Matrix optimization:

recent progress on algorithm foundation

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

ISMP2018, Bordeaux, France
July 6, 2018

中国科学院
CHINESE ACADEMY OF SCIENCES

Acknowledgements

Based on the joint work with **Ying Cui** at *USC* and **Xinyuan Zhao** at *BISEC, Beijing*.

- Y. Cui, C. Ding and X.Y. Zhao, *Quadratic Growth Conditions for Convex Matrix Optimization Problems Associated with Spectral Functions*, **SIAM Journal on Optimization** 27, 2332–2355 (2017).
- Y. Cui and C. Ding, *Nonsmooth Composite Matrix Optimizations: Strong Regularity, Constraint Nondegeneracy and Beyond*, **preprint** (2018).

USC University of Southern California

BISEC
BEIJING INSTITUTE FOR SCIENTIFIC AND ENGINEERING COMPUTING
北京科学与工程计算研究院

Outline

Matrix optimization problem

Augmented Lagrangian method

- **The convex case**
- **The non-convex case**

Two perturbation results

- **The calmness of X**
- **The strong regularity of the general MOP**

Matrix optimization problem

Matrix Optimization Problem (MOP)

MOP:

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x) + \theta(g(x)) \\ \text{s.t.} \quad & h(x) \in \mathcal{Q} \end{aligned}$$

Matrix Optimization Problem (MOP)

MOP:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{X}} & f(x)+\theta(g(x)) \\ \text{s.t.} & h(x)\in\mathcal{Q} \end{array}$$

- $\mathbb{X}$: a finite dimensional Euclidean space

Matrix Optimization Problem (MOP)

MOP:

$$\begin{aligned} \min_{x \in \mathbb{X}} \quad & f(x) + \theta(g(x)) \\ \text{s.t.} \quad & h(x) \in \mathcal{Q} \end{aligned}$$

- $\mathbb{X}$: a finite dimensional Euclidean space
- f: a smooth function (e.g., the data-fitting term)

Matrix Optimization Problem (MOP)

MOP:

$$\begin{aligned} \min_{x \in \mathbb{X}} \quad & f(x) + \theta(g(x)) \\ \text{s.t.} \quad & h(x) \in \mathcal{Q} \end{aligned}$$

- $\mathbb{X}$: a finite dimensional Euclidean space
- f: a smooth function (e.g., the data-fitting term)
- θ: a function defined on the finite dimensional matrix space (symmetric/Hermitian or non-symmetric/non-Hermitian)

Matrix Optimization Problem (MOP)

MOP:

$$\begin{array}{ll} \min_{x \in \mathbb{X}} & f(x) + \theta(g(x)) \\ \text{s.t.} & h(x) \in \mathcal{Q} \end{array}$$

- $\mathbb{X}$: a finite dimensional Euclidean space
- f: a smooth function (e.g., the data-fitting term)
- θ: a function defined on the finite dimensional matrix space (symmetric/Hermitian or non-symmetric/non-Hermitian)
  the regularization term (structured features, e.g., lower rank), e.g., $\delta_{\mathcal{S}^n_+}(\cdot)$, $\|\cdot\|_*$, $\|\cdot\|_{(k)}$, Schatten norm.
- g and h: smooth functions

Matrix Optimization Problem (MOP)

MOP:

$$\begin{array}{ll} \min_{x \in \mathbb{X}} & f(x) + \theta(g(x)) \\ \text{s.t.} & h(x) \in \mathcal{Q} \end{array}$$

- $\mathbb{X}$: a finite dimensional Euclidean space
- f: a smooth function (e.g., the data-fitting term)
- θ: a function defined on the finite dimensional matrix space (symmetric/Hermitian or non-symmetric/non-Hermitian) the regularization term (structured features, e.g., lower rank), e.g., $\delta_{\mathcal{S}^n_+}(\cdot)$, $\|\cdot\|_*$, $\|\cdot\|_{(k)}$, Schatten norm.
- g and h: smooth functions
- $\mathcal{Q}$: a convex polyhedral set

MOPs in Data Science

The **regularization term** and has the following **composite structure**:

$$\theta(X) = \phi(\sigma(X))$$

MOPs in Data Science

The **regularization term** and has the following **composite structure**:

$$\theta(X) = \phi(\sigma(X))$$

- $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is an (**absolutely**) **symmetric** function
- $\sigma(X) = (\sigma_1(X), \ldots, \sigma_m(X))$ with $\sigma_1(X) \geq \ldots \geq \sigma_m(X)$: **spectral** (**singular values**/**eigenvalues**) of X

MOPs in Data Science

The **regularization term** and has the following **composite structure**:

$$\theta(X) = \phi(\sigma(X))$$

- $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is an (**absolutely**) **symmetric** function
- $\sigma(X) = (\sigma_1(X), \ldots, \sigma_m(X))$ with $\sigma_1(X) \geq \ldots \geq \sigma_m(X)$: **spectral** (**singular values**/**eigenvalues**) of X

e.g., $\delta_{\mathcal{S}^n_+}$; δ_{EDM}; $\|\cdot\|_*$; the spectral norm $\|\cdot\|_2$; Ky Fan's k-norm $\|\cdot\|_{(k)}$

MOPs in Data Science

The **regularization term** and has the following **composite structure**:

$$\theta(X) = \phi(\sigma(X))$$

- $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is an (**absolutely**) **symmetric** function
- $\sigma(X) = (\sigma_1(X), \ldots, \sigma_m(X))$ with $\sigma_1(X) \geq \ldots \geq \sigma_m(X)$: **spectral** (**singular values**/**eigenvalues**) of X

e.g., $\delta_{\mathcal{S}^n_+}$; $\delta_{\rm EDM}$; $\|\cdot\|_*$; the spectral norm $\|\cdot\|_2$; Ky Fan's k-norm $\|\cdot\|_{(k)}$

Matrix completion:

$$\min \left\{ \|P_\Omega(X) - P_\Omega(M)\|_F^2 + \lambda \|X\|_* \mid \mathcal{A}(X) = b,\ X \in \mathcal{P} \right\}$$

— $P_\Omega(\cdot)$: the observation operator

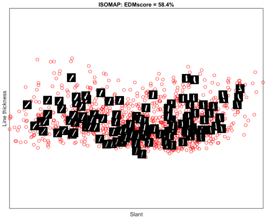
More applications

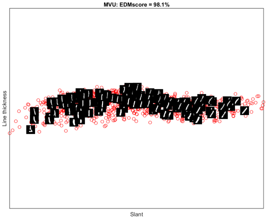
- SDP
- Fastest mixing Markov chain problem (fast load balancing of paralleled systems)
- Fastest distributed linear averaging problem
- The reduced rank approximations of transition matrices
- The low rank approximations of doubly stochastic matrices
- Low-rank approximation of matrices with linear structures
- ......

