Local convergence analysis of augmented Lagrangian method for nonlinear semidefinite programming

丁超
应用数学研究所
中国科学院数学与系统科学研究院

中国运筹学会数学规划分会第十三届数学优化大会会议，青岛
2021年10月18日

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

中国科学院
CHINESE ACADEMY OF SCIENCES

Acknowledgements

Based on the joint work with **Shiwei Wang** @ *Institute of Applied Mathematics, AMSS*

♣ S.W. Wang and D., *Local convergence analysis of augmented Lagrangian method for nonlinear semidefinite programming*, arXiv: 2110.10594, 2021.

Outline

Augmented Lagrangian method

Convergence analysis for ALM

Main results

Augmented Lagrangian method

Augmented Lagrangian function

Consider

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

Augmented Lagrangian function

Consider

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

Augmented Lagrangian function[1]:

$$L_\sigma(x; y) := \Phi(x) + \langle y, h(x) \rangle + \frac{\sigma}{2} \|h(x)\|^2$$

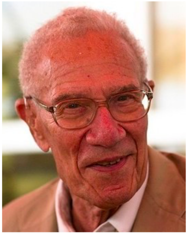
where $\sigma > 0$

[1] Arrow, K.J., Solow, R.M.: Gradient methods for constrained maxima with weakened assumptions. In: Arrow, K.J., Hurwicz, L., Uzawa, H., (eds.) Studies in Linear and Nonlinear Programming. Stanford University Press, Stanford, pp. 165-176 (1958)

K. Arrow and R. Solow

Kenneth Joseph “Ken” Arrow
(23 August 1921 – 21 February 2017)

John Bates Clark Medal (1957); Nobel Prize in Economics (1972); von Neumann Theory Prize (1986); National Medal of Science (2004); ForMemRS (2006)



Robert Merton Solow
(August 23, 1924 –)

John Bates Clark Medal (1961); Nobel Memorial Prize in Economic Sciences (1987); National Medal of Science (1999); Presidential Medal of Freedom (2014); ForMemRS (2006)

Augmented Lagrangian method[2]

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

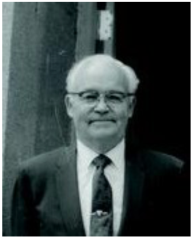
$$\begin{cases} x^{k+1} \approx \operatorname{argmin}\left\{ L_\sigma(x; y^k) \right\} \\ y^{k+1} = y^k + \sigma h(x^{k+1}) \end{cases}$$

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

Augmented Lagrangian method[2]

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

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



Magnus Rudolph Hestenes

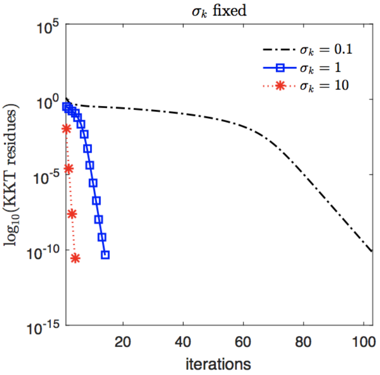
(February 13 1906 – May 31 1991)

Michael James David Powell

(29 July 1936 – 19 April 2015)

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

In particular, for SDP:



Source: Cui, Sun and Toh, MP 19'[3]

[3] Y. Cui, D.F. Sun, K.C. Toh, "On the R-superlinear convergence of the KKT residuals generated by the augmented Lagrangian method for convex composite conic programming", Mathematical Programming 178 (2019) 381–415

ALMs in practice (cont'd)

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

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

ALMs in practice (cont'd)

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

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

ALMs in practice (cont'd)

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

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

Powell, 69'[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

[4] M.J.D. Powell. A method for nonlinear constraints in minimization problems. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

ALMs in practice (cont'd)

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

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

Powell, 69'[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

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

[4] M.J.D. Powell. A method for nonlinear constraints in minimization problems. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

ALMs in practice (cont'd)

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

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

Powell, 69'[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

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

Newton method for $\Psi(y) := \phi(x(y,\sigma)) = 0$:

$$y \leftarrow y - J\Psi(y)^{-1}\Psi(y) = y - J\Psi(y)^{-1}\phi(x(y,\sigma))$$

[4] M.J.D. Powell. A method for nonlinear constraints in minimization problems. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

ALMs in practice (cont'd)

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

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

Powell, 69'[4]: **ALM $\Longleftrightarrow$ Approximate Newton method**

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

Newton method for $\Psi(y) := \phi(x(y,\sigma)) = 0$:

$$y \leftarrow y - J\Psi(y)^{-1}\Psi(y) = y - J\Psi(y)^{-1}\phi(x(y,\sigma))$$

Powell, 69' shows that the Jacobian of Ψ satisfies

$$\|-J\Psi - I\| = O(\frac{1}{\sigma - c})$$

[4] M.J.D. Powell. A method for nonlinear constraints in minimization problems. in *Optimization*, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298

Convergence analysis for ALM

Local convergence rate of ALM (I)

Convex OPs:

- **Rockafellar, 76'**:

  Dual upper Lipschitz continuity + dual boundedness + stopping criteria

  $\Longrightarrow$ dual Q-linear

Local convergence rate of ALM (I)

