Quadratic Convergence of SQP-Like Methods for Convex-Composite Optimization

James V Burke
Mathematics, University of Washington

Joint work with
Abraham Engle, Amazon

Vietnam-USA Joint Mathmatical Meeting
ICISE, Quy Nhon - Vietnam
June 10–13 , 2019

Convex-Composite Optimization

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \qquad (\mathbf{P})$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex
$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth

Convex-Composite Optimization

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \qquad (\mathbf{P})$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — The Model

$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth — The Data

Convex-Composite Optimization

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \quad + g(x) \qquad (\mathbf{P})$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — The Model

$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth — The Data

$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — Regularization

used to induce solution properties

Convex-Composite Optimization

$$\min_{x\in\mathbb{R}^n} f(x) := h(c(x)) \quad + g(x) \qquad \textbf{(P)}$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex
$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth
$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex   Regularization
used to induce solution properties

70's
Fletcher, Powel, Osborne

80-90's
Burke, Ferris, Fletcher, Kawasaki, Masden, Poliquin, Powel, Osborne, Rockafellar, Womersley, Wright, Yuan

Recent (15-19's)
Aravkin, Bell, B, Chang, Cui, Duchi, Davis, Drusvyatskiy, Hoheisel, Hong, Lewis, Ioffe, Mordukhovich, Pang, Ruan

Examples: 70 - 90's

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Examples: 70 - 90's

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility: $c(x) \in C : \quad \min \operatorname{dist}(c(x) \mid C),$

where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and
$\operatorname{dist}(y \mid C) := \inf \{\|y - z\| \mid z \in C\}$.

Examples: 70 - 90's

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility: $c(x) \in C: \quad \min \operatorname{dist}(c(x) \mid C),$

where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and
$\operatorname{dist}(y \mid C) := \inf\{\|y - z\| \mid z \in C\}$.

Exact Penalization: $\min \varphi(x) + \alpha \operatorname{dist}(\hat{c}(x) \mid C)$

Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \alpha \operatorname{dist}(y \mid C)$

Examples: 70 - 90's

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility: $c(x) \in C$: $\quad \min \operatorname{dist}(c(x) \mid C)$,
where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and
$\operatorname{dist}(y \mid C) := \inf \{\|y - z\| \mid z \in C\}$.

Exact Penalization: $\min \varphi(x) + \alpha \operatorname{dist}(\hat{c}(x) \mid C)$
Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \alpha \operatorname{dist}(y \mid C)$

Non-linear programming: $\min \varphi(x) + \delta_C(\hat{c}(x))$.
Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \delta_C(y)$, where
$\delta_C(y) = 0$ if $y \in C$ and $+\infty$ otherwise.

More Recent Examples

Optimal Value Composition:

$$h(c) := \min \left\{ b^\top y \mid Ay \leq c \right\}$$

More Recent Examples

Optimal Value Composition:

$$h(c) := \min \left\{ b^\top y \mid Ay \leq c \right\}$$

Quadratic support functions:

$$h(c) := \sup_{u \in U} \langle u, Bc \rangle - \frac{1}{2} u^T M u$$

with $U \subset \mathbb{R}^k$ non-empty, closed, convex, $M \in \mathbb{S}^n$ is positive semi-definite.

More Recent Examples

Optimal Value Composition:

$$h(c) := \min \left\{ b^\top y \mid Ay \leq c \right\}$$

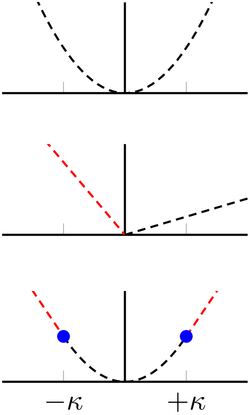
Quadratic support functions:

$$h(c) := \sup_{u \in U} \langle u, Bc \rangle - \frac{1}{2} u^T M u$$

with $U \subset \mathbb{R}^k$ non-empty, closed, convex, $M \in \mathbb{S}^n$ is positive semi-definite.

Piecewise linear-quadratic (PLQ) penalties:
(Rockfellar-Wets (97))
Quadratic support functions with $U \subset \mathbb{R}^k$ non-empty, closed and convex polyhedron.

Dual representation of PLQs

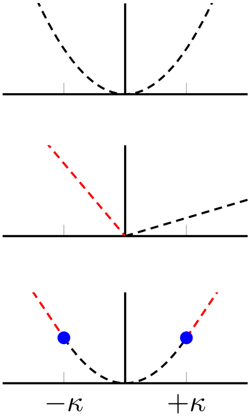


$$\frac{1}{2}x^2 = \sup_{u \in \mathbb{R}} \langle u, x \rangle - \frac{1}{2}u^2$$

$$Q_{0.8}(x) = \sup_{u \in [-0.8, 0.2]} \langle u, x \rangle$$

$$\rho_h(x) = \sup_{u \in [-\kappa, \kappa]} \langle u, x \rangle - \frac{1}{2}u^2$$

Dual representation of PLQs



$$\frac{1}{2}x^2 = \sup_{u \in \mathbb{R}} \langle u, x \rangle - \frac{1}{2}u^2$$

$$Q_{0.8}(x) = \sup_{u \in [-0.8, 0.2]} \langle u, x \rangle$$

$$\rho_h(x) = \sup_{u \in [-\kappa, \kappa]} \langle u, x \rangle - \frac{1}{2}u^2$$

PLQ penalties closed under addition and affine composition.

