy

$$\varphi(x, u) = \sup_y \{l(x, y) + \langle y, u \rangle\}, \quad \varphi^\rho(x, u) = \sup_y \{l^\rho(x, y) + \langle y, u \rangle\}$$
$$\varphi_{R_x}(\xi, u) = \sup_y \{l_{R_x}(\xi, y) + \langle y, u \rangle\}, \quad \varphi^\rho_{R_x}(\xi, u) = \sup_y \{l^\rho_{R_x}(\xi, y) + \langle y, u \rangle\}.$$

The localization of NManOPs (con't)

Proposition

The following statements are equivalent:

(i) $(\bar{x},\bar{y})$ *satisfies the first-order optimality condition of the NManOP* $(0,\bar{y}) \in \partial\varphi(\bar{x},0)$;

(ii) $(0_{\bar{x}},\bar{y})$ *satisfies the first-order optimality condition of localized NManOP;*

(iii) *For any* $\rho > 0$, $(\bar{x},\bar{y})$ *satisfies* $(0,\bar{y}) \in \partial\varphi^{\rho}(\bar{x},0)$;

(iv) *For any* $\rho > 0$, $(0_{\bar{x}},\bar{y})$ *satisfies* $(0_{\bar{x}},\bar{y}) \in \partial\varphi^{\rho}_{R_{\bar{x}}}(0_{\bar{x}},0)$;

(v) $\mathrm{grad}_x\, l(\bar{x},\bar{y}) = 0$, $0 \in \partial_y[-l](\bar{x},\bar{y})$, *or* $\mathrm{grad}_x\, L(\bar{x},\bar{y}) = 0$, $\bar{y} \in \partial\theta(g(\bar{x}))$;

(vi) $\nabla_{\xi} l_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0$, $0 \in \partial_y[-l_{R_{\bar{x}}}](0_{\bar{x}},\bar{y})$, *or* $\nabla_{\xi} L_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0$, $\bar{y} \in \partial\theta(g_{R_{\bar{x}}}(0_{\bar{x}}))$;

(vii) $\mathrm{grad}_x\, l^{\rho}(\bar{x},\bar{y}) = 0$, $0 \in \nabla_y l^{\rho}(\bar{x},\bar{y})$, *or* $\mathrm{grad}_x\, L(\bar{x},\bar{y}) = 0$, $\nabla\, \mathrm{env}_{\rho}\, \theta(g(\bar{x}) + \rho^{-1}\bar{y}) = \bar{y}$, *where* $\mathrm{env}_{\rho}\,\theta$ *is the Moreau-Yosida regularization of* θ *defined by* $\mathrm{env}_{\rho}\,\theta(p) := \min_{y\in\mathbb{Y}} \theta(y) + \frac{\rho}{2}\|p-y\|^2$;

(viii) $\nabla_{\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0$, $0 \in \nabla_y l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}},\bar{y})$, *or* $\nabla_{\xi} L_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0$, $\nabla\, \mathrm{env}_{\rho}\, \theta(g_{R_{\bar{x}}}(0_{\bar{x}}) + \rho^{-1}\bar{y}) = \bar{y}$.

(Strong) variational sufficiency for Riemannian optimization

Recall the **NManOP** and **localized NManOP**:

$$\begin{array}{ll} \min & f(x) + \theta(g(x)) \\ \text{s.t.} & x \in \mathcal{M}, \end{array} \quad \text{and} \quad \begin{array}{ll} \min & f_{R_x}(\xi) + \theta(g_{R_x}(\xi)) \\ \text{s.t.} & \xi \in T_x\mathcal{M}. \end{array}$$

For any $\bar{x}$ and $\bar{y}$ satisfying the first-order condition of NManOP and any given retraction $R_{\bar{x}}$, we define the (strong) variational sufficiency for **NManOP** by

Recall the **NManOP** and **localized NManOP**:

$$\begin{array}{ll} \min & f(x) + \theta(g(x)) \\ \text{s.t.} & x \in \mathcal{M}, \end{array} \quad \text{and} \quad \begin{array}{ll} \min & f_{R_x}(\xi) + \theta(g_{R_x}(\xi)) \\ \text{s.t.} & \xi \in T_x\mathcal{M}. \end{array}$$