Convex OPs:

- **Rockafellar, 76'**:

  Dual upper Lipschitz continuity + dual boundedness + stopping criteria

  $\Longrightarrow$ dual Q-linear

- **Luque, 84'**:

  An error bound condition + stopping criteria $\Longrightarrow$ dual Q-linear

Local convergence rate of ALM (I)

Convex OPs:

- **Rockafellar, 76'**:

  Dual upper Lipschitz continuity + dual boundedness + stopping criteria

  $\Longrightarrow$ dual Q-linear

- **Luque, 84'**:

  An error bound condition + stopping criteria $\Longrightarrow$ dual Q-linear

- Cui, Sun and Toh, **MP** 19':

  Dual calmness + stopping criteria

  $\Longrightarrow$ dual Q-(asymptotically super-)linear

  + KKT residual R-(asymptotically super-)linear

Local convergence rate of ALM (I)

Convex OPs:

- **Rockafellar, 76'**:

  Dual upper Lipschitz continuity $+$ dual boundedness $+$ stopping criteria

  $\Longrightarrow$ dual Q-linear

- **Luque, 84'**:

  An error bound condition $+$ stopping criteria $\Longrightarrow$ dual Q-linear

- Cui, Sun and Toh, **MP** 19':

  Dual calmness $+$ stopping criteria

  $\Longrightarrow$ dual Q-(asymptotically super-)linear

  $+$ KKT residual R-(asymptotically super-)linear

Key fact: the strong connection with dual PPA

Local convergence rate of ALM (II)

NLP (non-convex):

- cf. e.g., Bertsekas, 82'; Nocedal and Wright, 06':

  SOSC + **LICQ** + **strict complementarity** $\Longrightarrow$ primal R-linear+dual Q-linear

Local convergence rate of ALM (II)

NLP (non-convex):

- cf. e.g., Bertsekas, 82'; Nocedal and Wright, 06':

  SOSC + **LICQ** + **strict complementarity** $\Longrightarrow$ primal R-linear+dual Q-linear

- Conn, Gould and Toint, 91'; Contesse-Becker 93'; Ito & Kunisch 91':

  SOSC + **LICQ** $\Longrightarrow$ primal R-linear + dual Q-linear

Local convergence rate of ALM (II)

NLP (non-convex):

- cf. e.g., Bertsekas, 82'; Nocedal and Wright, 06':

  SOSC $+$ **LICQ** $+$ **strict complementarity** $\Longrightarrow$ primal R-linear+dual Q-linear

- Conn, Gould and Toint, 91'; Contesse-Becker 93'; Ito & Kunisch 91':

  SOSC $+$ **LICQ** $\Longrightarrow$ primal R-linear $+$ dual Q-linear

- Fernández and Solodov, **SIOPT** 12':

  SOSC $+$ **initial multiplier sufficiently close** $\Longrightarrow$ primal-dual linear

  Key fact: For NLP, automatically,

  SOSC $\Longrightarrow$ a primal-dual error bound (Hoffman's error bound)

Local convergence rate of ALMs (III)

Non-polyhedral & non-convex:

- NLSDP (Sun, Sun and Zhang, **MP** 08'):

  strong SOSC + **LICQ** + initial sufficiently close $\Longrightarrow$ primal-dual linear

Local convergence rate of ALMs (III)

Non-polyhedral & non-convex:

- NLSDP (Sun, Sun and Zhang, **MP** 08'):

  strong SOSC + **LICQ** + initial sufficiently close $\Longrightarrow$ primal-dual linear

- C^2-cone reducible problems (Kanzow and Steck, **MP** 18'):

  $$\text{SOSC} + \textbf{SRCQ} \quad \Longrightarrow \quad \text{primal-dual linear}$$

  Key fact: (**D.**, Sun and Zhang, **SIOPT** 17')

  $$\text{SOSC} + \text{SRCQ} \iff \text{robust isolated calmness}$$

Local convergence rate of ALMs (III)

Non-polyhedral & non-convex:

- NLSDP (Sun, Sun and Zhang, **MP** 08'):

  strong SOSC + **LICQ** + initial sufficiently close $\Longrightarrow$ primal-dual linear

- C^2-cone reducible problems (Kanzow and Steck, **MP** 18'):

  SOSC + **SRCQ** $\quad\Longrightarrow\quad$ primal-dual linear

  Key fact: (**D.**, Sun and Zhang, **SIOPT** 17')

  SOSC + SRCQ $\Longleftrightarrow$ robust isolated calmness

For **non-convex** and **non-polyhedral** cases (MOPs),

- all results are obtained under the solution **uniqueness** assumption
- unlike the polyhedral case (NLPs), Hoffman's error bound does not hold in general

Local convergence rate of ALM (IV)

- Piecewise linear quadratic (Hang & Sarabi, **SIOPT** 21'):

  SOSC + **initial multiplier sufficiently close** $\Longrightarrow$ primal-dual linear

Local convergence rate of ALM (IV)

- Piecewise linear quadratic (Hang & Sarabi, **SIOPT** 21'):

  SOSC + **initial multiplier sufficiently close** $\Longrightarrow$ primal-dual linear