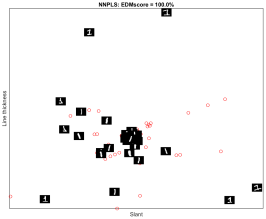
More applications

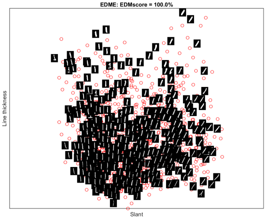
- SDP
- Fastest mixing Markov chain problem (fast load balancing of paralleled systems)
- Fastest distributed linear averaging problem
- The reduced rank approximations of transition matrices
- The low rank approximations of doubly stochastic matrices
- Low-rank approximation of matrices with linear structures
- ......
- **Unsupervised learning**

MOPs in unsupervised learning









MNIST dataset. Source: **D. and Qi, (2017)**

Solving MOPs: an example

The fastest mixing Markov chain problem:[1]

- find the edge transition probabilities that give the fastest mixing Markov chain
- equivalent to minimize the **second largest** eigenvalue modulus of P

$$\begin{array}{ll} \min & \|P\|_{(2)} \\ \text{s.t.} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad P = P^T \\ & P_{ij} = 0, \quad (i,j) \notin \mathcal{E} \end{array}$$

[1] S. Boyd, P. Diaconis, and L. Xiao, SIAM Review, 46 (2004), pp. 667–689.

Numerical results

Table 1: The performance of the **augmented Lagrangian method** (**ALM**) and the ADMM for solving the FMMC problems. The computational time is in the format of "hh:mm:ss".

		iteration	η	time
problem	d; n	ALM | ADMM	ALM| ADMM	ALM | ADMM
cage	2562 ; 366	6;6;200 | 1925	0.0-7 | 9.1-7	05 | 37
G3	19176 ; 800	32;57;88 | 599	3.0-7 | 8.3-7	1:37 | 1:21
G6	9665 ; 800	30;44;145 | 989	8.5-7 | 9.6-7	1:09 | 1:52
G15	4661 ; 800	31;51;200 | 6122	3.4-8 | 7.7-7	1:05 | 11:27
G46	9990 ; 1000	30;44;134 | 1619	5.6-7 | 9.9-7	1:34 | 5:33
G54	5916 ; 1000	22;62;200 | 8928	7.7-7 | 9.9-7	2:23 | 27:16
G43	9990 ; 1000	24;96;90 | 2073	2.9-7 | 9.3-7	2:37 | 6:03
delaunayn10	3056 ; 1024	61;359;200 | 25000	6.8-9 | 7.2-5	10:25 | 1:04:29
G22	19990 ; 2000	31;46;56 | 2918	2.3-8 | 9.9-7	5:57 | 41:31
G24	19990 ; 2000	41;296;200 | 6808	2.7-7 | 9.9-7	53:57 | 1:38:37
G26	19990 ; 2000	29;87;200 | 2954	1.4-7 | 9.9-7	16:15 | 42:18
minnesota	3303 ; 2642	25;24;123 | 258	0.0-10 | 9.1-7	6:27 | 6:29
G48	6000 ; 3000	40;79;200 | 9470	9.2-7 | 9.2-7	19:39 | 4:40:08
G49	6000 ; 3000	25;38;200 | 7488	6.6-7 | 8.5-7	10:58 | 3:36:58
G50	6000 ; 3000	26;42;74 | 6370	4.8-8 | 7.8-7	9:36 | 2:59:34
USpowerGrid	6594 ; 4941	27;120;200 | 25000	1.4-7 | 1.0-5	3:11:22 | 56:47:18

Why ALM?

Why ALM?

A quick answer: **it just works!**

Why ALM?

A quick answer: **it just works!**

ALM has the **asymptotically superlinear** convergence / **linearly convergent** of an **arbitrary** order

Perturbed MOPs

Canonically perturbed MOPs with parameters (a,b,c):

$$\begin{aligned} \text{min} \quad & f(x) - \langle \mathbf{a}, x \rangle + \theta(g(x) + \mathbf{b}) \\ \text{s.t.} \quad & h(x) + \mathbf{c} \in \mathcal{Q} \end{aligned}$$

Perturbed MOPs

Canonically perturbed MOPs with parameters (a,b,c):

$$\begin{aligned} \min \quad & f(x) - \langle \mathbf{a}, x \rangle + \theta(g(x) + \mathbf{b}) \\ \text{s.t.} \quad & h(x) + \mathbf{c} \in \mathcal{Q} \end{aligned}$$

Under suitable CQs, the Karush-Kuhn-Tucker (KKT) optimality condition for perturbed problem takes the following form:

$$\begin{cases} \mathbf{a} = \nabla f(x) + g'(x)^* y + h'(x)^* z \\ y \in \partial\theta(g(x) + \mathbf{b}) \\ z \in \mathcal{N}_{\mathcal{K}}(h(x) + \mathbf{c}) \end{cases} \quad \Longleftrightarrow \quad \textbf{GE}: \quad (a, b, c) \in \mathcal{T}_L(x, y, z)$$

Perturbed MOPs

Canonically perturbed MOPs with parameters (a,b,c):

$$\begin{aligned} \min \quad & f(x) - \langle \mathbf{a}, x \rangle + \theta(g(x) + \mathbf{b}) \\ \text{s.t.} \quad & h(x) + \mathbf{c} \in \mathcal{Q} \end{aligned}$$

Under suitable CQs, the Karush-Kuhn-Tucker (KKT) optimality condition for perturbed problem takes the following form:

$$\begin{cases} \mathbf{a} = \nabla f(x) + g'(x)^* y + h'(x)^* z \\ y \in \partial\theta(g(x) + \mathbf{b}) \\ z \in \mathcal{N}_{\mathcal{K}}(h(x) + \mathbf{c}) \end{cases} \iff \mathbf{GE}: \quad (a, b, c) \in \mathcal{T}_L(x, y, z)$$

