Convergence rate analysis of SCP_{ls} for a class of constrained difference-of-convex optimization problems

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong

64th Annual Meeting of the Australian Mathematical Society
December 2020
(Joint work with Zhaosong Lu and Peiran Yu)

Motivating applications

Structured constrained optimization problems:

- Compressed sensing with Gaussian noise:

$$\min_x \ \|x\|_1 - \|x\| \ \text{ subject to } \ \|Ax - b\| \leq \sigma,$$

  where $A \in \mathbb{R}^{q \times n}$, $b \in \mathbb{R}^q$, $\sigma \in (0, \|b\|)$.

- Compressed sensing with Cauchy noise:

$$\min_x \ \|x\|_1 - \|x\| \ \text{ subject to } \ \|Ax - b\|_{LL_2,\gamma} \leq \sigma,$$

  where $A \in \mathbb{R}^{q \times n}$, $b \in \mathbb{R}^q$, $\sigma \in (0, \|b\|_{LL_2,\gamma})$, and $\|\cdot\|_{LL_2,\gamma}$ is the Lorentzian norm.

$$\|y\|_{LL_2,\gamma} = \sum_{i=1}^{m} \log\left(1 + \frac{y_i^2}{\gamma^2}\right).$$

General model

Consider

$$\min_{x\in\mathbb{R}^n} F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot)\le 0}(x),$$

where:

- f has Lipschitz gradient.
- P_1 and P_2 are convex continuous.
- $g(x) = (g_1(x), \ldots, g_m(x))$ with each g_i having Lipschitz gradient.
- $\{x : \; g(x) \le 0\} \neq \emptyset$.
- F is level-bounded.

General model

Consider

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot)\le 0}(x),$$

where:

- f has Lipschitz gradient.
- P_1 and P_2 are convex continuous.
- $g(x) = (g_1(x), \dots, g_m(x))$ with each g_i having Lipschitz gradient.
- $\{x : \ g(x) \le 0\} \neq \emptyset$.
- F is level-bounded.
- The MFCQ holds at every point satisfying $g(x) \le 0$.

General model

Consider

$$\min_{x\in\mathbb{R}^n} F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot)\leq 0}(x),$$

where:

- f has Lipschitz gradient.
- P_1 and P_2 are convex continuous.
- $g(x) = (g_1(x), \ldots, g_m(x))$ with each g_i having Lipschitz gradient.
- $\{x: \ g(x) \leq 0\} \neq \emptyset$.
- F is level-bounded.
- The MFCQ holds at every point satisfying $g(x) \leq 0$.

Algorithmic ideas:

- Augmented Lagrangian.

General model

Consider

$$\min_{x \in \mathbb{R}^n} \; F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot) \leq 0}(x),$$

where:

- f has Lipschitz gradient.
- P_1 and P_2 are convex continuous.
- $g(x) = (g_1(x), \ldots, g_m(x))$ with each g_i having Lipschitz gradient.
- $\{x : \; g(x) \leq 0\} \neq \emptyset$.
- F is level-bounded.
- The MFCQ holds at every point satisfying $g(x) \leq 0$.

Algorithmic ideas:

- Augmented Lagrangian.
- Moving balls approximation.

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_{x} \; f(x) \;\; \text{subject to } g(x) \leq 0.$$

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_x \; f(x) \;\text{ subject to } g(x) \le 0.$$

Key update: Starting with an x^t satisfying $g(x^t) \le 0$, compute

$$\begin{aligned} x^{t+1} = \arg\min_x \quad & f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \tfrac{L_f}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle + \tfrac{L_{g_i}}{2}\|x - x^t\|^2 \le 0 \;\; \forall i. \end{aligned}$$

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_x \; f(x) \; \text{ subject to } g(x) \le 0.$$

Key update: Starting with an x^t satisfying $g(x^t) \le 0$, compute

$$\begin{aligned} x^{t+1} = \arg\min_x \quad & f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \tfrac{L_f}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle + \tfrac{L_{g_i}}{2}\|x - x^t\|^2 \le 0 \;\; \forall i. \end{aligned}$$

- The above algorithm is well defined and any accumulation point of $\{x^t\}$ is stationary. (Auslender, Shefi, Teboulle '10)
- Convergence of $\{x^t\}$ under convexity (Auslender, Shefi, Teboulle '10) or semialgebraicity (Bolte, Pauwels '16) is known.
- Variants with line-search scheme have been proposed (Lu '12) (Bolte, Chen, Pauwels '19).

Subproblem needs iterative solver except for $m = 1$.

Sequential convex programming with line search

Algorithm 1: SCP$_{ls}$ (Lu '12)

Pick $c > 0$, $0 < \underline{L} < \overline{L}$, $\tau > 1$, x^0 with $g(x^0) \le 0$. Set $t = 0$.