PLQ penalties in practice

Application	Objective	PLQs
Regression	$\|Ax - b\|^2$	L_2
Robust regression	$\rho_H(Ax - b)$	Huber
Quantile regression	$Q(Ax - b)$	Asym. L_1
Lasso	$\|Ax - b\|^2 + \lambda\|x\|_1$	$L_2 + L_1$
Robust lasso	$\rho_H(Ax - b) + \lambda\|x\|_1$	Huber + L_1
SVM	$\frac{1}{2}\|w\|^2 + H(\mathbf{1} - Ax)$	L_1 + hinge loss
SVR	$\rho_V(Ax - b)$	Vapnik loss
Kalman smoother	$\|Gx - w\|^2_{Q^{-1}} + \|Hx - z\|^2_{R^{-1}}$	$L_2 + L_2$
Robust trend smoothing	$\|Gx - w\|_1 + \rho_H(Hx - z)$	L_1 + Huber

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y)$$

- The conjugate of h given by the support function for epi(h),

$$h^*(y) := \sup_x [\langle y, x\rangle - h(x)]$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x)) \quad + g(x)$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \quad + g(x)$$

- The conjugate of h given by the support function for epi(h),

$$h^*(y) := \sup_x [\langle y, x\rangle - h(x)]$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x \in \mathbb{R}^n} h(c(x)) \quad + g(x)$$

- The Lagrangian for $\mathbf{P}$: (B. (87))

$$L(x, y, v) := \langle y, c(x) \rangle - h^*(y) \quad + \langle v, x \rangle - g^*(v)$$

- The conjugate of h given by the support function for epi(h),

$$h^*(y) := \sup_x [\langle y, x \rangle - h(x)]$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

- The conjugate of h given by the support function for epi(h),

$$h^*(y) := \sup_x [\langle y, x\rangle - h(x)]$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x \in \mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x) \rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

- The conjugate of h given by the support function for epi(h),

$$h^*(y) := \sup_x [\langle y, x \rangle - h(x)] \qquad = \sup_{(x,\mu) \in \text{epi}(h)} \langle (y, -1), (x, \mu) \rangle$$

Algorithms

$$\mathbf{P}_k \qquad \min_x h\left(c(x^k)+\nabla c(x^k)[x-x^k]\right)+\frac{1}{2}(x-x^k)^\top H_k(x-x^k),$$

- H_k approximates the Hessian of a Lagrangian for $\mathbf{P}$ at (x^k, y^k)
- Newton's method: $H_k := \nabla^2_{xx} L(x^k, y^k) = \sum_{k=1}^m y_i^k \nabla^2_{xx} c_i(x^k)$
- $\mathbf{P}_k$ may or may not be convex depending on whether $H_k \succeq 0$.
- A example is the Gauss-Newton method: $h = \|\cdot\|_2^2$

$$\min_x \left\|c(x^k) + c'(x^k)(x - x^k)\right\|_2^2$$

Algorithm for NLP

NLP minimize $\phi(x)$

subject to $f_i(x) = 0,\ i = 1, \ldots, s,\ f_i(x) \leq 0,\ i = s+1, \ldots, m.$

Algorithm for NLP

$$\text{NLP} \quad \text{minimize} \;\; \phi(x)$$
$$\text{subject to} \;\; f_i(x) = 0,\; i = 1, \ldots, s,\;\; f_i(x) \le 0,\; i = s+1, \ldots, m.$$

- Convex-Composite Framework

$$h(\mu, y) = \mu + \delta_K(y), \qquad K := \{0\}^s \times \mathbb{R}_-^{m-s}$$
$$c(x) = (\phi(x), f(x))$$
$$L(x, y) = \phi(x) + \sum_{k=1}^{m} y_i f_i(x) \; - \delta_{K^\circ}(y), \quad K^\circ = \mathbb{R}^s \times \mathbb{R}_+^{m-s}$$

Algorithm for NLP

NLP minimize $\phi(x)$

subject to $f_i(x)=0,\ i=1,\ldots,s,\ f_i(x)\le 0,\ i=s+1,\ldots,m.$

- Convex-Composite Framework

$$h(\mu,y)=\mu+\delta_K(y), \qquad K:=\{0\}^s\times\mathbb{R}^{m-s}_-$$

$$c(x)=(\phi(x),f(x))$$

$$L(x,y)=\phi(x)+\sum_{k=1}^{m} y_i f_i(x)\ -\delta_{K^\circ}(y), \quad K^\circ=\mathbb{R}^s\times\mathbb{R}^{m-s}_+$$

- Subproblems:

$\mathbf{P_k}$ minimize $\phi(x^k)+\nabla\phi(x^k)^T(x-x^k)+\frac{1}{2}[x-x^k]^\top H_k[x-x^k]$

subject to $f_i(x^k)+\nabla f_i(x^k)^T(x-x^k)=0,\ i=1,\ldots,s$

$f_i(x^k)+\nabla f_i(x^k)^T(x-x^k)=0,\ i=s+1,\ldots,m.$

Algorithm for NLP

NLP minimize $\phi(x)$

subject to $f_i(x)=0,\ i=1,\ldots,s,\ f_i(x)\le 0,\ i=s+1,\ldots,m.$

- Convex-Composite Framework

$$h(\mu,y)=\mu+\delta_K(y), \qquad K:=\{0\}^s\times\mathbb{R}_-^{m-s}$$

$$c(x)=(\phi(x),f(x))$$

$$L(x,y)=\phi(x)+\sum_{k=1}^{m} y_i f_i(x)\ -\delta_{K^\circ}(y), \quad K^\circ=\mathbb{R}^s\times\mathbb{R}_+^{m-s}$$

- Subproblems: Sequential quadratic programming (SQP)

$\mathbf{P_k}$ minimize $\phi(x^k)+\nabla\phi(x^k)^T(x-x^k)+\frac{1}{2}[x-x^k]^\top H_k[x-x^k]$

subject to $f_i(x^k)+\nabla f_i(x^k)^T(x-x^k)=0,\ i=1,\ldots,s$

$f_i(x^k)+\nabla f_i(x^k)^T(x-x^k)=0,\ i=s+1,\ldots,m.$

Convergence of Convex-Composite Newton's Method

Robinson (72):

Assumed $h = \delta_K$ with $K := \{0\}^s \times \mathbb{R}^{m-s}_-$ (NLP case).