- Second order cone (SOC) (Hang, Mordukhovich and Sarabi, **J. Global Opt.** 21'):

  SOSC + **semi-isolated calm** + initial sufficiently close
  + **multiplier unique** $\Longrightarrow$ primal-dual linear

Local convergence rate of ALM (IV)

- Piecewise linear quadratic (Hang & Sarabi, **SIOPT** 21'):

  SOSC + **initial multiplier sufficiently close** $\Longrightarrow$ primal-dual linear

- Second order cone (SOC) (Hang, Mordukhovich and Sarabi, **J. Global Opt.** 21'):

  SOSC + **semi-isolated calm** + initial sufficiently close
  + **multiplier unique** $\Longrightarrow$ primal-dual linear

- Fully amenable (Rockafellar, **MP** 21'): Piecewise linear quadratic & SOC

  strong variational sufficiency $\Longrightarrow$ primal R-linear + dual Q-linear

Main results

Local convergence analysis of ALM for NLSDP

NLSDP:

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & h(x) = 0, & \leftarrow \quad y \\ & G(x) \in \mathcal{S}^n_+, & \leftarrow \quad \Gamma \end{array}$$

Let $\lambda = (y, \Gamma)$, $\Phi(x) = (h(x), G(x))$ and $\mathcal{K} = \{0\} \times \mathcal{S}^n_+$.

Augmented Lagrangian function:

$$\mathcal{L}(x, \lambda, \rho) := f(x) + \frac{\rho}{2}\text{dist}^2(\Phi(x) + \frac{\lambda}{\rho}, \mathcal{K}) - \frac{\|\lambda\|^2}{2\rho}$$

KKT:

$$S_{KKT}(a_1, a_2, b) = \left\{ (x, y, \Gamma) \in \mathcal{X} \times \Re^e \times \mathcal{S}^n : \begin{array}{l} \nabla_x L(x, y, \Gamma) - a_1 = 0, \\ h(x) - a_2 = 0, \\ 0 \preceq (G(x) - b) \perp \Gamma \preceq 0. \end{array} \right\}$$

For a stationary point $\bar{x}$, the set of multipliers satisfying KKT system:

$$\mathcal{M}(\bar{x}) = \{(y, \Gamma) \in \Re^e \times \mathcal{S}^n \mid (\bar{x}, y, \Gamma) \in S_{\text{KKT}}(0, 0, 0)\}.$$

Algorithm 1

Input: Let $(x^0, \lambda^0) \in \mathcal{X} \times \mathcal{H}$, $\rho_0 > 0$, $\varsigma > 1$, $\xi \in (0,1)$, $\{\epsilon_k\}_{k \geq 0}$ with $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$ and set $k := 0$.

Output: x, λ

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

2: Compute $x^{k+1} \approx \mathrm{argmin}\{\mathcal{L}(x, \lambda^k, \rho^k)\}$ such that $\|\nabla_x \mathcal{L}(\cdot, \lambda^k, \rho^k)\| \leq \epsilon_k$.

3: Update the vector of multipliers to

$$\lambda^{k+1} := \rho^k \big[\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k})\big].$$

Update ρ^{k+1} by $\rho^{k+1} = \rho^k$ or $\rho^{k+1} = \varsigma\rho^k$ according to certain rules.

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

An assumption

Assumption

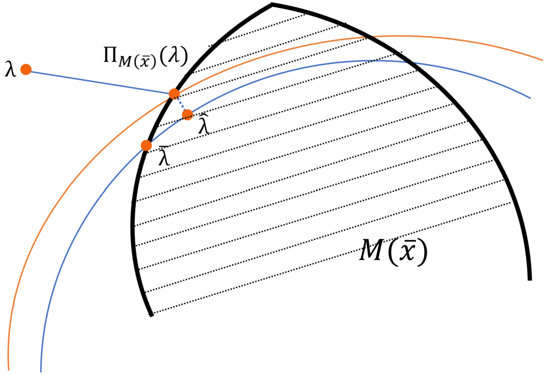
For all $\lambda = (y, \Gamma) \notin \mathcal{M}(\bar{x})$ *sufficiently close to* $\bar{\lambda} \in \operatorname{rbd} \mathcal{M}(\bar{x})$ *and* x *sufficiently close to* $\bar{x}$*, there also exists* $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ *with* $\pi(\widehat{\Gamma}) = \pi(\overline{\Gamma})$ *such that*

$$\|\Pi_{\mathcal{M}(\bar{x})}(\lambda) - \widehat{\lambda}\| = O(R(x, \lambda)),$$

where $R(x, \lambda)$ *is the residual function defined by*

$$R(x, \lambda) = \|\nabla_x L(x, \lambda)\| + \|\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \lambda)\|.$$

An assumption (cont'd)



Main result

Theorem

Let $\bar{x} \in \mathcal{X}$ be a stationary point to NLSDP and $\bar{\lambda} \in \mathcal{M}(\bar{x})$. Suppose SOSC holds at $(\bar{x}, \bar{\lambda})$ and semi-isolated calmness of S_{KKT} holds at $(0, (\bar{x}, \bar{\lambda}))$.