For any $\bar{x}$ and $\bar{y}$ satisfying the first-order condition of NManOP and any given retraction $R_{\bar{x}}$, we define the (strong) variational sufficiency for **NManOP** by

Manifold (strong) variational sufficiency under $R_{\bar{x}}$

$\Longleftrightarrow$ **(strong) variational sufficiency for the localized NManOP**

(Strong) variational sufficiency for Riemannian optimization

Recall the **NManOP** and **localized NManOP**:

$$\begin{array}{ll} \min & f(x) + \theta(g(x)) \\ \text{s.t.} & x \in \mathcal{M}, \end{array} \quad \text{and} \quad \begin{array}{ll} \min & f_{R_x}(\xi) + \theta(g_{R_x}(\xi)) \\ \text{s.t.} & \xi \in T_x\mathcal{M}. \end{array}$$

For any $\bar{x}$ and $\bar{y}$ satisfying the first-order condition of NManOP and any given retraction $R_{\bar{x}}$, we define the (strong) variational sufficiency for **NManOP** by

Manifold (strong) variational sufficiency under $R_{\bar{x}}$

$\Longleftrightarrow$ **(strong) variational sufficiency for the localized NManOP**

For **NManOP**, we are able to study

- **Local augmented duality**
- **PPA & RALM**

The local duality for manifold optimizations

For a given $\bar{\rho} > 0$, the local augmented primal problem for (NManOP):

$$\min_{x \in R_{\bar{x}}(\mathcal{W})} \sup_{y \in \mathcal{Y}} l^{\bar{\rho}}(x, y) \text{ for } x \in R_{\bar{x}}(\mathcal{W}), \tag{P}$$

the local augmented dual problem for (NManOP):

$$\max_{y \in \mathcal{Y}} \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\bar{\rho}}(x, y) \text{ for } y \in \mathcal{Y}. \tag{D}$$

Theorem

Suppose that $(\bar{x}, \bar{y})$ is a first-order stationary point of NManOP and the manifold variational sufficiency condition holds at $(\bar{x}, \bar{y})$. Then, the problems (P) and (D) defined in the neighborhood $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$ of $(\bar{x}, \bar{y})$ have optimal solutions with $\min(P) = \max(D)$, and

$$x^* \text{ solves } (P) \quad \Longleftrightarrow \quad x^* \text{ minimizes in (NManOP) relative to } R_{\bar{x}}(\mathcal{W}).$$

Moreover the following conditions are equivalent:

(a) x^* *minimizes in (P) and* y^* *maximizes in (D),*

(b) (x^*, y^*) *is a saddle point of* $l^{\bar{\rho}}$ *on* $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$*,*

(c) (x^*, y^*) *is a saddle point of* l^{ρ} *on* $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$ *for any* $\rho \geq \bar{\rho}$*.*

Characterization of strong variational sufficiency for NManOPs (I)

Theorem

Let $\bar{x}$ and $\bar{y}$ satisfy the first-order optimality condition. Suppose that θ is a **polyhedral convex function** *or the indicator function of* **second-order cone** *or* **SDP cone***. Then, the following two conditions are equivalent.*

- *the* **manifold strong variational sufficient condition** *with respect to $(\bar{x}, \bar{y})$ under retraction $R_{\bar{x}}$ holds at $(\bar{x}, \bar{y})$;*
- *the* **manifold strong second-order sufficient condition** *(***M-SSOSC***) holds at $(\bar{x}, \bar{y})$, i.e., for any $Dg(\bar{x})\xi \in \operatorname{aff} \mathcal{C}_{\theta,g}(\bar{x}, \bar{y}) \setminus \{0\}$,*

$$\langle \xi, \operatorname{Hess}_x L(\bar{x}; \bar{y}) \xi \rangle - \sigma\left(\bar{y}, \mathcal{T}^2_{\mathcal{K}}(g(\bar{x}), Dg(\bar{x})\xi)\right) > 0.$$

Characterization of strong variational sufficiency for NManOPs (I)