Established quadratic convergence in the NLP case under linear independence of the active constraint gradients, strict complementarity, and strong second-order sufficiency.

Robinson (80):

Introduced the revolutionary notion of **generalized equations** *which, among many other consequences, re-established quadratic convergence for NLP. The generalized equations approach is much more powerful as it allows access to a very rich sensitivity theory including metric regularity properties of solution mappings.*

Convergence of Convex-Composite Newton's Method

Womersley (85):

Assumed h is finite-valued piecewise linear convex.

> *Established quadratic convergence under NLP-like conditions: linear independence of the active constraint gradients, strict complementarity, and strong second-order sufficiency.*

B-Ferris (95):

Assumed h is finite-valued closed, proper, convex.

> *Established quadratic convergence when* $C := \arg\min h$ *is a set of weak sharp minima for* h*, and* $\arg\min f = \{x \mid c(x) \in C\}$.

Cibulka-Dontchev-Kruger (16):

Assumed h is piecewise linear convex.

> *Established super-linear convergence under the Dennis-Moré conditions using generalized equations.*

The Program

A long standing open problem:

Can one establish second-order rates using the rich history of second-order ideas for convex-composite functions?

(B(87), Kawazaki(88), Ioffe(88), B-Poliquin(92), Rochafellar-Wets(92), Nguyen(17-19))

The Program

A long standing open problem:

Can one establish second-order rates using the rich history of second-order ideas for convex-composite functions?

(B(87), Kawazaki(88), Ioffe(88), B-Poliquin(92), Rochafellar-Wets(92), Nguyen(17-19))

Solution Proposal:

> *Develop a generalized equations approach for the PLQ class using PLQ second-order theory and partial smoothness to establish second-order rates under hypotheses motivated by those used for NLP.*

The Program

A long standing open problem:

Can one establish second-order rates using the rich history of second-order ideas for convex-composite functions?

(B(87), Kawazaki(88), Ioffe(88), B-Poliquin(92), Rochafellar-Wets(92), Nguyen(17-19))

Solution Proposal:

> *Develop a generalized equations approach for the PLQ class using PLQ second-order theory and partial smoothness to establish second-order rates under hypotheses motivated by those used for NLP.*

Key new ingredient is *partial smoothness* due to (Lewis (02)).

PLQ Functions

$h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is called piecewise linear-quadratic (PLQ) if $\operatorname{dom} h \neq \emptyset$ and, for $\mathcal{K} \geq 1$,

$$\operatorname{dom} h = \bigcup_{k=1}^{\mathcal{K}} C_k,$$

where the sets C_k are convex polyhedrons,

$$C_k = \{c \mid \langle a_{kj}, c\rangle \leq \alpha_{kj}, \text{ for all } j \in \{1, \ldots, s_k\}\},$$

and relative to which $h(c)$ is given by an expression of the form

$$h(c) = \frac{1}{2}\langle c, Q_k c\rangle + \langle b_k, c\rangle + \beta_k \quad \forall\, c \in C_k$$

with $\beta_k \in \mathbb{R}$, $b_k \in \mathbb{R}^n$, and $Q_k \in \mathbb{S}^m$.

Variational Analysis of PLQ-Composite Functions

Assume $f := h \circ c$ with h convex PLQ and c in $\mathcal{C}^2(\mathbb{R}^n)$.

Active Set: For $c \in \operatorname{dom} h$, the active set at c is $\mathcal{K}(c) := \{k \mid c \in C_k\}$.

Basic Constraint Qualification: (BCQ)

$$\ker \nabla c(\bar{x})^\top \cap N_{\operatorname{dom} h}(c(\bar{x})) = \{0\}$$

Subdifferential: Under the BCQ

$$\partial f(x) = c'(x)^T \partial h(c(x)).$$

Directional Derivative: Under BCQ

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} = h'(c(x); c'(x)d)$$

with

$$h'(\bar{c}; w) = \langle Q_k \bar{c} + b_k, w \rangle \quad \forall\, k \in \mathcal{K}(\bar{c}) \text{ and } w \in T_{C_k}(\bar{c}).$$

Directions of Non-Ascent and Multipliers

Directions of non-ascent:

$$\begin{aligned} D(x) &:= \left\{ d \in \mathbb{R}^n \;\middle|\; f'(x : d) \leq 0 \right\} \\ &= \left\{ d \in \mathbb{R}^n \;\middle|\; h'(c(x); \nabla c(x) d) \leq 0 \right\} \qquad \text{(BCQ)} \end{aligned}$$

The Multiplier Set:

$$M(\bar{x}) := \ker \nabla c(\bar{x})^\top \cap \partial h(c(\bar{x})) = \left\{ y \;\middle|\; \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial_x L(\bar{x}, y) \\ \partial_y(-L)(\bar{x}, y) \end{pmatrix} \right\}$$

The Second Directional Derivative

The PLQ second directional derivative:
(Rockafellar-Wets (97))

$$0 \leq h''(\bar{c}; w) := \lim_{t \searrow 0} \frac{h(\bar{c} + tw) - h(\bar{c}) - th'(\bar{c}; w)}{\frac{1}{2}t^2}$$
$$= \begin{cases} \langle w, Q_k w \rangle & \text{when } w \in T_{C_k}(\bar{c}), \\ \infty & \text{when } w \notin T_{\operatorname{dom} h}(\bar{c}). \end{cases}$$

and $h''(\bar{c}; \cdot)$ is PLQ, but not necessarily convex.