$X(a, b, c)$: the set of all locally optimal solutions

Perturbed MOPs

Canonically perturbed MOPs with parameters (a,b,c):

$$\begin{aligned} \min \quad & f(x) - \langle \mathbf{a}, x\rangle + \theta(g(x) + \mathbf{b}) \\ \text{s.t.} \quad & h(x) + \mathbf{c} \in \mathcal{Q} \end{aligned}$$

Under suitable CQs, the Karush-Kuhn-Tucker (KKT) optimality condition for perturbed problem takes the following form:

$$\left\{ \begin{aligned} & \mathbf{a} = \nabla f(x) + g'(x)^* y + h'(x)^* z \\ & y \in \partial\theta(g(x) + \mathbf{b}) \\ & z \in \mathcal{N}_{\mathcal{K}}(h(x) + \mathbf{c}) \end{aligned} \right. \quad \Longleftrightarrow \quad \textbf{GE}: \quad (a, b, c) \in \mathcal{T}_L(x, y, z)$$

$X(a, b, c)$: the set of all locally optimal solutions

$S_{\mathrm{KKT}}(a, b, c)$: the set of all solutions (x, y, z) to the KKT system

Perturbed MOPs (cont'd)

Let Ψ be a **set-valued mapping** with $(\overline{X}, \overline{Y}) \in \operatorname{gph} \Psi$. **Lipschitz-like** properties:

Perturbed MOPs (cont'd)

Let Ψ be a **set-valued mapping** with $(\overline{X}, \overline{Y}) \in \operatorname{gph} \Psi$. **Lipschitz-like** properties:

- **Calmness**:
  there exist neighborhoods U of $\overline{X}$, V of $\overline{Y}$ and constant $\kappa > 0$ such that
  $$\Psi(X) \cap V \subseteq \Psi(\overline{X}) + \kappa \|X - \overline{X}\| \mathbb{B} \quad \forall X \in U.$$

Perturbed MOPs (cont'd)

Let Ψ be a **set-valued mapping** with $(\overline{X}, \overline{Y}) \in \mathrm{gph}\,\Psi$. **Lipschitz-like** properties:

- **Calmness**:
  there exist neighborhoods U of $\overline{X}$, V of $\overline{Y}$ and constant $\kappa > 0$ such that
  $$\Psi(X) \cap V \subseteq \Psi(\overline{X}) + \kappa\|X - \overline{X}\|\mathbb{B} \quad \forall\, X \in U.$$
- **Robust isolated calmness**:
  there exist neighborhoods U of $\overline{X}$, V of $\overline{Y}$ and constant $\kappa > 0$ such that
  $$\emptyset \neq \Psi(X) \cap V \subseteq \{\overline{Y}\} + \kappa\|X - \overline{X}\|\mathbb{B} \quad \forall\, X \in U.$$

Let Ψ be a **set-valued mapping** with $(\overline{X}, \overline{Y}) \in \operatorname{gph} \Psi$. **Lipschitz-like** properties:

- **Calmness**:
  there exist neighborhoods U of $\overline{X}$, V of $\overline{Y}$ and constant $\kappa > 0$ such that
  $$\Psi(X) \cap V \subseteq \Psi(\overline{X}) + \kappa \|X - \overline{X}\| \mathbb{B} \quad \forall X \in U.$$
- **Robust isolated calmness**:
  there exist neighborhoods U of $\overline{X}$, V of $\overline{Y}$ and constant $\kappa > 0$ such that
  $$\emptyset \neq \Psi(X) \cap V \subseteq \{\overline{Y}\} + \kappa \|X - \overline{X}\| \mathbb{B} \quad \forall X \in U.$$

Strong regularity:

$S_{\mathrm{KKT}}(a, b, c)$ is locally Lipschitz continuous.

Why it matters?

Why it matters?

- **Theory**: perturbation analysis of OPs

Why it matters?

- **Theory**: perturbation analysis of OPs
- **Algorithm**: e.g., convergence rate of augmented Lagrangian method

Augmented Lagrangian method

Augmented Lagrangian method (ALM)

Initiated by **Hestenes (1969)**; **Powell (1969)**

Augmented Lagrangian method (ALM)

Initiated by **Hestenes (1969)**; **Powell (1969)**

- Convex case: PPA
  - global convergence
  - local (super)linear convergence

  Rockafellar (1976a), (1976b); **Pennanen (2002)**; ...

Augmented Lagrangian method (ALM)

Initiated by **Hestenes (1969)**; **Powell (1969)**

- Convex case: PPA
  - global convergence
  - local (super)linear convergence

  Rockafellar (1976a), (1976b); **Pennanen (2002)**; ...
- Non-convex case: for NLP, under suitable conditions, local convergence at a linear rate
  Powell (1972); **Rockafellar (1973a), (1973b)**; **Tretyakov (1973)**, **Bertsekas (1976)**; **Conn et al. (1991)**, **Contesse-Back (1993)**; **Ito and Kunisch (1990)**; ...

Augmented Lagrangian method (ALM)

Initiated by **Hestenes (1969)**; **Powell (1969)**

- Convex case: PPA
  - global convergence
  - local (super)linear convergence

  Rockafellar (1976a), (1976b); **Pennanen (2002)**; ...
- Non-convex case: for NLP, under suitable conditions, local convergence at a linear rate
  Powell (1972); **Rockafellar (1973a), (1973b)**; **Tretyakov (1973)**, **Bertsekas (1976)**; **Conn et al. (1991)**, **Contesse-Back (1993)**; **Ito and Kunisch (1990)**; ...

A survey paper:

R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35: 183–238, 1993.