Theorem

Let $\bar{x}$ and $\bar{y}$ satisfy the first-order optimality condition. Suppose that θ is a **polyhedral convex function** *or the indicator function of* **second-order cone** *or* **SDP cone**. *Then, the following two conditions are equivalent.*

- *the* **manifold strong variational sufficient condition** *with respect to $(\bar{x}, \bar{y})$ under retraction $R_{\bar{x}}$ holds at $(\bar{x}, \bar{y})$;*
- *the* **manifold strong second-order sufficient condition** *(***M-SSOSC***) holds at $(\bar{x}, \bar{y})$, i.e., for any $Dg(\bar{x})\xi \in \operatorname{aff} \mathcal{C}_{\theta,g}(\bar{x}, \bar{y}) \setminus \{0\}$,*

$$\langle \xi, \operatorname{Hess}_x L(\bar{x}; \bar{y}) \xi \rangle - \sigma\left(\bar{y}, \mathcal{T}^2_{\mathcal{K}}(g(\bar{x}), Dg(\bar{x})\xi)\right) > 0.$$

- **M-SSOSC** is **independent** of the choice of the retraction $R_{\bar{x}}$
- Actually, the **manifold strong variational sufficient condition** is **independent** of the retraction $R_{\bar{x}}$;
- It is possible to define the **retractionally (strongly) convex function**. Also, the strong retractional convexity is **independent** of the choice of the retraction $R_{\bar{x}}$ at least near the critical point $\bar{x}$.

Local convergence rate of RALM and its subproblem

Local convergence rate of RALM

Recall the **inexact RALM** iteration takes the form of

$$\begin{cases} x^{k+1} \approx \operatorname*{argmin}_{x \in \mathcal{U}} l^{\rho_k}\left(x, y^k\right), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), \end{cases}$$

where $\rho_k,\ \tilde{\rho}_k > 0$

Follow one of the following rules:

$$\left(2\tilde{\rho}_k\left[l^{\rho_k}\left(x^{k+1}, y^k\right) - \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\rho_k}\left(x, y^k\right)\right]\right)^{1/2} \le \begin{cases} \text{(a)} & \varepsilon_k \\ \text{(b)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|\right\} \\ \text{(c)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|^2\right\} \end{cases}$$

Local convergence rate of RALM (con't)

Theorem

Suppose that the **M-SSOSC** *for NManOP holds at a local optimal solution* $\bar{x}$. *Let* $\{(x^k, y^k)\}$ *be the sequence generated by ALM. Then, under some suitable conditions,* $\operatorname{dist}\left(y^k, \mathcal{M}(\bar{x})\right) \to 0$ *Q-linearly at a rate* $0 < \tau < 1$, *i.e.,*

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

where $\mathcal{M}(\bar{x})$ *is the Lagrange multiple set of* $\bar{x}$. *Moreover,* $x^k \to \bar{x}$ *R-linearly at that rate as long as the stopping criterion in approximate minimization is supplemented by the proviso that*

$$\left\|\operatorname{grad}_x l^{\rho_k}\left(x^{k+1}, y^k\right)\right\| \leq c \left\|y^{k+1} - y^k\right\| \quad \textit{for some fixed } c.$$

Semismooth Newton method for the subproblem

Recall the RALM subproblem:

$$x^{k+1} \approx \underset{x \in \mathcal{U}}{\operatorname{argmin}}\, l^{\rho_k}\left(x, y^k\right)$$

Consider the **embedded submanifold**:

- $\mathcal{M} \subset \overline{\mathcal{M}} \equiv \mathbb{R}^{m \times n}$
- $\Phi(x) := \operatorname{grad} l^{\rho_k}\left(x, y^k\right)$

Semismooth Newton method for the subproblem

Recall the RALM subproblem:

$$x^{k+1} \approx \operatorname*{argmin}_{x \in \mathcal{U}} l^{\rho_k}\left(x, y^k\right)$$

Consider the **embedded submanifold**:

- $\mathcal{M} \subset \overline{\mathcal{M}} \equiv \mathbb{R}^{m \times n}$
- $\Phi(x) := \operatorname{grad} l^{\rho_k}\left(x, y^k\right)$

Vector fields: the Riemannian gradient $\operatorname{grad} f(x) : \mathcal{M} \to T\mathcal{M}$ is given by

$$\operatorname{grad} f(x) = \Pi_x(\nabla f(x))$$

- Π_x is the projection onto $T_x\mathcal{M}$.