The Second Directional Derivative

The PLQ second directional derivative:
(Rockafellar-Wets (97))

$$0 \leq h''(\bar{c}; w) := \lim_{t \searrow 0} \frac{h(\bar{c} + tw) - h(\bar{c}) - th'(\bar{c}; w)}{\frac{1}{2}t^2}$$
$$= \begin{cases} \langle w, Q_k w \rangle & \text{when } w \in T_{C_k}(\bar{c}), \\ \infty & \text{when } w \notin T_{\operatorname{dom} h}(\bar{c}). \end{cases}$$

and $h''(\bar{c}; \cdot)$ is PLQ, but not necessarily convex.

Moreover, there exists a neighborhood V of $\bar{c}$ such that

$$h(c) = h(\bar{c}) + h'(\bar{c}; c - \bar{c}) + \frac{1}{2} h''(\bar{c}; c - \bar{c}) \text{ for } c \in V \cap \operatorname{dom} h.$$

PLQ-Composite 2$^{\text{nd}}$-Order Nec. and Suff. Conditions

(Rockafellar-Wets (97))

Let $\bar{x} \in \operatorname{dom} f$ such that f satisfies BCQ at $\bar{x}$.

(1) (Nec.) If f has a local minimum at $\bar{x}$, then
$0 \in \nabla c(\bar{x})^\top \partial h(c(\bar{x}))$ and, $\forall\ d \in D(\bar{x})$,

$$h''(c(\bar{x}); \nabla c(\bar{x})d) + \max\left\{\left\langle d, \nabla^2_{xx} L(\bar{x}, y)d\right\rangle \mid y \in M(\bar{x})\right\} \geq 0\ .$$

(2) (Suff.) If $0 \in \nabla c(\bar{x})^\top \partial h(c(\bar{x}))$ and, $\forall\ d \in D(\bar{x}) \setminus \{0\}$,

$$h''(c(\bar{x}); \nabla c(\bar{x})d) + \max\left\{\left\langle d, \nabla^2_{xx} L(\bar{x}, y)d\right\rangle \mid y \in M(\bar{x})\right\} > 0,$$

then $\bar{x}$ is a strong local minimizer of f,
that is, there exists $\varepsilon > 0, \mu > 0$ such that

$$f(x) \geq f(\bar{x}) + \frac{\mu}{2}\|x - \bar{x}\|_2^2 \quad \forall\ x \in B(\bar{x}, \varepsilon).$$

Convex-Composite Generalized Equations

Let $f := h \circ c$ be convex-composite, and define the set-valued mapping $g + G : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{n+m}$ by

$$g(x,y) = \begin{pmatrix} \nabla c(x)^\top y \\ -c(x) \end{pmatrix}, \quad G(x,y) = \begin{pmatrix} \{0\}^n \\ \partial h^\star(y) \end{pmatrix}.$$

The associated generalized equation for **P** is $g + G \ni 0$.

Convex-Composite Generalized Equations

Let $f := h \circ c$ be convex-composite, and define the set-valued mapping $g + G : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{n+m}$ by

$$g(x,y) = \begin{pmatrix} \nabla c(x)^\top y \\ -c(x) \end{pmatrix}, \quad G(x,y) = \begin{pmatrix} \{0\}^n \\ \partial h^\star(y) \end{pmatrix}.$$

The associated generalized equation for **P** is $g + G \ni 0$.

For a fixed $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, define the linearization mapping

$$\mathcal{G} : (x,y) \mapsto g(\bar{x},\bar{y}) + \nabla g(\bar{x},\bar{y}) \begin{pmatrix} x - \bar{x} \\ y - \bar{y} \end{pmatrix} + G(x,y),$$

where $\nabla g(\bar{x},\bar{y}) = \begin{pmatrix} \nabla^2(\bar{y}c)(\bar{x}) & \nabla c(\bar{x})^\top \\ -\nabla c(\bar{x}) & 0 \end{pmatrix}.$

Newton's Method for Generalized Equations

- Let $f := h \circ c$ be convex-composite.
- For $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ set $\widehat{H} := \nabla^2_{xx} L(\hat{x}, \hat{y})$.
- Assume f satisfies BCQ at $\hat{x}$.

Then, $(\tilde{x}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfy the optimality conditions for

$$\min_{x \in \mathbb{R}^n} h(c(\hat{x}) + \nabla c(\hat{x})(x - \hat{x}) + \frac{1}{2}(x - \hat{x})^\top \widehat{H}(x - \hat{x})$$

if and only if $(\tilde{x}, \tilde{y})$ solves the Newton equations for $g + G$:

$$0 \in g(\hat{x}, \hat{y}) + \nabla g(\hat{x}, \hat{y}) \begin{pmatrix} x - \hat{x} \\ y - \hat{y} \end{pmatrix} + G(x, y).$$

Strong Metric Subregularity

A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically subregular* at $\bar{u}$ for $\bar{v}$ if $(\bar{u}, \bar{v}) \in \text{graph}\,(S)$ and there exists $\kappa \geq 0$ and a neighborhood U of $\bar{u}$ such that

$$\|u - \bar{u}\| \leq \kappa \text{dist}\,(\bar{v} \mid S(u)) \text{ for all } u \in U.$$

Strong Metric Subregularity

A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically subregular* at $\bar{u}$ for $\bar{v}$ if $(\bar{u}, \bar{v}) \in \operatorname{graph}(S)$ and there exists $\kappa \geq 0$ and a neighborhood U of $\bar{u}$ such that

$$\|u - \bar{u}\| \leq \kappa \operatorname{dist}(\bar{v} \mid S(u)) \text{ for all } u \in U.$$

Theorem: (B-Engel(18)) $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ convex PLQ and $f := h \circ c$ satisfies BCQ at $\bar{x} \in \operatorname{dom} f$. Then, the following are equivalent:

(1) The multiplier set $M(\bar{x}) := \ker \nabla c(\bar{x})^\top \cap \partial h(c(\bar{x}))$ is a singleton $\{\bar{y}\}$ and the second-order sufficient conditions are satisfied at $\bar{x}$.