Augmented Lagrangian method

The convex case

The non-convex case

The convex case

Consider the following quadratic MOP and its dual:

$$\begin{array}{lll} (P) & \min & \frac{1}{2}\|\mathcal{F}X - d\|^2 + \langle C, X\rangle + \theta(X) \\ & \text{s.t.} & \mathcal{A}X = b,\ X \in \mathcal{P} \end{array}$$

$$\begin{array}{lll} (D) & \min & \frac{1}{2}\|w\|^2 - \langle w, d\rangle - \langle b, y\rangle + \theta^*(-S) + \delta_P^*(-Z) \\ & \text{s.t.} & \mathcal{F}^* w + \mathcal{A}^* y + S + Z = C \end{array}$$

The convex case

Consider the following quadratic MOP and its dual:

$$\begin{aligned}(P) \quad \min \quad & \tfrac{1}{2}\|\mathcal{F}X - d\|^2 + \langle C, X\rangle + \theta(X) \\ \text{s.t.} \quad & \mathcal{A}X = b,\ X \in \mathcal{P}\end{aligned}$$

$$\begin{aligned}(D) \quad \min \quad & \tfrac{1}{2}\|w\|^2 - \langle w, d\rangle - \langle b, y\rangle + \theta^*(-S) + \delta_P^*(-Z) \\ \text{s.t.} \quad & \mathcal{F}^* w + \mathcal{A}^* y + S + Z = C\end{aligned}$$

The **Lagrange function** of (D):

$$\begin{aligned}L(w, y, S, Z, X) := \quad & \tfrac{1}{2}\|w\|^2 - \langle b, y\rangle - \langle d, w\rangle + \theta^*(-S) + \delta_P^*(-Z) \\ & + \langle X, \mathcal{F}^* w + \mathcal{A}^* y + S + Z - C\rangle,\end{aligned}$$

ALM (cont'd)

The **augmented Lagrangian function** associated with (D)

$$L_\sigma(w, y, S, Z, X) := L(w, y, S, Z, X) + \frac{\sigma}{2}\|\mathcal{F}^* w + \mathcal{A}^* y + S + Z - C\|^2.$$

ALM for solving (D):

Given a sequence of scalars $\sigma_k \uparrow \sigma_\infty \leq \infty$ and a starting point $X^0 \in \mathbb{R}^{m\times n}$, the $(k+1)$-th iteration of the **inexact** ALM:

$$\begin{cases} z^{k+1} := (w^{k+1}, y^{k+1}, S^{k+1}, Z^{k+1}) \approx \arg\min \zeta_k(z) := L_{\sigma_k}(w, y, S, Z, X^k) \\ X^{k+1} = X^k + \sigma_k(\mathcal{F}^* w^{k+1} + \mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C) \end{cases}$$

ALM (cont'd)

The **augmented Lagrangian function** associated with (D)

$$L_\sigma(w, y, S, Z, X) := L(w, y, S, Z, X) + \frac{\sigma}{2}\|\mathcal{F}^* w + \mathcal{A}^* y + S + Z - C\|^2.$$

ALM for solving (D):

Given a sequence of scalars $\sigma_k \uparrow \sigma_\infty \leq \infty$ and a starting point $X^0 \in \mathbb{R}^{m\times n}$, the $(k+1)$-th iteration of the **inexact** ALM:

$$\begin{cases} z^{k+1} := (w^{k+1}, y^{k+1}, S^{k+1}, Z^{k+1}) \approx \arg\min \zeta_k(z) := L_{\sigma_k}(w, y, S, Z, X^k) \\ X^{k+1} = X^k + \sigma_k(\mathcal{F}^* w^{k+1} + \mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C) \end{cases}$$

Terminate the subproblem for solving z^{k+1} by one of following criteria:

$$(A) \quad \zeta_k(z^{k+1}) - \inf \zeta_k(z) \leq \varepsilon_k^2/2\sigma_k, \ \varepsilon_k \geq 0, \ \sum_{k=0}^{\infty} \varepsilon_k < \infty$$

$$(B) \quad \zeta_k(z^{k+1}) - \inf \zeta_k(z) \leq (\eta_k^2/2\sigma_k)\|X^{k+1} - X^k\|^2, \ \eta_k \geq 0, \ \sum_{k=0}^{\infty} \eta_k < \infty$$

ALM & PPA

ALM for (D) $\Longleftrightarrow$ **PPA** (**proximal point algorithm**) for (P)

ALM & PPA

ALM for (D) $\Longleftrightarrow$ **PPA** (**proximal point algorithm**) for (P)

The **essential objective function** of (P)

$F(X) := \inf \left\{ L(w, y, S, Z, X) \mid (w, y, S, Z) \in \Re^d \times \Re^e \times \Re^{m\times n} \times \Re^{m\times n} \right\}$

$\mathcal{T}_{\mathbf{F}} := \partial F$ — a maximal monotone operator

Optimal solution set of (P): $X(0) = \{X \in \mathcal{X} \mid 0 \in \mathcal{T}_F(X)\}$

ALM & PPA

ALM for (D) $\Longleftrightarrow$ **PPA** (**proximal point algorithm**) for (P)

The **essential objective function** of (P)

$F(X) := \inf \left\{ L(w, y, S, Z, X) \mid (w, y, S, Z) \in \Re^d \times \Re^e \times \Re^{m\times n} \times \Re^{m\times n} \right\}$

$\mathcal{T}_{\mathbf{F}} := \partial F$ — a maximal monotone operator

Optimal solution set of (P): $X(0) = \{X \in \mathcal{X} \mid 0 \in \mathcal{T}_F(X)\}$

Given $\sigma_k > 0$, the **proximal mapping** associated with $\sigma_k \mathcal{T}_F$:

$$\mathcal{P} := (\mathcal{I} + \sigma_k \mathcal{T}_F)^{-1}$$

The **proximal point algorithm** (PPA):

$$X^{k+1} \approx \mathcal{P}_k(X^k), \quad \mathcal{P}_k := (\mathcal{I} + \sigma_k \mathcal{T}_F)^{-1}$$

ALM & PPA

ALM for (D) $\Longleftrightarrow$ **PPA** (**proximal point algorithm**) for (P)

The **essential objective function** of (P)

$F(X) := \inf \left\{ L(w, y, S, Z, X) \mid (w, y, S, Z) \in \Re^d \times \Re^e \times \Re^{m\times n} \times \Re^{m\times n} \right\}$

$\mathcal{T}_{\mathbf{F}} := \partial F$ — a maximal monotone operator

Optimal solution set of (P): $X(0) = \{X \in \mathcal{X} \mid 0 \in \mathcal{T}_F(X)\}$

Given $\sigma_k > 0$, the **proximal mapping** associated with $\sigma_k \mathcal{T}_F$:

$$\mathcal{P} := (\mathcal{I} + \sigma_k \mathcal{T}_F)^{-1}$$

The **proximal point algorithm** (PPA):

$$X^{k+1} \approx \mathcal{P}_k(X^k), \quad \mathcal{P}_k := (\mathcal{I} + \sigma_k \mathcal{T}_F)^{-1}$$

Criteria for approximate calculation of $\mathcal{P}_k(X^k)$:

$$(A) \quad \|X^{k+1} - \mathcal{P}_k(X^k)\| \le \varepsilon_k^2 / 2\sigma_k$$

$$(B) \quad \|X^{k+1} - \mathcal{P}_k(X^k)\| \le (\eta_k^2 / 2\sigma_k) \|X^{k+1} - X^k\|^2$$

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

Robust isolated calmness: requires the solution of problem (P) to be **unique** — **Restrictive**

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.
[3] F.J. Luque. *SIAM Journal on Control and Optimization*, 22(2): 277-293, 1984.