Usually, we have $\nabla f(x)$ is (globally) **Lipschitz continuous** and even **(strongly) semismooth** in the ambient space $\bar{\mathcal{M}}$

Semismooth Newton method for the subproblem

Recall the RALM subproblem:

$$x^{k+1} \approx \underset{x \in \mathcal{U}}{\operatorname{argmin}} \; l^{\rho_k}\left(x, y^k\right)$$

Consider the **embedded submanifold**:

- $\mathcal{M} \subset \overline{\mathcal{M}} \equiv \mathbb{R}^{m \times n}$
- $\Phi(x) := \operatorname{grad} l^{\rho_k}\left(x, y^k\right)$

Vector fields: the Riemannian gradient $\operatorname{grad} f(x) : \mathcal{M} \to T\mathcal{M}$ is given by

$$\operatorname{grad} f(x) = \Pi_x(\nabla f(x))$$

- Π_x is the projection onto $T_x\mathcal{M}$.

Usually, we have $\nabla f(x)$ is (globally) **Lipschitz continuous** and even **(strongly) semismooth** in the ambient space $\bar{\mathcal{M}}$

We typically interest in solving the following nonsmooth equation:

$$\Phi(x) := \operatorname{grad} f(x) = 0$$

A natural and fundamental problem:

When the vector field $\Phi(x)$ is **semismooth**?

A natural and fundamental problem:

When the vector field $\Phi(x)$ is **semismooth**?

Zhou, Bao, **D.** and Zhu, **MP** (2023):

$\bar{\Phi} \equiv \nabla f$ in $\overline{\mathcal{M}} \equiv \mathbb{R}^{m\times n}$:		Φ in $\mathcal{M} \subset \overline{\mathcal{M}}$:
• **Lipschitz continuous**	$\Longrightarrow$	• **Lipschitz continuous**
• **(strongly) semismooth**		• **(strongly) semismooth**

Local convergence of semismooth Newton method

- To solve $\Phi = \operatorname{grad} f(p) = 0$.

Choose $H_k \in \mathcal{K}(p_k)$, use CG to find $V_k \in T_{p_k}\mathcal{M}$, such that

$$\|(H_k + \omega_k I)V_k + \Phi(p_k)\| \le \tilde{\eta}_k$$

where $\omega_k = \|\Phi(p_k)\|^\nu$ and $\tilde{\eta}_k$ is a sequence converges to 0.

Local convergence of semismooth Newton method

- To solve $\Phi = \operatorname{grad} f(p) = 0$.

Choose $H_k \in \mathcal{K}(p_k)$, use CG to find $V_k \in T_{p_k}\mathcal{M}$, such that

$$\|(H_k + \omega_k I)V_k + \Phi(p_k)\| \leq \tilde{\eta}_k$$

where $\omega_k = \|\Phi(p_k)\|^\nu$ and $\tilde{\eta}_k$ is a sequence converges to 0.

Theorem

Let $\mathcal{K} = \partial\Phi$. Denote p_* be any accumulation point of $\{p_k\}$. If Φ is semismooth at p_* with order ν with respect to $\mathcal{K}$, and all elements of $\mathcal{K}(p_*)$ are **positive definite**, *then we have $p_k \to p_*$ as $k \to \infty$ and for sufficiently large k, it holds*

$$d(p_{k+1}, p_*) \leq O\left(d(p_k, p_*)^{1+\min\{\nu,\bar{\nu}\}}\right).$$

Characterization of strong variational sufficiency for NManOPs (II)

Theorem

Let $\bar{x}$ and $\bar{y}$ satisfy the first-order optimality condition. Suppose that θ is a **polyhedral convex function** *or the indicator function of* **second-order cone** *or* **positive semidefinite cone**. *Then, the following conditions are equivalent.*

- *the manifold strong variational sufficient condition with respect to $(\bar{x}, \bar{y})$;*
- *the manifold strong second-order sufficient condition (***M-SSOSC***) holds at $(\bar{x}, \bar{y})$;*
- *any $H \in \mathcal{K}(\bar{x})$ is* **positive definite**.