(i) *If $\bar{\lambda} \in \mathrm{ri}\, \mathcal{M}(\bar{x})$, then there exist positive constants $\overline{r}$, $\overline{\zeta}$, $\overline{\varrho}$ such that for any starting point $(x^0, \lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x}, \bar{\lambda})$ the $\{(x^k, \lambda^k)\}_{k \geq 0}$ generated by ALM with $\rho^k \geq \overline{\varrho}$ and $\epsilon_k = o(R(x^k, \lambda^k))$ for all k satisfies*

$$\|x^{k+1} - x^k\| + \|\lambda^{k+1} - \lambda^k\| \leq \overline{\zeta} R(x^k, \lambda^k).$$

(ii) *If $\bar{\lambda} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$ and Assumption holds, then there exist positive constants $\overline{r}$, $\overline{\zeta}$, $\overline{\varrho}$ such that for any starting point $(x^0, \lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x}, \bar{\lambda})$ the $\{(x^k, \lambda^k)\}_{k \geq 0}$ generated by ALM with $\rho^k \geq \overline{\varrho}$ and $\epsilon_k = o(R(x^k, \lambda^k))$ and $\lambda^k \notin \mathcal{M}(\bar{x})$ for all k also satisfies the above inequality.*

Moreover, for each case, the sequence is convergent to $(\bar{x}, \widehat{\lambda})$ for some $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ and its rate of convergence is linear, i.e., for k sufficiently large,

$$\|(x^{k+1}, \lambda^{k+1}) - (\bar{x}, \widehat{\lambda})\| \leq \tau^k \|(x^k, \lambda^k) - (\bar{x}, \widehat{\lambda})\|,$$

where $\tau^k = 2\sqrt{2\zeta}\kappa_1 \kappa_2^2 (R_k^{-1} \epsilon_k + (\rho^k)^{-1} \overline{\zeta})$.

Spectral operators on matrices

Spectral operators: a class of matrix-valued functions defined on the spectral. (**D.** PhD thesis; **D.**, et. al, **MP** 18' and **SIOPT** 20')

For instance, the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot) : \mathcal{S}^n \to \mathcal{S}^n$ over $\mathcal{S}^n_+$:

$$\Pi_{\mathcal{S}^n_+}(X) = P \begin{bmatrix} (\lambda_1(X))_+ & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & (\lambda_n(X))_+ \end{bmatrix} P^T, \quad X \in \mathcal{S}^n,$$

where $\lambda_1(X) \geq \ldots \geq \lambda_n(X)$ are eigenvalues and P is the corresponding eigenvector matrix, i.e.,

$$X = P \begin{bmatrix} \lambda_1(X) & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_n(X) \end{bmatrix} P^T.$$

Spectral operators: a class of matrix-valued functions defined on the spectral. (**D.** PhD thesis; **D.**, et. al, **MP** 18' and **SIOPT** 20')

For instance, the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot) : \mathcal{S}^n \to \mathcal{S}^n$ over $\mathcal{S}^n_+$:

$$\Pi_{\mathcal{S}^n_+}(X) = P \begin{bmatrix} (\lambda_1(X))_+ & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & (\lambda_n(X))_+ \end{bmatrix} P^T, \quad X \in \mathcal{S}^n,$$

where $\lambda_1(X) \geq \ldots \geq \lambda_n(X)$ are eigenvalues and P is the corresponding eigenvector matrix, i.e.,

$$X = P \begin{bmatrix} \lambda_1(X) & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_n(X) \end{bmatrix} P^T.$$

Let $\pi(X) := \cup_{i=1}^d \alpha^l(X)$ be the partition of eigenvalues $\lambda(X)$ with $\alpha^l(X) := \{i : \lambda_i(X) = v_i(X)\}$, where $v_1(X) > \cdots > v_d(X)$ are different eigenvalues of X.

A refined perturbation analysis of eigenvalue decompositions

Given a fixed $\overline{A} \in \mathcal{S}^n$. Let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For any $H \in \mathcal{S}^n$ and $A \in \mathbb{B}_r(\overline{A})$, let U be an orthogonal matrix such that

$$U^T\big(\Lambda(A) + H\big)U = \Lambda\big(\Lambda(A) + H\big).$$

Lemma

Then, for any $H \to 0$, we have

$$\begin{cases} U_{\bar{\alpha}_k\bar{\alpha}_l} = O(\|H\|), & k, l = 1, \cdots, \bar{d}, k \neq l \\ U_{\bar{\alpha}_k\bar{\alpha}_k}U^T_{\bar{\alpha}_k\bar{\alpha}_k} = I_{|\bar{\alpha}_k|} + O(\|H\|^2), & k = 1, \cdots, \bar{d} \end{cases}$$

Furthermore, for each $k \in \{1, \ldots, \bar{d}\}$, there exists $Q_k \in \mathcal{O}^{|\bar{\alpha}_k|}$ such that

$$U_{\bar{\alpha}_k\bar{\alpha}_k} = Q_k + O(\|H\|^2)$$

and

$$Q_k^T H_{\bar{\alpha}_k\bar{\alpha}_k} Q_k = \Lambda_{\bar{\alpha}_k\bar{\alpha}_k}(\Lambda(X) + H) - Q_k^T \Lambda(A)_{\bar{\alpha}_k\bar{\alpha}_k} Q_k + O(\|H\|^2).$$

The $O(\|H\|)$ and $O(\|H\|^2)$ above are uniform for all $A \in \mathbb{B}_r(\overline{A})$.