(2) The mapping $g + G$ is strongly metrically subregular at $(\bar{x}, \bar{y})$ for 0 and $\bar{x}$ is a strong local minimizer of f.

Strong Metric Subregularity

A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically subregular* at $\bar{u}$ for $\bar{v}$ if $(\bar{u}, \bar{v}) \in \operatorname{graph}(S)$ and there exists $\kappa \geq 0$ and a neighborhood U of $\bar{u}$ such that

$$\|u - \bar{u}\| \leq \kappa \operatorname{dist}(\bar{v} \mid S(u)) \text{ for all } u \in U.$$

Theorem: (B-Engel(18)) $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ convex PLQ and $f := h \circ c$ satisfies BCQ at $\bar{x} \in \operatorname{dom} f$. Then, the following are equivalent:

(1) The multiplier set $M(\bar{x}) := \ker \nabla c(\bar{x})^\top \cap \partial h(c(\bar{x}))$ is a singleton $\{\bar{y}\}$ and the second-order sufficient conditions are satisfied at $\bar{x}$.

(2) The mapping $g + G$ is strongly metrically subregular at $(\bar{x}, \bar{y})$ for 0 and $\bar{x}$ is a strong local minimizer of f.

Corollary: The matrix secant method converges superlinearly if the Dennis-Móre condition holds.

Partial Smoothness: Lewis (02)

- $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a closed and proper function.
- $\mathcal{M}$ a $\mathcal{C}^2$-smooth manifold and $\bar{c} \in \mathcal{M} \subset \mathbb{R}^m$.

The function h is *partly smooth* at $\bar{c}$ relative to $\mathcal{M}$ if $\mathcal{M}$ the following four properties hold:

(1) (restricted smoothness) the restriction $h|_{\mathcal{M}}$ is smooth around $\bar{c}$, in that there exists a neighborhood V of $\bar{c}$ and a $\mathcal{C}^2$-smooth function g defined on V such that $h = g$ on $V \cap \mathcal{M}$;

(2) (existence of subgradients) at every point $c \in \mathcal{M}$ close to $\bar{c}$, $\partial h(c) \neq \emptyset$;

(3) (normals and subgradients parallel) $\mathrm{par}\partial h(\bar{c}) = N_{\mathcal{M}}(\bar{c})$;

(4) (subgradient inner semicontinuity) the subdifferential map ∂h is inner semicontinuous at $\bar{c}$ relative to $\mathcal{M}$.

Partial Smoothness: Lewis (02)

- $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a closed and proper function.
- $\mathcal{M}$ a $\mathcal{C}^2$-smooth manifold and $\bar{c} \in \mathcal{M} \subset \mathbb{R}^m$.

The function h is *partly smooth* at $\bar{c}$ relative to $\mathcal{M}$ if $\mathcal{M}$ the following four properties hold:

(1) (restricted smoothness) the restriction $h|_{\mathcal{M}}$ is smooth around $\bar{c}$, in that there exists a neighborhood V of $\bar{c}$ and a $\mathcal{C}^2$-smooth function g defined on V such that $h = g$ on $V \cap \mathcal{M}$;

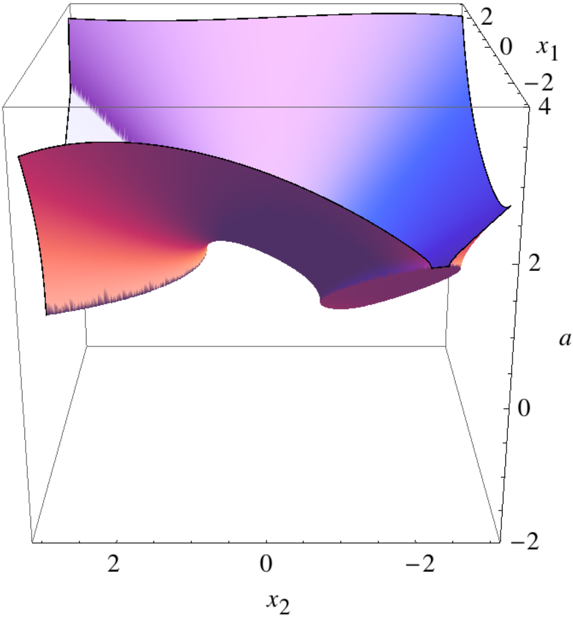
(2) (existence of subgradients) at every point $c \in \mathcal{M}$ close to $\bar{c}$, $\partial h(c) \neq \emptyset$;

(3) (normals and subgradients parallel) $\mathrm{par}\partial h(\bar{c}) = N_{\mathcal{M}}(\bar{c})$;

(4) (subgradient inner semicontinuity) the subdifferential map ∂h is inner semicontinuous at $\bar{c}$ relative to $\mathcal{M}$.