Step 1. Choose any $\xi^t \in \partial P_2(x^t)$.

Step 2. Pick $L_f^{t,0} \in [\underline{L}, \overline{L}]$ and $L_g^{t,0} \in [\underline{L}, \overline{L}]^m$. Set $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^{t,0}, L_g^{t,0})$.

Step 3. Compute

$$\begin{aligned} \widetilde{x} = \arg\min_x \ & \left\{ \langle \nabla f(x^t) - \xi^t, x - x^t \rangle + \frac{\widetilde{L}_f}{2} \|x - x^t\|^2 + P_1(x) \right\} \\ \text{s.t.} \quad & g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle + \frac{(\widetilde{L}_g)_i}{2} \|x - x^t\|^2 \le 0 \ \ \forall i. \end{aligned}$$

Step 3a) If $g(\widetilde{x}) \not\le 0$, let $\widetilde{L}_g \leftarrow \tau \widetilde{L}_g$ and go to Step 3. Else, go to Step 3b).

Step 3b) If $F(\widetilde{x}) \le F(x^t) - \frac{c}{2}\|\widetilde{x} - x^t\|^2$, go to Step 4. Else, let $\widetilde{L}_f \leftarrow \tau \widetilde{L}_f$ and go to step 3.

Step 4. If a termination criterion is not met, set $(L_f^t, L_g^t) = (\widetilde{L}_f, \widetilde{L}_g)$ and $x^{t+1} = \widetilde{x}$. Update $t \leftarrow t + 1$ and go to **Step 1.**

SCP_{ls}: Subsequential convergence

Theorem 1. (Lu '12), (Yu, P., Lu '20)

(i) SCP_{ls} is well defined.

(ii) The Slater condition holds for each subproblem.

(iii) Let λ^t be a Lagrange multiplier for the subproblem at the end of the tth iteration. Then $\{\lambda^t\}$ is bounded.

(iv) Let $\{x^t\}$ be the sequence generated by SCP_{ls}. Then the sequence $\{x^t\}$ is bounded, $\lim\limits_{t\to\infty} \|x^{t+1} - x^t\| = 0$, and any accumulation point x^* is a stationary point, in the sense that

$$0 \in \nabla f(x^*) + \partial P_1(x^*) - \partial P_2(x^*) + N_{g(\cdot)\le 0}(x^*).$$

Outline

Aim: Analyze the convergence rate SCP_{ls}, using suitable Kurdyka-Łojasiewicz (KL) type assumptions.

Outline:

- Convergence rate analysis in general nonconvex settings.
- Convergence rate analysis in convex settings.
- Explicit KL exponent for some models.
- Applications and future directions.

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^{\alpha} \leq \operatorname{dist}(0, \partial h(x))$$

  whenever $x \in \operatorname{dom} \partial h$, $\|x - \bar{x}\| \leq \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^{\alpha} \le \operatorname{dist}(0, \partial h(x))$$

  whenever $x \in \operatorname{dom} \partial h$, $\|x - \bar{x}\| \le \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.
- If h has the KL property at every $\bar{x} \in \operatorname{dom} \partial h$ with the same α, then h is said to be a KL function with exponent α.

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^{\alpha} \leq \mathrm{dist}(0, \partial h(x))$$

  whenever $x \in \operatorname{dom} \partial h$, $\|x - \bar{x}\| \leq \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.
- If h has the KL property at every $\bar{x} \in \operatorname{dom} \partial h$ with the same α, then h is said to be a KL function with exponent α.

Examples.

- Proper closed semialgebraic functions are KL functions with exponent $\alpha \in [0, 1)$. (Bolte et al. '07)

Convergence rate: Nonconvex settings

Define

$$\bar{F}(x,y,w) := f(x) + P_1(x) - P_2(x) + \delta_{\bar{G}(\cdot)\le 0}(x,y,w),$$

where

$$\bar{G}(x,y,w) := \begin{pmatrix} g_1(y) + \langle \nabla g_1(y), x-y\rangle + \frac{w_1}{2}\|x-y\|^2 \\ \vdots \\ g_m(y) + \langle \nabla g_m(y), x-y\rangle + \frac{w_m}{2}\|x-y\|^2 \end{pmatrix}$$

Convergence rate: Nonconvex settings

Define

$$\bar{F}(x,y,w) := f(x) + P_1(x) - P_2(x) + \delta_{\bar{G}(\cdot)\le 0}(x,y,w),$$

where

$$\bar{G}(x,y,w) := \begin{pmatrix} g_1(y) + \langle \nabla g_1(y), x-y\rangle + \frac{w_1}{2}\|x-y\|^2 \\ \vdots \\ g_m(y) + \langle \nabla g_m(y), x-y\rangle + \frac{w_m}{2}\|x-y\|^2 \end{pmatrix}$$

