On the Convergence Properties of Infeasible Inexact Proximal Alternating Linearized Minimization

Yukuan Hu

State Key Laboratory of Scientific and Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, China

Joint work with **Xin Liu** (AMSS)

Group Seminar
Aug. 11, 2023

Outline

Outline

Block-Structured Optimization Problems

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}_1, \ldots, \mathbf{x}_n), \\ \text{s. t.} \quad & \mathbf{x}_i \in \mathcal{S}_i := \{\mathbf{w}_i \in \mathbb{R}^{m_i} : \mathbf{h}_i(\mathbf{w}_i) \le 0\}, \; i \in [n]. \end{aligned} \tag{P}$$

- $\mathbf{x} = (\mathbf{x}_i)_{i=1}^n := (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \bigotimes_{i=1}^n \mathbb{R}^{m_i}$, $[n] := \{1, \ldots, n\}$.
- $f : \bigotimes_{i=1}^n \mathbb{R}^{m_i} \to \mathbb{R}$: differentiable, possibly nonconvex.
- $\mathbf{h}_i := (h_{i,1}, \ldots, h_{i,p_i})^\top : \mathbb{R}^{m_i} \to \mathbb{R}^{p_i}$: convex differentiable, $i \in [n]$.
- Applications:
  - 3D anisotropic frictional contact; [Kučera 2008]
  - image processing; [Bonettini et al. 2018]
  - topology optimization; [Liu et al. 2020]
  - electronic structure calculation; [H. et al. 2023]
  - ...

Proximal Alternating Linearized Minimization (PALM)

Framework 1: PALM for solving the problem (P). [Bolte et al. 2014]

Input: $\mathbf{x}^{(0)} = (\mathbf{x}_i^{(0)})_{i=1}^n \in \bigotimes_{i=1}^n \mathbb{R}^{m_i}$, $\{\sigma_i^{(0)} > 0\}_{i=1}^n$.

Set $k := 0$.

while *certain conditions not satisfied* **do**

  for $i \in [n]$ **do**

    Solve the i-th proximal linearized subproblem

$$\min_{\mathbf{x}_i \in \mathcal{S}_i} \left\langle \nabla_i f(\mathbf{x}_{<i}^{(k+1)}, \mathbf{x}_{\geq i}^{(k)}), \mathbf{x}_i - \mathbf{x}_i^{(k)} \right\rangle + \frac{\sigma_i^{(k)}}{2} \|\mathbf{x}_i - \mathbf{x}_i^{(k)}\|^2 \tag{1}$$

    to obtain $\mathbf{x}_i^{(k+1)} \in \mathbb{R}^{m_i}$ fulfilling *certain conditions*.

    Update the i-th proximal parameter $\sigma_i^{(k)}$ to $\sigma_i^{(k+1)} > 0$ if necessary.

  end

  Set $k := k + 1$.

end

Output: $\mathbf{x}^{(k)} = (\mathbf{x}_i^{(k)})_{i=1}^n \in \bigotimes_{i=1}^n \mathbb{R}^{m_i}$.

- $\nabla_i := \nabla_{\mathbf{x}_i}$, $\mathbf{x}_{<i} := (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$, $\mathbf{x}_{\geq i} := (\mathbf{x}_i, \ldots, \mathbf{x}_n)$, $i \in [n]$.
- Solving the subproblem (1) $\Leftrightarrow$ Projecting

$$\tilde{\mathbf{x}}_i^{(k)} := \mathbf{x}_i^{(k)} - \frac{1}{\sigma_i^{(k)}} \nabla_i f(\mathbf{x}_{<i}^{(k+1)}, \mathbf{x}_{\geq i}^{(k)}) \tag{2}$$

  onto $\mathcal{S}_i$.

Related Works

PALM-E [Razaviyayn et al. 2013; Xu-Yin 2013; Bolte et al. 2014; Wang et al. 2018; ...]

- Subproblem (1) is **exactly** solved.
- Sufficient reduction:

$$a_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \leq f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k+1)}).$$

- Subgradient norm lower bound:

$$\operatorname{dist}(0, \partial F(\mathbf{x}^{(k+1)})) \leq a_2\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|,$$

where $F := f + \sum_{i=1}^{n} \delta_{\mathcal{S}_i}$, $\delta_{\mathcal{S}_i}$ is the indicator function for $\mathcal{S}_i$.