Generalizes classical notions of *nondegeneracy, strict complementarity, and active constraint identification.*

Partial Smoothness



Rockafellar-Wets Representation (RWR)

h is PLQ and $\operatorname{int}(\operatorname{dom} h) \neq \emptyset$. Then, WLOG, the polyhedral sets $\{C_k\}_{k=1}^{\mathcal{K}}$ are given in terms of a common set of $s > 0$ hyperplanes $\mathcal{H} := \{(a_j, \alpha_j)\}_{j=1}^{s} \subset (\mathbb{R}^m \setminus \{0\}) \times \mathbb{R}$, so that $\forall\ k \in \{1, \ldots, \mathcal{K}\}$,

$$C_k = \{c \mid \langle \omega_{kj} a_j, c \rangle \leq \omega_{kj} \alpha_j,\ \text{for all } j \in \{1, \ldots, s\}\},$$

with $\omega_{kj} \in \{\pm 1\}$,

$$I_k(c) = \{j \mid \langle \omega_{kj} a_j, c \rangle = \omega_{kj} \alpha_j\} = \{j \mid \langle a_j, c \rangle = \alpha_j\} \subset \{1, \ldots, s\},$$

and

(i) $\emptyset \neq \operatorname{int}(C_k) = \left\{ c \;\middle|\; \begin{array}{l} \langle \omega_{kj} a_j, c \rangle < \omega_{kj} \alpha_j, \\ \forall\ j \in \{1, \ldots, s_k\} \end{array} \right\}, \ \forall\, k \in \{1, \ldots, \mathcal{K}\},$

(ii) $\operatorname{int}(C_{k_1}) \cap \operatorname{int}(C_{k_2}) = \emptyset$ when $k_1 \neq k_2$.

Condition (b) implies that if $c \in C_{k_1} \cap C_{k_2}$, then $c \in \operatorname{bdry} C_{k_1} \cap \operatorname{bdry} C_{k_2}$ when $k_1 \neq k_2$.

The Active Manifold

- $\mathcal{M}$ Active set: $\mathcal{K}(c) := \{k \in \mathbb{R}^m \mid c \in C_k,\ k \in \{1, 2, \dots, \mathcal{K}\}\}$

- Active Manifold: $\mathcal{M}_{\bar{c}} := \operatorname{ri} \bigcap_{k \in \mathcal{K}(\bar{c})} C_k$

- Active set (RWR) for

$$C_k = \{c \mid \langle \omega_{kj} a_j, c \rangle \leq \omega_{kj} \alpha_j,\ \text{for all } j \in \{1, \dots, s\}\},$$

with $\omega_{kj} \in \{\pm 1\}$, is

$$I_k(c) = \{j \mid \langle \omega_{kj} a_j, c \rangle = \omega_{kj} \alpha_j\} = \{j \mid \langle a_j, c \rangle = \alpha_j\} \subset \{1, \dots, s\}.$$

The Active Manifold

Lemma: Let $\bar{c} \in \operatorname{dom} f$ and assume $\operatorname{dom} h$ is given by an RWR. Then, for all $c \in \mathcal{M}_{\bar{c}}$ and $k \in \mathcal{K}(\bar{c})$,

$$\mathcal{K}(c) = \mathcal{K}(\bar{c}),\ \mathcal{M}_c = \mathcal{M}_{\bar{c}} \text{ and } I_k(c) = I_k(\bar{c}).$$

Moreover,

$$\mathcal{M}_{\bar{c}} = \left\{ c \;\middle|\; \begin{array}{c} \langle c, a_j \rangle = \alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \in I_k(\bar{c}) \\ \langle c, \omega_{kj} a_j \rangle < \omega_{kj} \alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \notin I_k(\bar{c}) \end{array} \right\}$$

The Active Manifold

Lemma: Let $\bar{c} \in \operatorname{dom} f$ and assume $\operatorname{dom} h$ is given by an RWR. Then, for all $c \in \mathcal{M}_{\bar{c}}$ and $k \in \mathcal{K}(\bar{c})$,

$$\mathcal{K}(c) = \mathcal{K}(\bar{c}),\ \mathcal{M}_c = \mathcal{M}_{\bar{c}} \text{ and } I_k(c) = I_k(\bar{c}).$$

Moreover,

$$\mathcal{M}_{\bar{c}} = \left\{ c \;\middle|\; \begin{array}{c} \langle c, a_j \rangle = \alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \in I_k(\bar{c}) \\ \langle c, \omega_{kj} a_j \rangle < \omega_{kj}\alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \notin I_k(\bar{c}) \end{array} \right\}$$

For $k \in \mathcal{M}_{\bar{c}}$ set $A := A_k(\bar{c})$ whose columns are $\{a_j \mid k \in \mathcal{K}(\bar{c}), j \in I_k(\bar{c})\}$.
Then $\exists$ diagonal P_j with entries ± 1 on the diagonal such that

$$AP_j = A_{k_j}(c) \quad \forall\ c \in \mathcal{M}_{\bar{c}},$$

and, for any $k \in \mathcal{K}(\bar{c})$ and $c \in \mathcal{M}_{\bar{c}}$,

$$T_{\mathcal{M}_{\bar{c}}}(c) = \ker A^{\top},\ \text{ and } N_{\mathcal{M}_{\bar{c}}}(c) = \operatorname{Ran}(A).$$

The Subdifferential of h

We let $\bar{k} = |\mathcal{K}(\bar{c})$ and $\ell := |I_k(\bar{c})| = |I_{k'}(\bar{c})|$ for all $k, k' \in \mathcal{K}(\bar{c})$, so that $A \in \mathbb{R}^{m\times\ell}, P_j \in \mathbb{R}^{\ell\times\ell}, P_{\bar{k}} = I_\ell$, and define block matrices $\hat{\mathcal{Q}} := \operatorname{diag}(Q_k), \hat{\mathcal{A}} := \operatorname{diag} AP_j$

$$\mathcal{A} := \begin{pmatrix} (1-\bar{k})AP_1 & AP_2 & \cdots & A \\ AP_1 & (1-\bar{k})AP_2 & \cdots & A \\ \vdots & \ddots & \ddots & \vdots \\ AP_1 & AP_2 & \cdots & (1-\bar{k})A \end{pmatrix},$$

$$\mathcal{Q} := \begin{bmatrix} Q_{k_1} \\ Q_{k_2} \\ \vdots \\ Q_{k_{\bar{k}}} \end{bmatrix}, \ \mathcal{B} := \begin{bmatrix} b_{k_1} \\ b_{k_2} \\ \vdots \\ b_{k_{\bar{k}}} \end{bmatrix}, \ J := \begin{bmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{bmatrix}$$

and averaged quantities

$$\bar{Q} = (1/\bar{k})J^\top \hat{\mathcal{Q}} J, \quad \bar{A} = (1/\bar{k})J^\top \hat{\mathcal{A}}, \quad \bar{b} = (1/\bar{k})J^\top \mathcal{B}, \quad \lambda_0(\bar{c}) = \bar{Q}\bar{c} + \bar{b}.$$