Numerical experiments

Convergence rate: robust matrix completion

For a given $A \in \mathbb{R}^{m\times n}$, $\mathcal{M} = \mathrm{Fr}(m,n,r) := \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) = r\}$, consider

$$\begin{array}{ll} \min_{X\in\mathbb{R}^{m\times n}} & \|P_\Omega(X-A)\|_{\ell_1} \\ \text{s.t.} & X \in \mathrm{Fr}(m,n,r), \end{array}$$

Consider a basic example: Ω is the full index set. Let

$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix}^T$, $V = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & -0.8 & 0 & 0 \\ 0 & 0.8 & 0.6 & 0 & 0 \end{bmatrix}^T$ and

$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The observed matrix is set to $A = A_{\text{ex}} + E_{\text{out}}$, where $A_{\text{ex}} = USV^T$ is the assumed ground truth and E_{out} is a matrix with random entries added only in the lower right 2×2 submatrix. Since A_{ex} is of rank $r = 3$, $X^* = A_{\text{ex}}$ is a solution of this problem and $y^*_{ij} = \mathrm{sgn}(E^{ij}_{\text{out}})$ is a corresponding multiplier.

It can be checked directly that the **M-SSOSC** holds at (X^*, y^*)!

Riemannian ALM:

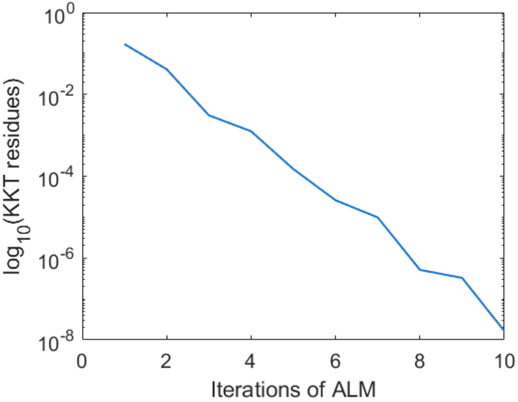
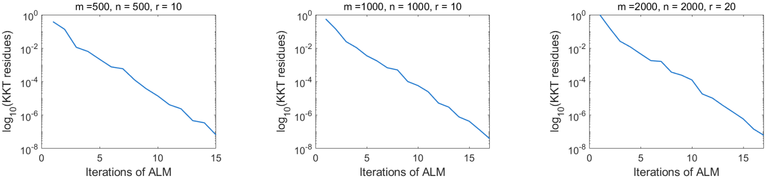


Figure 3: the KKT residues generated by inexact ALM

Convergence rate: robust matrix completion (con't)



(a) m=n=500, r=10 **(b)** m=n=1000, r=10 **(c)** m=n=2000, r=20

Figure 4: The KKT residues of robust matrix completion problems generated by inexact ALM

The robust matrix completion

The performance of Riemannian augmented Lagrangian method for the robust matrix completion problem.

m	n	r	it	time(sec)	KKT residual	$\|X - A_{\text{ex}}\|$
500	500	10	15	5.24	4.8339e-08	1.6186e-08
1000	1000	10	17	14.87	3.5088e-08	8.3608e-08
2000	2000	20	17	33.47	2.5007e-08	2.4877e-08
5000	5000	20	30	430.02	6.6165e-09	3.4093e-09