Theorem 2. (Yu, P., Lu '20)
Suppose $g \in C^2$, ∇P_2 is Lipschitz around the set of stationary points, and $\bar{F}$ is KL. Let $\{x^t\}$ be generated by SCP$_{ls}$. Then $x^t \to x^*$ for some x^*. If in addition $\bar{F}$ is KL with exponent $\alpha \in [0,1)$, then

(i) if $\alpha = 0$, then $\{x^t\}$ converges finitely;

(ii) if $\alpha \in (0, \frac{1}{2}]$, then $\{x^t\}$ converges locally linearly;

(iii) if $\alpha \in (\frac{1}{2}, 1)$, then $\|x^t - x^*\| = O(t^{-\frac{1-\alpha}{2\alpha-1}})$.

Convergence rate: Convex settings

Recall

$$F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot)\leq 0}(x).$$

Theorem 3. (Yu, P., Lu '20)
Suppose $\{f, g_1, \ldots, g_m\}$ are convex and $P_2 = 0$. Let $\{x^t\}$ be the sequence generated by SCP$_{ls}$. Then $x^t \to x^*$ for some x^*. If in addition F is KL with exponent $\alpha \in [0, 1)$, then:

(i) if $\alpha \in [0, \frac{1}{2}]$, then $\{x^t\}$ converges locally linearly;

(ii) if $\alpha \in (\frac{1}{2}, 1)$, then $\|x^t - x^*\| = O(t^{-\frac{1-\alpha}{2\alpha-1}})$.

Convergence rate: Convex settings

Recall

$$F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot)\leq 0}(x).$$

Theorem 3. (Yu, P., Lu '20)
Suppose $\{f, g_1, \ldots, g_m\}$ are convex and $P_2 = 0$. Let $\{x^t\}$ be the sequence generated by SCP$_{ls}$. Then $x^t \to x^*$ for some x^*. If in addition F is KL with exponent $\alpha \in [0, 1)$, then:

(i) if $\alpha \in [0, \frac{1}{2}]$, then $\{x^t\}$ converges locally linearly;

(ii) if $\alpha \in (\frac{1}{2}, 1)$, then $\|x^t - x^*\| = O(t^{-\frac{1-\alpha}{2\alpha-1}})$.

Remark: It actually holds that for some $\kappa > 0$

(i) if $\alpha \in [0, \frac{1}{2}]$, then $\{F(x^t) - F^* + \kappa\,\mathrm{dist}^2(x^t, \operatorname{Arg\,min} F)\}$ converges locally Q-linearly;

(ii) if $\alpha \in (\frac{1}{2}, 1)$, then $F(x^t) - F^* + \kappa\,\mathrm{dist}^2(x^t, \operatorname{Arg\,min} F) = O(t^{-\frac{1-\alpha}{2\alpha-1}})$.

Explicit KL exponent

Consider

$$\begin{array}{ll} \min\limits_{x} & P(x) \\ \text{s.t.} & h_i(A_i x) \le 0 \quad \text{for} \;\; i = 1, \cdots, m, \end{array}$$

where $P : \mathbb{R}^n \to \mathbb{R}$ is convex, $A_i \in \mathbb{R}^{q_i \times n}$ and $h_i : \mathbb{R}^{q_i} \to \mathbb{R}$ is *strictly* convex. Suppose the feasible set is nonempty.

Explicit KL exponent

Consider

$$\min_{x} \quad P(x)$$
$$\text{s.t.} \quad h_i(A_i x) \le 0 \quad \text{for} \ \ i = 1, \cdots, m,$$

where $P : \mathbb{R}^n \to \mathbb{R}$ is convex, $A_i \in \mathbb{R}^{q_i \times n}$ and $h_i : \mathbb{R}^{q_i} \to \mathbb{R}$ is *strictly* convex. Suppose the feasible set is nonempty.

Theorem 5. (Yu, P., Lu '20)
Let $F(x) := P(x) + \sum_{i=1}^m \delta_{h_i(A_i \cdot) \le 0}(x)$ and $\bar{x} \in \operatorname{Arg\,min} F$. Suppose

(i) There exists a Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^m_+$ and $x \mapsto P(x) + \sum_{i=1}^m \bar{\lambda}_i h_i(A_i x)$ is KL with exponent $\alpha \in (0, 1)$.

(ii) The strict complementarity condition holds at $(\bar{x}, \bar{\lambda})$.

Then F satisfies KL property at $\bar{x}$ with exponent α.

Applications

Consider

$$\begin{array}{ll} \min_x & \|x\|_1 - \mu\|x\| \\ \text{s.t.} & \frac{1}{2}\|Ax - b\|^2 \le \delta \end{array}$$

Suppose

- $\mu \in [0, 1]$;
- $\delta \in (0, \frac{1}{2}\|b\|^2)$;
- $A \in \mathbb{R}^{q\times n}$ has *full row rank*;
- A does not have zero columns when $\mu = 1$.