The Subdifferential of h

For any $c \in \mathcal{M}_{\bar{c}}$, $\partial h(c)$ can be given by two equivalent formulations:

$$\partial h(c) = \left\{ y \,\middle|\, \begin{array}{l} \exists\, \mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \geq 0 \\ \text{such that } Jy = \mathcal{Q}c + \mathcal{B} + \hat{\mathcal{A}}\mu \end{array} \right\} = \lambda_0(c) + \bar{A}\mathcal{U}(c),$$

where

$$\mathcal{U}(c) := \left\{ \mu \geq 0 \,\middle|\, \mathcal{A}\mu = \bar{k} \left[\mathcal{Q}c + \mathcal{B} - J(\bar{Q}c + \bar{b}) \right] \right\}.$$

The Subdifferential of h

For any $c \in \mathcal{M}_{\bar{c}}$, $\partial h(c)$ can be given by two equivalent formulations:

$$\partial h(c) = \left\{ y \,\middle|\, \begin{array}{l} \exists\, \mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \geq 0 \\ \text{such that } Jy = \mathcal{Q}c + \mathcal{B} + \hat{\mathcal{A}}\mu \end{array} \right\} = \lambda_0(c) + \bar{A}\mathcal{U}(c),$$

where

$$\mathcal{U}(c) := \left\{ \mu \geq 0 \,\middle|\, \mathcal{A}\mu = \bar{k}\left[\mathcal{Q}c + \mathcal{B} - J(\bar{Q}c + \bar{b})\right] \right\}.$$

Structure Functional of Osborne (01)

The Subdifferential of h

For any $c \in \mathcal{M}_{\bar{c}}$, $\partial h(c)$ can be given by two equivalent formulations:

$$\partial h(c) = \left\{ y \;\middle|\; \begin{array}{l} \exists\, \mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \geq 0 \\ \text{such that } Jy = \mathcal{Q}c + \mathcal{B} + \hat{\mathcal{A}}\mu \end{array} \right\} = \lambda_0(c) + \bar{A}\mathcal{U}(c),$$

where

$$\mathcal{U}(c) := \left\{ \mu \geq 0 \;\middle|\; \mathcal{A}\mu = \bar{k}\left[\mathcal{Q}c + \mathcal{B} - J(\bar{Q}c + \bar{b})\right] \right\}.$$

Nondegeneracy: We say $\mathcal{M}_{\bar{c}}$ satisfies the nondegenercy condition if $\ker(A) = \{0\}$.

The Subdifferential of h

For any $c \in \mathcal{M}_{\bar{c}}$, $\partial h(c)$ can be given by two equivalent formulations:

$$\partial h(c) = \left\{ y \;\middle|\; \begin{array}{l} \exists\, \mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \geq 0 \\ \text{such that } Jy = \mathcal{Q}c + \mathcal{B} + \hat{\mathcal{A}}\mu \end{array} \right\} = \lambda_0(c) + \bar{A}\mathcal{U}(c),$$

where

$$\mathcal{U}(c) := \left\{ \mu \geq 0 \;\middle|\; \mathcal{A}\mu = \bar{k}\left[\mathcal{Q}c + \mathcal{B} - J(\bar{Q}c + \bar{b})\right] \right\}.$$

Nondegeneracy: We say $\mathcal{M}_{\bar{c}}$ satisfies the nondegenercy condition if $\ker(A) = \{0\}$.

Lemma: Let $c \in \mathcal{M}_{\bar{c}}$. If $\ker A = \{0\}$, then, for every $y \in \partial h(c)$, there is a unique $\mu(c, y) \in \mathcal{U}(c)$ such that $y = \lambda_0(c) + \bar{A}\mu(c, y)$.

k-Strict Complementarity

Let $\bar{c} \in \operatorname{dom} h$. We say *$k$-strict complementarity* holds at $(c, y) \in \operatorname{graph}(\partial h)$ for $\mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \in \mathcal{U}(c)$ wrt $\mathcal{M}_{\bar{c}}$ if

(1) $c \in \mathcal{M}_{\bar{c}}$ and $y = \lambda_0(c) + \bar{A}\mu$,

(2) $\exists\ k \in \mathcal{K}(\bar{c})$ with $\mu_k > 0$,

(3) if $j \in \mathcal{K}(c) \setminus \{k\}$ and $i \in \{1, \ldots, \ell\}$ with $(\mu_j)_i = 0$, then the scalars $(P_{j'})_{ii} = 1$ for all $j' \in \mathcal{K}(c)$.

k-Strict Complementarity

Let $\bar{c} \in \operatorname{dom} h$. We say *$k$-strict complementarity* holds at $(c, y) \in \operatorname{graph}(\partial h)$ for $\mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \in \mathcal{U}(c)$ wrt $\mathcal{M}_{\bar{c}}$ if

(1) $c \in \mathcal{M}_{\bar{c}}$ and $y = \lambda_0(c) + \bar{A}\mu$,

(2) $\exists\ k \in \mathcal{K}(\bar{c})$ with $\mu_k > 0$,

(3) if $j \in \mathcal{K}(c) \setminus \{k\}$ and $i \in \{1, \ldots, \ell\}$ with $(\mu_j)_i = 0$, then the scalars $(P_{j'})_{ii} = 1$ for all $j' \in \mathcal{K}(c)$.