Given $\overline{A} \in \mathcal{S}^n$ and let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For any $H \in \mathcal{S}^n$ and $A \in \mathbb{B}_r(\overline{A})$, let U be an orthogonal matrix such that

$$U^T\big(A + H\big)U = \Lambda(A + H).$$

Lemma

For all $l \in \{1, \ldots, \bar{d}\}$, there exist $Q_l \in \mathcal{O}^{|\bar{\alpha}^l|}$ (depends on H) such that for all $H \to 0$,

$$(P^T U)_{\bar{\alpha}_k \bar{\alpha}_l} = \Theta_{kl} \circ (\widetilde{H}_{\bar{\alpha}_k \bar{\alpha}_l} Q_l) + O(\|H\|^2), \quad k \neq l$$

where $O(\|H\|^2)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$,
$(\Theta_{kl})_{ij} = 1/\big((\Lambda(A)_{\bar{\alpha}_l \bar{\alpha}_l})_{ii} - (\Lambda(A)_{\bar{\alpha}_k \bar{\alpha}_k})_{jj}\big)$ and $\widetilde{H} = P^T H P$, $P \in \mathcal{O}^n(A)$.

Proposition

Let $\overline{A} \in \mathcal{S}^n$ be given. The metric projection operator $\Pi_{\mathbb{S}^n_+}(\cdot)$ over $\mathbb{S}^n_+$ is **uniformly** *B-differentiable of order 2 for any $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$, i.e., for $\mathcal{S}^n \ni H \to 0$,*

$$\Pi_{\mathcal{S}^n_+}(A+H) - \Pi_{\mathcal{S}^n_+}(A) - \Pi'_{\mathcal{S}^n_+}(A;H) = O\left(\|H\|^2\right)$$

and $O\left(\|H\|^2\right)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$.

$\star$ In literature, we only know that $\Pi_{\mathcal{S}^n_+}(\cdot)$ is B-differentiable of order 2 (**D.**, et al, **MP** 14').

Uniformly second order expansion for Moreau envelop of SDP

Proposition

Let $\overline{A} \in \mathcal{S}^n$ be given. For any $A \in \mathbb{B}_r(\overline{A})$ with $\pi(A) = \pi(\overline{A})$, we have for all $H \to 0$,

$$e\delta_{\mathcal{S}^n_+}(A+H) - e\delta_{\mathcal{S}^n_+}(A) = \langle \Pi_{\mathcal{S}^n_-}(A), H\rangle + \frac{1}{2} e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)(H) + O(\|H\|^3),$$

where $O(\|H\|^3)$ is **uniform** *for all $A \in \mathbb{B}_r(\overline{A})$ with $\pi(A) = \pi(\overline{A})$ and $\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)$.*

- $\pi(A) = \pi(\overline{A})$ means $\alpha_l(A) = \alpha_l(\overline{A})$ for all $k = 1, \ldots, d$
- (Poliquin & Rockafellar , **SIOPT** 96') generated the non-uniform version of $o(\|H\|^2)$

Uniform quadratic growth condition for AL function

Theorem

Let $\bar{x} \in \mathcal{X}$ be a locally optimal solution to the NLSDP and $\bar{\lambda} \in \mathcal{M}(\bar{x})$.

(i) *If $\bar{\lambda} \in \mathrm{ri}\, \mathcal{M}(\bar{x})$, then the SOSC holds at $(\bar{x}, \bar{\lambda})$ if and only if there are positive constants ρ_3, γ, ε, l such that for all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$ and all $\rho \geq \rho_3$ the uniform quadratic growth condition*

$$\mathcal{L}(x, \lambda, \rho) \geq f(\bar{x}) + l\|x - \bar{x}\|^2 \quad \text{for all } x \in \mathbb{B}_\gamma(\bar{x}) \tag{1}$$

is satisfied.

(ii) *If $\bar{\lambda} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$, then the SOSC holds at $(\bar{x}, \bar{\lambda})$ if and only if there are positive constants ρ_3, γ, ε, l such that (1) holds uniformly for all $\rho \geq \rho_3$ and all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$.*

Uniform quadratic growth condition for AL function

Theorem

Let $\bar{x} \in \mathcal{X}$ be a locally optimal solution to the NLSDP and $\bar{\lambda} \in \mathcal{M}(\bar{x})$.

(i) *If $\bar{\lambda} \in \operatorname{ri} \mathcal{M}(\bar{x})$, then the SOSC holds at $(\bar{x}, \bar{\lambda})$ if and only if there are positive constants ρ_3, γ, ε, l such that for all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$ and all $\rho \geq \rho_3$ the uniform quadratic growth condition*

$$\mathcal{L}(x, \lambda, \rho) \geq f(\bar{x}) + l\|x - \bar{x}\|^2 \quad \text{for all } x \in \mathbb{B}_\gamma(\bar{x}) \tag{1}$$

is satisfied.