Then our convergence theorem can be applied.

Consider

$$\begin{array}{ll} \min_x & \|x\|_1 - \mu\|x\| \\ \text{s.t.} & \|Ax - b\|_{LL_{2,\gamma}} \le \delta \end{array}$$

Suppose

- $\mu \in [0, 1]$;
- $\delta \in (0, \|b\|_{LL_{2,\gamma}})$;
- $A \in \mathbb{R}^{q\times n}$ has *full row rank*;
- A does not have zero columns when $\mu = 1$.

Then our convergence theorem can be applied.

Applications

Consider

$$\min_x \quad \|x\|_1 - \mu\|x\|$$
$$\text{s.t.} \quad \frac{1}{2}\|Ax - b\|^2 \le \delta$$

Suppose

- $\mu \in [0, 1]$;
- $\delta \in (0, \frac{1}{2}\|b\|^2)$;
- $A \in \mathbb{R}^{q \times n}$ has *full row rank*;
- A does not have zero columns when $\mu = 1$.

Then our convergence theorem can be applied. Moreover, if $\mu = 0$, we can deduce local linear convergence.

Consider

$$\min_x \quad \|x\|_1 - \mu\|x\|$$
$$\text{s.t.} \quad \|Ax - b\|_{LL_{2,\gamma}} \le \delta$$

Suppose

- $\mu \in [0, 1]$;
- $\delta \in (0, \|b\|_{LL_{2,\gamma}})$;
- $A \in \mathbb{R}^{q \times n}$ has *full row rank*;
- A does not have zero columns when $\mu = 1$.

Then our convergence theorem can be applied.

Conclusion and future work

Conclusion:

- Convergence rate of SCP_{ls} is analyzed in both convex and nonconvex settings, under two different KL assumptions.
- Explicit KL exponent of some constrained optimization models are obtained.

Conclusion and future work

Conclusion:

- Convergence rate of SCP_{ls} is analyzed in both convex and nonconvex settings, under two different KL assumptions.
- Explicit KL exponent of some constrained optimization models are obtained.

Future work:

- Minimizing $\|x\|_{\frac{1}{2}}^{\frac{1}{2}}$ over $\|Ax - b\| \leq \sigma$?
- General nonlinear cone constraints?

References:

- P. Yu, T. K. Pong and Z. Lu. *Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems*. Submitted, January 2020.

Thanks for coming! $\ddot\smile$

Explicit KL exponent II

Consider

$$\begin{array}{ll} \min\limits_{x} & P(x) \\ \text{s.t.} & h_1(A_1x) \le 0, \end{array}$$

where $P : \mathbb{R}^n \to \mathbb{R}$ is convex, $A_1 \in \mathbb{R}^{q_1 \times n}$ and $h_1 : \mathbb{R}^{q_1} \to \mathbb{R}$ is *strictly* convex. Suppose the feasible set is nonempty.

Corollary 1. (Yu, P., Lu '20)
Let $F(x) := P(x) + \delta_{h_1(A_1\cdot)\le 0}(x)$ and $\bar{x} \in \operatorname{Arg\,min} F$. Suppose

(i) $\inf F > \inf P$.

(ii) There exists a Lagrange multiplier $\bar{\lambda}_1 \in \mathbb{R}_+$ and $x \mapsto P(x) + \bar{\lambda}_1 h_1(A_1x)$ is KL with exponent $\alpha \in (0, 1)$.

Then F satisfies KL property at $\bar{x}$ with exponent α.

Explicit KL exponent II

Consider

$$\begin{array}{cl} \min\limits_x & P(x) \\ \text{s.t.} & h_1(A_1 x) \le 0, \end{array}$$

where $P : \mathbb{R}^n \to \mathbb{R}$ is convex, $A_1 \in \mathbb{R}^{q_1 \times n}$ and $h_1 : \mathbb{R}^{q_1} \to \mathbb{R}$ is *strictly* convex. Suppose the feasible set is nonempty.

Corollary 1. (Yu, P., Lu '20)
Let $F(x) := P(x) + \delta_{h_1(A_1 \cdot) \le 0}(x)$ and $\bar{x} \in \mathrm{Arg\,min}\, F$. Suppose

(i) $\inf F > \inf P$.

(ii) There exists a Lagrange multiplier $\bar{\lambda}_1 \in \mathbb{R}_+$ and $x \mapsto P(x) + \bar{\lambda}_1 h_1(A_1 x)$ is KL with exponent $\alpha \in (0, 1)$.

Then F satisfies KL property at $\bar{x}$ with exponent α.

Remark: Can be applied with $P = \|\cdot\|_1$ and $h_1 = \frac{1}{2}\|\cdot - b\|^2 - \delta$ or Poisson/logistic loss.