PALM-F [Hua-Yamashita 2016; Bonettini et al. 2018; Ochs 2019; Gur et al. 2022; ...]

- Subproblem (1) is **inexactly** solved but $\mathbf{x}_i^{(k)} \in \mathcal{S}_i$ throughout iterations.
- (Approx.) sufficient reduction + (Approx.) subgradient norm lower bound.
- Objective monotonicity is enforced via, e.g., line search.

Related Works

PALM-E [Razaviyayn et al. 2013; Xu-Yin 2013; Bolte et al. 2014; Wang et al. 2018; ...]

- Subproblem (1) is **exactly** solved.
- Sufficient reduction:

$$a_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \leq f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k+1)}).$$

- Subgradient norm lower bound:

$$\operatorname{dist}(0, \partial F(\mathbf{x}^{(k+1)})) \leq a_2\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|,$$

where $F := f + \sum_{i=1}^{n} \delta_{\mathcal{S}_i}$, $\delta_{\mathcal{S}_i}$ is the indicator function for $\mathcal{S}_i$.

PALM-F [Hua-Yamashita 2016; Bonettini et al. 2018; Ochs 2019; Gur et al. 2022; ...]

- Subproblem (1) is **inexactly** solved but $\mathbf{x}_i^{(k)} \in \mathcal{S}_i$ throughout iterations.
- (Approx.) sufficient reduction + (Approx.) subgradient norm lower bound.
- Objective monotonicity is enforced via, e.g., line search.

What if infeasible inexactness emerges?

Two Illustrative Examples

Example 1 (Linear constraints)

The feasible region $\mathcal{S}_i$ is the **Birkhoff polytope**

$$\mathcal{S}_i := \left\{ W \in \mathbb{R}^{m_i \times m_i} : W\mathbf{1} = \mathbf{1},\ W^\top \mathbf{1} = \mathbf{1},\ W \geq 0 \right\},$$

where "$\mathbf{1}$" is the all-one vector in $\mathbb{R}^{m_i}$.

- Applications: optimal transport problems, electronic structure calculation, ...
  [Chen et al. 2014; Peyré-Cuturi 2019; H. et al. 2023; ...]
- # variables $= m_i^2$ vs # equality constraints $= 2m_i \Rightarrow$ Dual perspective.
- Semismooth Newton method for dual problem $\Rightarrow$ Infeasibility.
  [Li et al. 2020; H. et al. 2023]

Two Illustrative Examples (Cont.)

Example 2 (Nonlinear constraints)

The feasible region $\mathcal{S}_i$ is an **ellipsoid** in $\mathbb{R}^{m_i}$

$$\mathcal{S}_i := \left\{ \mathbf{w} \in \mathbb{R}^{m_i} : \frac{1}{2}\mathbf{w}^\top A_i \mathbf{w} + \mathbf{b}_i^\top \mathbf{w} \leq \alpha_i \right\},$$

where $I \neq A_i \in \mathbb{S}^{m_i}_{++}$, $\mathbf{b}_i \in \mathbb{R}^{m_i}$, and $\alpha_i > 0$.

- Applications: topology optimization, 3D anisotrpic friction contact, polynomial optimization, sensor network localization, ...
  [Kučera 2008; He et al. 2010; Luo-Zhang 2010; Li et al. 2012; Liu et al. 2020; ...]
- Hybrid projection algorithm $\Rightarrow$ Feasibility. [Dai 2006]
- Alternating direction method of multipliers $\Rightarrow$ Infeasibility. [Jia et al. 2017]

Related Works (Cont.)

PALM-I

- Subproblem (1) is **inexactly** solved and $\mathbf{x}_i^{(k)} \notin \mathcal{S}_i$ can happen.
- Difficulties: objective nonmonotonicity $\not\Rightarrow$ stationarity.
- The only existing work: [Frankel et al. 2015]
  - hypothesis on iterates: there exist $\beta_1, \beta_2 > 0$ such that, for any $i \in [n]$, $k \geq 0$,

$$\begin{cases} \sum_{j=1}^{i-1} \|\mathbf{x}_j^{(k+1)} - \bar{\mathbf{x}}_j^{(k+1)}\| + \sum_{j=i}^{n} \|\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_j^{(k)}\| \leq \beta_1 \|\bar{\mathbf{x}}_i^{(k+1)} - \bar{\mathbf{x}}_i^{(k)}\|, \\ \left\langle \mathbf{x}_i^{(k)} - \bar{\mathbf{x}}_i^{(k)}, \bar{\mathbf{x}}_i^{(k+1)} - \bar{\mathbf{x}}_i^{(k)} \right\rangle \leq \beta_2 \|\bar{\mathbf{x}}_i^{(k+1)} - \bar{\mathbf{x}}_i^{(k)}\|^2, \end{cases}$$