(ii) *If $\bar{\lambda} \in \operatorname{rbd} \mathcal{M}(\bar{x})$, then the SOSC holds at $(\bar{x}, \bar{\lambda})$ if and only if there are positive constants ρ_3, γ, ε, l such that (1) holds uniformly for all $\rho \geq \rho_3$ and all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$.*

- While in polyhedral case, uniform quadratic growth condition can be get directly from SOSC

Comparison: Rockafellar's variational sufficiency

Consider a general parametrical optimization

$$\min \varphi(x,u) \quad \text{s.t. } u = 0,$$

where u is a perturbation parameter. Let $\varphi_\rho(x,u) = \varphi(x,u) + \frac{\rho}{2}|u|^2$.

Comparison: Rockafellar's variational sufficiency

Consider a general parametrical optimization

$$\min \varphi(x,u) \quad \text{s.t. } u = 0,$$

where u is a perturbation parameter. Let $\varphi_\rho(x,u) = \varphi(x,u) + \frac{\rho}{2}|u|^2$.

Definition (Rockafellar, MP 21' [5])

The **(strongly) variational sufficiency** for local optimality holds with respect to $(\bar{x},\bar{y}) \in \operatorname{gph}\partial\varphi_\rho$ satisfying the first-order condition if there exists $\rho > 0$ such that is **variationally (strongly) convex** with respect to $((\bar{x},0),(0,\bar{y}))$ in $\operatorname{gph}\partial\varphi_\rho$, i.e., there exist open convex neighborhoods $\mathcal{W}$ of $(\bar{x},0)$ and $\mathcal{Z}$ of $(0,\bar{y})$ such that

there exists a proper lsc **(strongly) convex function** $\psi \le \varphi_\rho$ on $\mathcal{W}$ such that

$$[\mathcal{W}\times\mathcal{Z}) \cap \operatorname{gph}\partial\psi = [\mathcal{W}\times\mathcal{Z}) \cap \operatorname{gph}\partial\varphi_\rho$$

and, for $(x,u;v,y)$ belonging to this common set, $\psi(x,u) = \varphi_\rho(x,u)$.

[5] R.T. Rockafellar. Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality. in *Math. Programming*, to appear.

Comparison: Rockafellar's variational sufficiency (cont'd)

Theorem (Rockafellar, MP 21')

With respect to $\bar{x}$ *and* $\bar{y}$ *satisfying the first-order optimality condition, the* **variational sufficiency** *for local optimality holds if and only if, for* $\rho > 0$ *sufficiently large, there is a closed convex neighborhood* $\mathcal{X} \times \mathcal{Y}$ *of* $(\bar{x}, \bar{y})$ *such that* $\mathcal{L}(x, y, \rho)$ *is* **convex in** $x \in \mathcal{X}$ *when* $y \in \mathcal{Y}$ *as well as concave* $y \in \mathcal{Y}$ *when* $x \in \mathcal{X}$.

Comparison: Rockafellar's variational sufficiency (cont'd)

Theorem (Rockafellar, MP 21')

With respect to $\bar{x}$ and $\bar{y}$ satisfying the first-order optimality condition, the **variational sufficiency** *for local optimality holds if and only if, for $\rho > 0$ sufficiently large, there is a closed convex neighborhood $\mathcal{X} \times \mathcal{Y}$ of $(\bar{x}, \bar{y})$ such that $\mathcal{L}(x, y, \rho)$ is* **convex in $x \in \mathcal{X}$** *when $y \in \mathcal{Y}$ as well as concave $y \in \mathcal{Y}$ when $x \in \mathcal{X}$.*

Theorem (Rockafellar, MP 21')

The **strongly variational sufficiency** *for local optimality implies* **augmented tilt stability**. *It corresponds equally to having the functions $\mathcal{L}(\cdot, y, \rho)$ on $\mathcal{X}$ for $y \in \mathcal{Y}$ be* **strongly convex**, *all with the same modulus of strong convexity.*

Both may fail easily even under SOSC!

The proof is in the pudding

Proposition

Let $(\bar{x}, \bar{\lambda})$ be a KKT point satisfying SOSC. Define

$$\mathcal{S}_\rho(\lambda) = \arg\min\{\mathcal{L}(x, \lambda, \rho) \mid x \in \mathbb{B}_{\widehat{r}}(\bar{x})\}.$$

Then there are positive constants $\tau, \rho_3, \widehat{r} > 0$ such that for every $\rho \geq \rho_3$ and $\lambda \in \mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$, the set $\mathcal{S}_\rho(\lambda)$ satisfies the uniform isolated calmness property, i.e.,

$$\mathcal{S}_\rho(\lambda) \subseteq \{\bar{x}\} + \tau\|\lambda - \bar{\lambda}\|\mathbb{B}$$

and satisfies $\varnothing \neq \mathcal{S}_\rho(\lambda) \subseteq \operatorname{int}\mathbb{B}_{\widehat{r}}(\bar{x})$.