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

Robust isolated calmness: requires the solution of problem (P) to be **unique** — **Restrictive**

Luque[3] relaxed this with an error bound type condition

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.
[3] F.J. Luque. *SIAM Journal on Control and Optimization*, 22(2): 277-293, 1984.

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

Robust isolated calmness: requires the solution of problem (P) to be **unique** — **Restrictive**

Luque[3] relaxed this with an error bound type condition

- allow the multi-solution case and is satisfied when $\mathcal{T}_F$ is **polyhedral** (e.g., LASSO)

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.
[3] F.J. Luque. *SIAM Journal on Control and Optimization*, 22(2): 277-293, 1984.

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

Robust isolated calmness: requires the solution of problem (P) to be **unique** — **Restrictive**

Luque[3] relaxed this with an error bound type condition

- allow the multi-solution case and is satisfied when $\mathcal{T}_F$ is **polyhedral** (e.g., LASSO)
- if $\mathcal{T}_F$ is **non-polyhedral**, it is difficult to check **Luque's condition** (counterexamples)

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.
[3] F.J. Luque. *SIAM Journal on Control and Optimization*, 22(2): 277-293, 1984.

Convergent rate

Rockafellar (1976)[2] established the convergence rate under the **robust isolated calmness**

Robust isolated calmness: requires the solution of problem (P) to be **unique** — **Restrictive**

Luque[3] relaxed this with an error bound type condition

- allow the multi-solution case and is satisfied when $\mathcal{T}_F$ is **polyhedral** (e.g., LASSO)
- if $\mathcal{T}_F$ is **non-polyhedral**, it is difficult to check **Luque's condition** (counterexamples)

> **Question**: For MOPs, how to ensure the convergence of ALM and how fast it can be?

[2] R.T. Rockafellar. *SIAM Journal on Control and Optimization*, 14(5): 877–898, 1976.
[3] F.J. Luque. *SIAM Journal on Control and Optimization*, 22(2): 277-293, 1984.

Convergence rate of ALM

Theorem (asymptotic superlinear for ALM)

Assume that $\mathcal{T}_F^{-1}(0) \neq \emptyset$. *(i) Let* $\{X^k\}$ *be a sequence generated by the PPA with stopping criterion* (A). *Then* $\{X^k\}$ *converges to some* $X^\infty \in X(0)$. *(ii) If the criterion (B) is also employed and optimal solution mapping* $X = \mathcal{T}_F^{-1}$ *is* **calm** *at the origin for* X^∞ *with modulus* $\kappa_p \geq 0$, *then there exists* $\bar{k} \geq 0$ *such that for all* $k \geq \bar{k}$, $\eta_k < 1$ *and*

$$\operatorname{dist}(X^{k+1}, \mathcal{T}_F^{-1}(0)) \leq \theta_k \operatorname{dist}(X^k, \mathcal{T}_F^{-1}(0)),$$

where

$$\theta_k = (\mu_k + 2\eta_k)(1-\eta_k)^{-1} \quad \textit{with} \quad \mu_k = \kappa_p / \sqrt{\kappa_p^2 + \sigma_k^2}\ ,$$

$$\theta_k \to \theta_\infty = \kappa_p / \sqrt{\kappa_p^2 + \sigma_\infty^2} \quad (\theta_\infty = 0 \text{ if } \sigma_\infty = \infty).$$

$\|\mathcal{F}^* w^{k+1} + \mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C\| \leq \tau_k^1 \operatorname{dist}(X^k, \mathcal{T}_F^{-1}(0))$ *and* $\Psi(z^{k+1}) - \Psi^* \leq \tau_k^2 \operatorname{dist}(X^k, \mathcal{T}_F^{-1}(0))$. *where* $\tau_k^1 \to \tau_\infty^1 = 1/\sigma_\infty$, $\tau_k^2 \to \tau_\infty^2 = \|X^\infty\|/\sigma_\infty$ *(*$\tau_\infty^1 = \tau_\infty^2 = 0$ if $\sigma_\infty = \infty$*).*

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)}$$

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)} \quad \left(\mathrm{CG}: \frac{\sqrt{\kappa_p} - 1}{\sqrt{\kappa_p} + 1} \approx 1\right)$$

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)} \quad \left(\mathrm{CG}: \frac{\sqrt{\kappa_p} - 1}{\sqrt{\kappa_p} + 1} \approx 1\right)$$

Powell (1969)[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

[4] M.J.D. Powell. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)} \quad \left(\mathrm{CG}: \frac{\sqrt{\kappa_p} - 1}{\sqrt{\kappa_p} + 1} \approx 1\right)$$

Powell (1969)[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \phi(x) = 0 \end{array} \qquad \textbf{ALM}: \quad \begin{cases} x^{k+1} = \mathrm{argmin}_x \; L_\sigma(x; y) \\ y^{k+1} = y^k + \sigma\phi(x^{k+1}) \end{cases}$$

[4] M.J.D. Powell. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)} \quad \left(\mathrm{CG}: \frac{\sqrt{\kappa_p} - 1}{\sqrt{\kappa_p} + 1} \approx 1\right)$$

Powell (1969)[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \phi(x) = 0 \end{array} \qquad \textbf{ALM}: \quad \begin{cases} x^{k+1} = \operatorname{argmin}_x \, L_\sigma(x; y) \\ y^{k+1} = y^k + \sigma\phi(x^{k+1}) \end{cases}$$

$y^{k+1} = y^k + \sigma\phi(x^{k+1}) \Leftarrow$ **gradient method** for $\Psi_\sigma(y) := \phi(x(y, \sigma)) = 0$

[4] M.J.D. Powell. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

Convergence rate of ALM (cont'd)