Lemma: Let $\bar{c} \in \operatorname{dom} h$. If $\mathcal{M}_{\bar{c}}$ is nondegenerate and for some $c \in \mathcal{M}_{\bar{c}}$ and there is a $(c, y) \in \operatorname{graph}(\partial h)$ such that k-strict complementarity holds at (c, y) wrt $\mathcal{M}_{\bar{c}}$, then $\mathcal{M}_{\bar{c}}$ is partly smooth.

k-Strict Complementarity

Let $\bar{c} \in \operatorname{dom} h$. We say *$k$-strict complementarity* holds at $(c, y) \in \operatorname{graph}(\partial h)$ for $\mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \in \mathcal{U}(c)$ wrt $\mathcal{M}_{\bar{c}}$ if

(1) $c \in \mathcal{M}_{\bar{c}}$ and $y = \lambda_0(c) + \bar{A}\mu$,

(2) $\exists\ k \in \mathcal{K}(\bar{c})$ with $\mu_k > 0$,

(3) if $j \in \mathcal{K}(c) \setminus \{k\}$ and $i \in \{1, \ldots, \ell\}$ with $(\mu_j)_i = 0$, then the scalars $(P_{j'})_{ii} = 1$ for all $j' \in \mathcal{K}(c)$.

Lemma: Let $\bar{c} \in \operatorname{dom} h$. If $\mathcal{M}_{\bar{c}}$ is nondegenerate and for some $c \in \mathcal{M}_{\bar{c}}$ and there is a $(c, y) \in \operatorname{graph}(\partial h)$ such that k-strict complementarity holds at (c, y) wrt $\mathcal{M}_{\bar{c}}$, then $\mathcal{M}_{\bar{c}}$ is partly smooth.

Moreover, if $\bar{x} \in \operatorname{dom} f$ and $\bar{y} \in \partial h(\bar{c})$ are such that $\bar{c} = c(\bar{x})$ and

$$\ker \nabla c(\bar{x})^\top \cap \operatorname{ri}(\partial h(\bar{c})) = \{\bar{y}\}, \qquad \text{(Strict Criticality (SC))}$$

then

$$D(\bar{x}) = \left\{ d \;\middle|\; h'(c(\bar{x}); \nabla c(\bar{x}) d) \le 0 \right\} = \ker A^\top \nabla c(\bar{x}).$$

Newton's Method Hypotheses

Let $f = h \circ c$ be PLQ convex composite, $\bar{x} \in \operatorname{dom} f$, $\bar{y} \in \partial h(c(\bar{x}))$, and set $\bar{c} := c(\bar{x})$.

Assumptions:

(a) c is $\mathcal{C}^3$-smooth,

(b) $\mathcal{M}_{\bar{c}}$ satisfies the nondegeneracy condition,

(c) f satisfies SC at $\bar{x}$ for $\bar{y}$,

(d) $\bar{x}$ satisfies the second-order sufficient conditions, i.e.,
$$h''(c(\bar{x}); \nabla c(\bar{x})d) + \left\langle d, \nabla^2_{xx} L(\bar{x}, \bar{y})d \right\rangle > 0 \quad \forall\, d \in \ker A^\top \nabla c(\bar{x}) \setminus \{0\},$$
where $M(\bar{x}) = \{\bar{y}\}$ and $D(\bar{x}) = \ker A^\top \nabla c(\bar{x})$.

Newton's Method Hypotheses

Let $f = h \circ c$ be PLQ convex composite, $\bar{x} \in \operatorname{dom} f$, $\bar{y} \in \partial h(c(\bar{x}))$, and set $\bar{c} := c(\bar{x})$.

Assumptions:

(a) c is $\mathcal{C}^3$-smooth,

(b) $\mathcal{M}_{\bar{c}}$ satisfies the nondegeneracy condition,

(c) f satisfies SC at $\bar{x}$ for $\bar{y}$,

(d) $\bar{x}$ satisfies the second-order sufficient conditions, i.e.,

$$h''(c(\bar{x}); \nabla c(\bar{x})d) + \left\langle d, \nabla^2_{xx} L(\bar{x}, \bar{y})d \right\rangle > 0 \quad \forall\, d \in \ker A^\top \nabla c(\bar{x}) \setminus \{0\},$$

where $M(\bar{x}) = \{\bar{y}\}$ and $D(\bar{x}) = \ker A^\top \nabla c(\bar{x})$.

NLP Analogues:

(b) linear independence of the active constraint gradients,

(c) strict complementary slackness, and

(d) strong second-order sufficiency condition.

Convergence of Newton's Method

There exists a neighborhood $\mathcal{N}$ of $(\bar{x}, \bar{y})$ such that if $(x^0, y^0) \in \mathcal{N}$, then there exists a unique sequence $\{(x^k, y^k)\}$ satisfying the optimality conditions of $\mathbf{P_k}$ with $H_k := \nabla^2_{xx} L(x^k, y^k)$ such that, for all $k \in \mathbb{N}$,

(i) $c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}] \in \mathcal{M}_{\bar{c}}$,

(ii) $y^k \in \operatorname{ri} \partial h(c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}])$,

(iii) $H_{k-1}[x^k - x^{k-1}] + \nabla c(x^{k-1})^\top y^k = 0$,

(iv) x^{k+1} is a strong local minimizer of $\mathbf{P_k}$.

Moreover, the sequence (x^k, y^k) converges to $(\bar{x}, \bar{y})$ at a quadratic rate.