The proof is in the pudding

Proposition

Let $(\bar{x}, \bar{\lambda})$ be a KKT point satisfying SOSC. Define

$$\mathcal{S}_\rho(\lambda) = \arg\min\{\mathcal{L}(x, \lambda, \rho) \mid x \in \mathbb{B}_{\widehat{r}}(\bar{x})\}.$$

Then there are positive constants $\tau, \rho_3, \widehat{r} > 0$ such that for every $\rho \geq \rho_3$ and $\lambda \in \mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$, the set $\mathcal{S}_\rho(\lambda)$ satisfies the uniform isolated calmness property, i.e.,

$$\mathcal{S}_\rho(\lambda) \subseteq \{\bar{x}\} + \tau\|\lambda - \bar{\lambda}\|\mathbb{B}$$

and satisfies $\varnothing \neq \mathcal{S}_\rho(\lambda) \subseteq \operatorname{int} \mathbb{B}_{\widehat{r}}(\bar{x})$.

- (Sun at al., **MP** 08'): nondegeneracy + strong SOSC $\Longrightarrow$ Lipschitz continuity
- (Rockafellar, **MP** 21'):
  strongly variational sufficiency $\Longrightarrow$ augmented tilt stability
  - ⋆ stronger than SOSC
  - ⋆ fully amenable function

  works for NLP, NLSOC; but **NLSDP** is **another kettle of fish**

Proof sketch

A refined perturbation analysis of eigenvalue decompositions

$\Downarrow$

SOSC $+$ **Uniform second Expansion of Moreau envelop**

Proof sketch

A refined perturbation analysis of eigenvalue decompositions

$$\Downarrow$$

SOSC $+$ Uniform second Expansion of Moreau envelop $\Longrightarrow$ Uniform quadratic growth condition

Proof sketch

A refined perturbation analysis of eigenvalue decompositions

$\Downarrow$

SOSC $+$ **Uniform second Expansion of Moreau envelop** $\Longrightarrow$ Uniform quadratic growth condition

$\Downarrow$ $+$

Uniform isolated calmness of AL subproblems

Proof sketch

A refined perturbation analysis of eigenvalue decompositions

$\Downarrow$

SOSC + Uniform second Expansion of Moreau envelop $\Longrightarrow$ Uniform quadratic growth condition

$\Downarrow$ +

Uniform isolated calmness of AL subproblems + Semi-isolated calmness + Assumption

Proof sketch

A refined perturbation analysis of eigenvalue decompositions

⇓

SOSC + Uniform second Expansion of Moreau envelop ⟹ Uniform quadratic growth condition

⇓ (from SOSC)

\+ (below Uniform quadratic growth condition)

Uniform isolated calmness of AL subproblems + Semi-isolated calmness + Assumption

⇓

Local linear convergence of ALM

Semi-isolated calmness

Definition

The semi-isolated clamness for the mapping S_{KKT} at $((0,0,0),(\bar{x},\bar{y},\overline{\Gamma}))$ holds if there exists $\kappa > 0$ and open neighborhoods $\mathbb{V}$ of $(0,0,0)$ and $\mathbb{U}$ of $(\bar{x},\bar{y},\overline{\Gamma})$ such that for all $(x,y,\Gamma) \in S_{\mathrm{KKT}}(a_1,a_2,b) \cap \mathbb{U}$,

$$\|x-\bar{x}\| + \mathrm{dist}((y,\Gamma),\mathcal{M}(\bar{x})) \leq \kappa_1 \|(a_1,a_2,b)\|$$

Sufficient conditions for semi-isolated calmness

Theorem

Let $\bar{x} \in \mathcal{X}$ *be a locally optimal solution to the NLSDP and* $(\bar{y}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$*. Suppose SOSC holds at* $(\bar{x}, \bar{y}, \overline{\Gamma})$ *and*

$$\mathcal{G}_1(\bar{x}) \cap \mathrm{ri}\, \mathcal{G}_2(\bar{x}) \neq \varnothing,$$

where $\mathcal{G}_1(\bar{x}) = \{(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n \mid \nabla f(\bar{x}) + \nabla h(\bar{x})^* y + \nabla G(\bar{x})^* \Gamma = 0\}$ *and* $\mathcal{G}_2(\bar{x}) = \{(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n \mid \Gamma \in \mathcal{N}_{\mathcal{S}^n_+}(G(\bar{x}))\}$*. Then there exist a constant* $\kappa_1 > 0$ *and a neighborhood* $\mathbb{U} := \mathbb{B}_{r_1}(\bar{x}, \bar{y}, \overline{\Gamma})$ *of* $(\bar{x}, \bar{y}, \overline{\Gamma})$ *such that for any* $(a_1, a_2, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$*,*

$$\|x - \bar{x}\| + \mathrm{dist}((y, \Gamma), \mathcal{M}(\bar{x})) \leq \kappa_1 \|(a_1, a_2, b)\| \quad \forall\, (x, y, \Gamma) \in S_{\mathrm{KKT}}(a_1, a_2, b) \cap \mathbb{U}.$$