Roughly speaking, $X^k \to X^\infty$ **linearly** with a rate bounded from above by

$$\frac{\kappa_p}{\sqrt{\kappa_p^2 + \sigma_\infty^2}} < 1 \quad \textbf{(fast linear)} \quad (\mathrm{CG}: \frac{\sqrt{\kappa_p} - 1}{\sqrt{\kappa_p} + 1} \approx 1)$$

Powell (1969)[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \phi(x) = 0 \end{array} \qquad \textbf{ALM}: \quad \begin{cases} x^{k+1} = \operatorname{argmin}_x \, L_\sigma(x; y) \\ y^{k+1} = y^k + \sigma\phi(x^{k+1}) \end{cases}$$

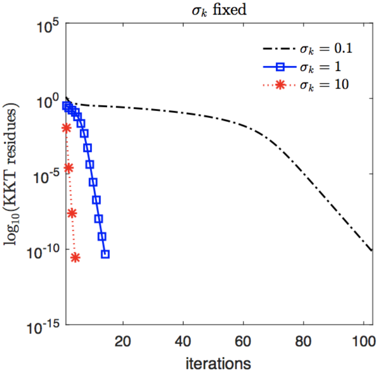
$y^{k+1} = y^k + \sigma\phi(x^{k+1}) \Leftarrow$ **gradient method** for $\Psi_\sigma(y) := \phi(x(y, \sigma)) = 0$

Powell (1969) shows that the Jacobian $-J_\sigma^*$ of Ψ_σ^* satisfies

$$\|-J_\sigma^* - I\| = O(\frac{1}{\sigma - c})$$

[4] M.J.D. Powell. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

Convergence rate of ALM (cont'd)



Source: **Cui, Sun and Toh (2018)**

Augmented Lagrangian method

The convex case

The non-convex case

The non-convex case

- For the nonlinear SDP problem, **Sun, Sun and Zhang (2007)**[5]: ALM converges linearly under the **strong regularity**, i.e., $S_{\rm KKT}$ is locally Lipschitz continuous

[5] D.F. Sun, J. Sun and L.W. Zhang. *Mathematical Programming*, 144: 349–391, 2007

The non-convex case

- For the nonlinear SDP problem, **Sun, Sun and Zhang (2007)**[5]: ALM converges linearly under the **strong regularity**, i.e., S_{KKT} is locally Lipschitz continuous
- For the general conic optimization, **Kanzow and Steck (2018)**[6]: a modified ALM converges linearly under the **robust isolated calmness** of S_{KKT}

---

[5] D.F. Sun, J. Sun and L.W. Zhang. *Mathematical Programming*, 144: 349–391, 2007
[6] C. Kanzow and D. Steck. *Mathematical Programming*, 2018.

The non-convex case

- For the nonlinear SDP problem, **Sun, Sun and Zhang (2007)**[5]: ALM converges linearly under the **strong regularity**, i.e., S_{KKT} is locally Lipschitz continuous
- For the general conic optimization, **Kanzow and Steck (2018)**[6]: a modified ALM converges linearly under the **robust isolated calmness** of S_{KKT}
- A characterization for S_{KKT} in general **non-polyhedral** and **non-convex** cases is provided by **D., Sun and Zhang (2017)**[7]

[5] D.F. Sun, J. Sun and L.W. Zhang. *Mathematical Programming*, 144: 349–391, 2007
[6] C. Kanzow and D. Steck. *Mathematical Programming*, 2018.
[7] C. Ding, D.F. Sun and L.W. Zhang. *SIAM Journal on Optimization*, 27: 67-90, 2017.

Two perturbation results

Calmness and strong regularity

- The **calmness** of the optimal solution X for the convex MOPs, i.e., there exist neighborhoods U of $\bar{u}$, V of $\overline{X} \in X(\bar{u})$ and constant $\kappa > 0$ such that

$$X(u) \cap V \subseteq X(\bar{u}) + \kappa \|u - \bar{u}\| \mathbb{B} \quad \forall\, u \in U.$$

Calmness and strong regularity

- The **calmness** of the optimal solution X for the convex MOPs, i.e., there exist neighborhoods U of $\bar{u}$, V of $\overline{X} \in X(\bar{u})$ and constant $\kappa > 0$ such that

$$X(u) \cap V \subseteq X(\bar{u}) + \kappa \|u - \bar{u}\| \mathbb{B} \quad \forall\, u \in U.$$

- The **strong regularity** of the KKT solution S_{KKT} for the general non-convex MOPs, i.e.,

S_{KKT} is a local Lipschitz continuous function.

Two perturbation results

The calmness of X

The strong regularity of the general MOP

The convex MOPs

Recall the following quadratic MOP and its dual:

$$(P)\quad \min \quad \tfrac{1}{2}\|\mathcal{F}X - d\|^2 + \langle C, X\rangle + \theta(X)$$
$$\text{s.t.} \quad \mathcal{A}X = b,\ X \in \mathcal{Q}$$

$$(D)\quad \min \quad \tfrac{1}{2}\|w\|^2 - \langle w, d\rangle - \langle b, y\rangle + \theta^*(-S) + \delta^*_{\mathcal{Q}}(-Z)$$
$$\text{s.t.} \quad \mathcal{F}^* w + \mathcal{A}^* y + S + Z = C$$

The convex MOPs

Recall the following quadratic MOP and its dual:

$$
\begin{aligned}
(P) \quad \min \quad & \tfrac{1}{2}\|\mathcal{F}X - d\|^2 + \langle C, X\rangle + \theta(X) \\
\text{s.t.} \quad & \mathcal{A}X = b,\ X \in \mathcal{Q}
\end{aligned}
$$

$$
\begin{aligned}
(D) \quad \min \quad & \tfrac{1}{2}\|w\|^2 - \langle w, d\rangle - \langle b, y\rangle + \theta^*(-S) + \delta_{\mathcal{Q}}^*(-Z) \\
\text{s.t.} \quad & \mathcal{F}^*w + \mathcal{A}^*y + S + Z = C
\end{aligned}
$$

The **Lagrange function** of (D):

$$
\begin{aligned}
L(w, y, S, Z, X) := \quad & \tfrac{1}{2}\|w\|^2 - \langle b, y\rangle - \langle d, w\rangle + \theta^*(-S) + \delta_{\mathcal{Q}}^*(-Z) \\
& + \langle X, \mathcal{F}^*w + \mathcal{A}^*y + S + Z - C\rangle,
\end{aligned}
$$