The CM problem revisited

Ozoliņš, Lai, Caflisch, and Osher **PNAS**, 2013: The Schrödinger equation of 1D free-electron model with periodic boundary condition:

$$-\frac{1}{2}\Delta\phi(x) = \lambda\phi(x), \quad x \in [0, 2]$$

Recall the compressed mode (CM) problem:

$$\min_{X \in \mathrm{St}(n,r)} \mathrm{tr}(X^T H X) + \mu\|X\|_1,$$

where H is a discretization of the Hamilton operator

		ManPG	AManPG	ARPG	SOC	LSe-I	LSe-II	LSq-I	LSq-II	LSp-I	LSp-II
		Running Time (s)									
n	200	11.54	3.86	6.43	1.78	1.20	1.70	1.21	1.76	**1.10**	1.54
	500	21.02	8.32	9.15	5.41	4.00	6.19	**3.96**	5.88	4.15	6.14
	1000	66.30	**8.60**	11.30	14.99	12.06	11.34	14.07	11.11	11.93	9.49
	1500	44.73	40.18	42.85	39.69	**24.00**	27.93	24.42	27.85	25.88	33.46
	2000	42.48	46.86	51.47	46.33	33.50	28.42	31.77	29.05	27.91	**26.22**
r	10	15.35	15.63	15.71	17.28	31.60	12.74	70.55	15.74	80.23	**11.26**
	15	35.35	**9.55**	11.17	15.47	49.14	12.07	53.59	11.74	51.58	12.01
	25	87.64	22.73	23.93	27.11	**17.39**	20.99	17.79	20.61	18.41	20.53
	30	83.13	27.41	29.35	18.10	11.49	15.49	11.40	15.13	**11.02**	14.42
μ	0.05	102.63	14.86	13.98	**6.34**	7.19	8.04	7.17	7.97	7.03	7.60
	0.15	65.08	17.09	21.48	32.51	14.22	16.03	14.13	15.69	**13.98**	15.59
	0.20	51.46	**12.96**	24.12	33.71	17.31	20.66	19.94	19.92	21.92	22.65
	0.25	42.74	**13.41**	25.43	34.80	18.08	20.84	61.23	53.44	35.93	44.30

Figure 5: Zhou, Bao, **D.** and Zhu, **MP** (2023)

The CM problem revisited (cont')

The minimum eigenvalue of the (generalized) Hessian matrix of $\Phi(x)$ in the CM problem. $(n, r, \mu) = (1000, 20, 0.1)$ and one of them varies. We report the results of 5 different runs.

		Minimum Eigenvalue ($\times 10$)				
n	200	3.04960	3.05820	3.05160	3.23330	1.75420
	500	0.97663	1.08610	1.14290	1.17090	1.01000
	1000	0.13422	2.06550	0.22452	1.88420	0.05863
	1500	0.65285	0.61208	0.63683	0.86749	0.80422
	2000	0.24362	0.24363	0.24154	0.24101	0.23518
r	10	0.00264	0.00570	0.00287	0.00901	0.00005
	15	0.06096	0.05403	0.01370	0.00334	0.06086
	25	0.60886	0.60154	0.36930	0.76388	0.06192
	30	3.01550	6.18060	2.96390	6.18660	5.35900
μ	0.05	0.58725	1.03500	1.01720	1.12870	1.15590
	0.15	0.57333	0.73569	0.39304	0.57299	0.53624
	0.20	0.34736	0.20896	0.45627	0.34404	0.03558
	0.25	0.01539	0.16556	0.05790	0.01262	0.05763

The CM problem revisited (cont')

Consider the Schrödinger equation of with boundary condition when $x \in [0,2]$. Discretize the domain $[0,2]$ into $n=4$ nodes. Let H be a discretization of the Hamilton operator, i.e.,

$$H = -\begin{bmatrix} -4 & 2 & 0 & 2 \\ 2 & -4 & 2 & 0 \\ 0 & 2 & -4 & 2 \\ 2 & 0 & 2 & -4 \end{bmatrix}$$

For $r=2$, $X^* = \begin{bmatrix} 0 & 0 & \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 \end{bmatrix}^\top$ is a local optimal of the CM problem if $\mu < 5\sqrt{2}$ and $y = \mu \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}^\top$ is a corresponding multiplier.

It can be checked directly that the **M-SSOSC** holds at (X^*, y^*)!

Summary

In this talk:

- The RALM for NManOP
- Characterization of Manifold strong variational sufficiency
- Fast linear local convergence rate of RALM without CQs
- Semismoothness on manifolds

Thank you.