Sufficient conditions for semi-isolated calmness

Theorem

Let $\bar{x} \in \mathcal{X}$ be a locally optimal solution to the NLSDP and $(\bar{y}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$. Suppose SOSC holds at $(\bar{x}, \bar{y}, \overline{\Gamma})$ and

$$\mathcal{G}_1(\bar{x}) \cap \operatorname{ri} \mathcal{G}_2(\bar{x}) \neq \varnothing,$$

where $\mathcal{G}_1(\bar{x}) = \{(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n \mid \nabla f(\bar{x}) + \nabla h(\bar{x})^* y + \nabla G(\bar{x})^* \Gamma = 0\}$ and $\mathcal{G}_2(\bar{x}) = \{(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n \mid \Gamma \in \mathcal{N}_{\mathcal{S}^n_+}(G(\bar{x}))\}$. Then there exist a constant $\kappa_1 > 0$ and a neighborhood $\mathbb{U} := \mathbb{B}_{r_1}(\bar{x}, \bar{y}, \overline{\Gamma})$ of $(\bar{x}, \bar{y}, \overline{\Gamma})$ such that for any $(a_1, a_2, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$,

$$\|x - \bar{x}\| + \operatorname{dist}((y, \Gamma), \mathcal{M}(\bar{x})) \leq \kappa_1 \|(a_1, a_2, b)\| \quad \forall\, (x, y, \Gamma) \in S_{\mathrm{KKT}}(a_1, a_2, b) \cap \mathbb{U}.$$

Other sufficient conditions:

- SOSC + SRCQ, reduced to isolated calmness
- SSOSC + nondegeneracy, reduced to strong regularity

Assumption: For all $\lambda = (y, \Gamma) \notin \mathcal{M}(\bar{x})$ sufficiently close to $\bar{\lambda} \in \operatorname{rbd} \mathcal{M}(\bar{x})$ and x sufficiently close to $\bar{x}$, there also exists $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ with $\pi(\widehat{\Gamma}) = \pi(\overline{\Gamma})$ such that

$$\|\Pi_{\mathcal{M}(\bar{x})}(\lambda) - \widehat{\lambda}\| = O(R(x, \lambda)).$$

Example 1:

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

The optimal solution is $\bar{x} = 0$ and $\mathcal{M}(\bar{x}) = \{\Gamma \mid \Gamma \in \mathcal{S}^3_-\}$. The SOSC and bounded linear regular hold at $(\bar{x}, \overline{\Gamma})$, trivially.

Pick $\overline{\Gamma} = \operatorname{Diag}(0, -1, -2)$. Then for all $\Gamma \in \mathbb{B}_{\min\{1/2, r_1\}}(\overline{\Gamma}) \backslash \mathcal{M}(\bar{x})$,

$$\Pi_{\mathcal{M}(\bar{x})}(\Gamma) = Q\operatorname{Diag}(\min\{0, \Gamma_1\}, \Gamma_2, \Gamma_3)Q^T,$$

where $Q \in \mathcal{O}^3(\Gamma)$. Let $\widehat{\Gamma} = Q\operatorname{Diag}(0, \Gamma_2, \Gamma_3)Q^T$. By the bounded linear regular property, we have

$$\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| = \operatorname{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma)).$$

Example 2:

$$\begin{array}{ll} \min & \frac{1}{2}x^2 + 2t \\ \text{s.t} & tA - x^2 I_2 \in \mathcal{S}^2_+, \quad \Leftarrow \Gamma \\ & t \geq 0, \qquad\qquad \Leftarrow y \end{array}$$

where $A = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$. This problem possesses the unique optimal solution $(\bar{t}, \bar{x}) = (0, 0)$. The corresponding multiplier is

$$\mathcal{M}(\bar{t}, \bar{x}) = \{(\Gamma, y) \in \mathcal{S}^2_- \times \Re \mid \langle A, -\Gamma \rangle \leq 2\}.$$

We can pick $\overline{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$. It is easy to see that $\begin{bmatrix} -1/2 & 0 \\ 0 & -1/2 \end{bmatrix} \in \mathcal{G}_1(\bar{t}, \bar{x}) \cap \mathrm{ri}\, \mathcal{G}_2(\bar{t}, \bar{x})$ and the SOSC holds at $(\bar{t}, \bar{x}, \overline{\Gamma})$.

For all $\Gamma \in \mathbb{B}_{\min\{r_1, 1/(2\sqrt{10})\}}(\overline{\Gamma})$, we know that $\langle A, -\Gamma \rangle < 2$, which implies that $\Pi_{\mathcal{M}(\bar{t}, \bar{x})}(\Gamma) = \Pi_{\mathcal{S}^2_-}(\Gamma)$. For $\Gamma = Q\mathrm{Diag}(\Gamma_1, \Gamma_2)Q^T$ with $Q \in \mathcal{O}^2(\Gamma)$, let $\widehat{\Gamma} = Q\mathrm{Diag}(0, \Gamma_2)Q^T$. $\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| = \mathrm{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma))$, which verifies Assumption.

Conclusions

In this talk:

- Convergence rate of ALM for NLSDP without the solution uniqueness assumption
- Uniformly second order expansion for Moreau envelop of SDP
- Sufficient conditions for semi-isolated calmness

Conclusions

In this talk:

- Convergence rate of ALM for NLSDP without the solution uniqueness assumption
- Uniformly second order expansion for Moreau envelop of SDP
- Sufficient conditions for semi-isolated calmness

Future work:

- Implementable stopping criterion
- dual Q-linear + primal R-linear

谢谢大家!