The convex MOPs

Recall the following quadratic MOP and its dual:

$$(P) \quad \min \quad \frac{1}{2}\|\mathcal{F}X - d\|^2 + \langle C, X\rangle + \theta(X)$$
$$\text{s.t.} \quad \mathcal{A}X = b,\ X \in \mathcal{Q}$$

$$(D) \quad \min \quad \frac{1}{2}\|w\|^2 - \langle w, d\rangle - \langle b, y\rangle + \theta^*(-S) + \delta^*_{\mathcal{Q}}(-Z)$$
$$\text{s.t.} \quad \mathcal{F}^* w + \mathcal{A}^* y + S + Z = C$$

The **Lagrange function** of (D):

$$L(w, y, S, Z, X) := \quad \frac{1}{2}\|w\|^2 - \langle b, y\rangle - \langle d, w\rangle + \theta^*(-S) + \delta^*_{\mathcal{Q}}(-Z) + \langle X, \mathcal{F}^* w + \mathcal{A}^* y + S + Z - C\rangle,$$

The **essential objective function** of (P)

$F(X) := \inf \left\{ L(w, y, S, Z, X) \mid (w, y, S, Z) \in \Re^d \times \Re^e \times \Re^{m\times n} \times \Re^{m\times n} \right\}$

$\mathcal{T}_{\mathbf{F}} := \partial F$ — a maximal monotone operator

Optimal solution set of (P): $X(0) = \{X \in \mathcal{X} \mid 0 \in \mathcal{T}_F(X)\}$

The calmness of X

Theorem (sufficient conditions for calmness)

$X = \mathcal{T}_F^{-1}$ is **calm** *at 0 for $\overline{X}$ if one of the following two conditions holds:*

The calmness of X

Theorem (sufficient conditions for calmness)

$X = \mathcal{T}_F^{-1}$ is **calm** *at 0 for $\overline{X}$ if one of the following two conditions holds:*

(i) the function θ is a $\mathcal{C}^2$-cone reducible function and the second order sufficient condition of (P) holds at $\overline{X}$.

The calmness of X

Theorem (sufficient conditions for calmness)

*$X = \mathcal{T}_F^{-1}$ is **calm** at 0 for $\overline{X}$ if one of the following two conditions holds:*

(i) the function θ is a $\mathcal{C}^2$-cone reducible function and the second order sufficient condition of (P) holds at $\overline{X}$.

*(ii) for any $v \in \partial\theta(X)$, $(\partial\theta)^{-1}$ is **calm** at v for X and there exists an optimal solution $\widehat{X}$ and Lagrange multipliers $(\hat{w}, \hat{y}, \widehat{S}, \widehat{Z})$ such that the strict complementarity holds at $(\widehat{X}, \hat{w}, \hat{y}, \widehat{S}, \widehat{Z})$.*

The calmness of $(\partial\theta)^{-1}$

Recall $\theta(X) = \phi(\sigma(X)) \; \forall \; X \in \mathbb{R}^{m\times n}$.

Theorem

Let $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ be an **absolutely symmetric function**. *For any $(\overline{X}, \overline{W}) \in$ gph $\partial\theta$, if the inverse of the subdifferential mapping $(\partial\phi)^{-1}$ is calm at $\sigma(\overline{W})$ for $\sigma(\overline{X})$, then $(\partial\theta)^{-1}$ is calm at $\overline{W}$ for $\overline{X}$.*

The calmness of $(\partial\theta)^{-1}$

Recall $\theta(X) = \phi(\sigma(X)) \; \forall \, X \in \mathbb{R}^{m\times n}$.

Theorem

Let $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ be an **absolutely symmetric function**. *For any $(\overline{X}, \overline{W}) \in gph \, \partial\theta$, if the inverse of the subdifferential mapping $(\partial\phi)^{-1}$ is calm at $\sigma(\overline{W})$ for $\sigma(\overline{X})$, then $(\partial\theta)^{-1}$ is calm at $\overline{W}$ for $\overline{X}$.*

- $\partial\phi$ is a **polyhedral mapping**, **S.M. Robinson** (1981)

The calmness of $(\partial\theta)^{-1}$

Recall $\theta(X) = \phi(\sigma(X)) \ \forall\ X \in \mathbb{R}^{m\times n}$.

Theorem

Let $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ be an **absolutely symmetric function**. *For any $(\overline{X}, \overline{W}) \in gph\, \partial\theta$, if the inverse of the subdifferential mapping $(\partial\phi)^{-1}$ is calm at $\sigma(\overline{W})$ for $\sigma(\overline{X})$, then $(\partial\theta)^{-1}$ is calm at $\overline{W}$ for $\overline{X}$.*

- $\partial\phi$ is a **polyhedral mapping**, **S.M. Robinson** (1981)
- ϕ is a **piecewise linear-quadratic function**, **J. Sun**'s Ph.D thesis (1986)

  e.g., $\delta_{\mathcal{S}^n_+}$; $\delta_{\rm EDM}$; $\|\cdot\|_2$; $\|\cdot\|_*$; $\|\cdot\|_{(k)}$; ...

Two perturbation results

The calmness of X

The strong regularity of the general MOP

The general MOPs

MOP:

$$\begin{aligned} \min_{x \in \mathbb{X}} \quad & f(x) + \theta(g(x)) \\ \text{s.t.} \quad & h(x) \in \mathcal{Q} \end{aligned}$$

The general MOPs

MOP:

$$\begin{array}{ll} \min_{x \in \mathbb{X}} & f(x) + \theta(g(x)) \\ \text{s.t.} & h(x) \in \mathcal{Q} \end{array}$$

- $\theta = \phi \circ \sigma$: a function defined on the finite dimensional matrix space

The general MOPs

MOP:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{X}} & f(x)+\theta(g(x)) \\ \text{s.t.} & h(x)\in\mathcal{Q} \end{array}$$

- $\theta = \phi \circ \sigma$: a function defined on the finite dimensional matrix space
- $\phi : \mathbb{R}^m \to (-\infty, \infty]$ is a convex piecewise linear function:

$$\phi(x) = \phi_1(x) + \phi_2(x), \quad x \in \mathbb{R}^m$$

with

$$\phi_1(x) := \max\{\langle \eta_1, x\rangle - \xi_1, \ldots, \langle \eta_l, x\rangle - \xi_l\}$$

and $\phi_2(x) := \delta_{\mathrm{dom}\ \phi}(x)$ where

$$\mathrm{dom}\ \phi := \{x \mid \langle \gamma_i, x\rangle - \iota_i \le 0,\ i = 1, \ldots, r\}$$

Second-order optimality condition

Let $\bar{x}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$.

Second-order optimality condition

Let $\bar{x}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$.

The "no-gap" **(strong) second order sufficient condition** holds at $\overline{X}$, if for any $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ ($d \in \widehat{\mathcal{C}}(\bar{x}) \setminus \{0\}$),

$$\sup_{(\bar{y},\overline{S}) \in \mathcal{M}(\bar{x})} \left\{ \langle d, L''_{xx}(\bar{x}, \bar{y}, \overline{S}) d \rangle - \Upsilon_{g(\bar{x})} \left(\overline{S}, g'(\bar{x}) d \right) \right\} > 0$$

Second-order optimality condition

Let $\bar{x}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$.

The "no-gap" **(strong) second order sufficient condition** holds at $\overline{X}$, if for any $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ ($d \in \widehat{\mathcal{C}}(\bar{x}) \setminus \{0\}$),

$$\sup_{(\bar{y},\overline{S}) \in \mathcal{M}(\bar{x})} \left\{ \langle d, L''_{xx}(\bar{x}, \bar{y}, \overline{S}) d \rangle - \Upsilon_{g(\bar{x})} \left(\overline{S}, g'(\bar{x}) d \right) \right\} > 0$$

- $\mathcal{C}(\bar{x})$: the **critical cone**

Second-order optimality condition

Let $\bar{x}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$.

The "no-gap" **(strong) second order sufficient condition** holds at $\overline{X}$, if for any $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ ($d \in \widehat{\mathcal{C}}(\bar{x}) \setminus \{0\}$),

$$\sup_{(\bar{y},\overline{S})\in\mathcal{M}(\bar{x})} \left\{ \langle d, L''_{xx}(\bar{x},\bar{y},\overline{S})d\rangle - \Upsilon_{g(\bar{x})}\left(\overline{S}, g'(\bar{x})d\right) \right\} > 0$$

- $\mathcal{C}(\bar{x})$: the **critical cone**
- $\widehat{\mathcal{C}}(\bar{x}) := \bigcap_{(y,S)\in\mathcal{M}(\bar{x})} \mathrm{app}(y,S)$ with $\mathrm{app}(y,S)$ is an **approximation** set of $\mathrm{aff}(\mathcal{C}(\bar{x}))$

Second-order optimality condition

Let $\bar{x}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$.

The "no-gap" **(strong) second order sufficient condition** holds at $\overline{X}$, if for any $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ ($d \in \widehat{\mathcal{C}}(\bar{x}) \setminus \{0\}$),

$$\sup_{(\bar{y},\overline{S}) \in \mathcal{M}(\bar{x})} \left\{ \langle d, L''_{xx}(\bar{x}, \bar{y}, \overline{S})d \rangle - \Upsilon_{g(\bar{x})}\left(\overline{S}, g'(\bar{x})d\right) \right\} > 0$$

- $\mathcal{C}(\bar{x})$: the **critical cone**
- $\widehat{\mathcal{C}}(\bar{x}) := \bigcap_{(y,S)\in\mathcal{M}(\bar{x})} \mathrm{app}(y, S)$ with $\mathrm{app}(y, S)$ is an **approximation** set of $\mathrm{aff}(\mathcal{C}(\bar{x}))$
- $\Upsilon_{g(\bar{x})}\left(\overline{S}, g'(\bar{x})d\right)$: the so-called "**$\sigma$-term**"

Constraint nondegeneracy

The **constraint nondegeneracy** for the MOP:

$$\left[\begin{array}{c} h'(\bar{x}) \\ g'(\bar{x}) \end{array}\right] \mathbb{X} + \left[\begin{array}{c} \mathrm{lin}\left(\mathcal{T}_{\mathcal{Q}}(h(\bar{x}))\right) \\ \mathcal{T}^{\mathrm{lin}}(g(\bar{x})) \end{array}\right] = \left[\begin{array}{c} \mathbb{R}^d \\ \mathbb{V} \end{array}\right],$$

where the linear space $\mathcal{T}^{\mathrm{lin}}$ is called the linearity space of the θ defined by

$$\mathcal{T}^{\mathrm{lin}}(Z) := \{H \mid \theta'(Z; H) = -\theta'(Z; -H)\}$$

Constraint nondegeneracy

The **constraint nondegeneracy** for the MOP:

$$\begin{bmatrix} h'(\bar{x}) \\ g'(\bar{x}) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathrm{lin}\left(\mathcal{T}_{\mathcal{Q}}(h(\bar{x}))\right) \\ \mathcal{T}^{\mathrm{lin}}(g(\bar{x})) \end{bmatrix} = \begin{bmatrix} \mathbb{R}^d \\ \mathbb{V} \end{bmatrix},$$

where the linear space $\mathcal{T}^{\mathrm{lin}}$ is called the linearity space of the θ defined by

$$\mathcal{T}^{\mathrm{lin}}(Z) := \{H \mid \theta'(Z;H) = -\theta'(Z;-H)\}$$

Most importantly, the **explicit formulas** for above are now available for the case ϕ is a convex piecewise linear function **Cui and Ding (2018)**.

Characterization of the strong regularity for MOPs

Theorem

Let $\bar{x} \in \mathbb{X}$ be a feasible solution to the MOP with $\mathcal{M}(\bar{x}) \neq \emptyset$. Suppose that the Robinson constraint qualification (RCQ) holds at $\bar{x}$ and $(\bar{y}, \overline{S}) \in \mathcal{M}(\bar{x})$. Then, the following statements are equivalent:

(i) *The strong second order sufficient condition and constraint nondegeneracy for the MOP both hold at $\bar{x}$;*

(ii) $(\bar{x}, \bar{y}, \overline{S})$ *is a strongly regular solution of the KKT equations.*

Conclusions

MOPs in Data Science

Conclusions

MOPs in Data Science

ALM for MOPs

Conclusions

MOPs in Data Science

ALM for MOPs

Sufficient conditions for the **calmness** of X

Conclusions

MOPs in Data Science

ALM for MOPs

Sufficient conditions for the **calmness** of X

Characterization of the **strong regularity** of the general MOP

Thank you!