    where $\bar{\mathbf{x}}_i^{(k+1)}$ is the solution of the subproblem (1);
  - sufficient reduction over $\{f(\bar{\mathbf{x}}^{(k)})\}$, where $\bar{\mathbf{x}}^{(k)} := (\bar{\mathbf{x}}_i^{(k)})_{i=1}^n \in \bigotimes_{i=1}^n \mathbb{R}^{m_i}$;
  - only theoretical values.

Our goal: convergence properties of PALM-I with implementable inexact criteria.

Outline

Complete Description of PALM-I

Algorithm 2: PALM-I for solving the problem (P).

Input: $\mathbf{x}^{(0)} \in \bigotimes_{i=1}^{n} \mathbb{R}^{m_i}$, $\bar{\varepsilon} > 0$, $\{\varepsilon^{(k)} \in [0, \bar{\varepsilon}]\}$, $M_u, M_l > 0$, $\{\sigma_i^{(0)} \in [M_l, M_u]\}_{i=1}^{n}$.

Set $k := 0$.

while *certain conditions not satisfied* **do**

  for $i \in [n]$ **do**

    Solve the i-th proximal linearized subproblem (1) to obtain $\mathbf{x}_i^{(k+1)} \in \mathbb{R}^{m_i}$ such that

$$\exists\, \boldsymbol{\lambda}_i^{(k+1)} \in \mathbb{R}_+^{p_i}, \text{ s. t. } \sqrt{r_i(\mathbf{x}_i^{(k+1)}, \boldsymbol{\lambda}_i^{(k+1)}, \tilde{\mathbf{x}}_i^{(k)})} \leq \varepsilon^{(k)},$$

    where $\tilde{\mathbf{x}}_i^{(k)}$ is computed via the equation (2).

    Update the i-th proximal parameter $\sigma_i^{(k)}$ to $\sigma_i^{(k+1)} \in [M_l, M_u]$ if necessary.

  end

  Set $k := k + 1$.

end

Output: $\mathbf{x}^{(k)} = (\mathbf{x}_i^{(k)})_{i=1}^{n} \in \bigotimes_{i=1}^{n} \mathbb{R}^{m_i}$.

For any $i \in [n]$, residual function $r_i : \mathbb{R}^{m_i} \times \mathbb{R}^{p_i} \times \mathbb{R}^{m_i} \to \mathbb{R}_+$ is defined as

$$r_i(\mathbf{x}_i, \boldsymbol{\lambda}_i, \tilde{\mathbf{x}}_i) := \underbrace{\max\left\{\left\langle \mathbf{x}_i, \mathbf{x}_i - \tilde{\mathbf{x}}_i + \nabla \mathbf{h}_i(\mathbf{x}_i)\boldsymbol{\lambda}_i \right\rangle, 0\right\} + \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_i + \nabla \mathbf{h}_i(\mathbf{x}_i)\boldsymbol{\lambda}_i \right\|_\infty}_{\text{substationarity}}$$

$$+ \underbrace{\left\| \max\left\{\mathbf{h}_i(\mathbf{x}_i), 0\right\} \right\|_\infty}_{\text{feasibility}} + \underbrace{\max\left\{-\left\langle \boldsymbol{\lambda}_i, \mathbf{h}_i(\mathbf{x}_i) \right\rangle, 0\right\}}_{\text{complementary slackness}}.$$

How to obtain $\boldsymbol{\lambda}_i^{(k+1)}$?

- Primal-dual subsolvers $\Rightarrow$ Dual variables.
- Solving an extra linear programming:

$$\min_{\boldsymbol{\lambda}_i \in \mathbb{R}_+^{p_i}} 0, \quad \text{s. t.} \begin{cases} \left\langle \mathbf{x}_i^{(k+1)}, \mathbf{x}_i^{(k+1)} - \tilde{\mathbf{x}}_i^{(k)} + \nabla \mathbf{h}_i(\mathbf{x}_i^{(k+1)}) \boldsymbol{\lambda}_i \right\rangle \leq \frac{(\varepsilon^{(k)})^2}{4}, \\ -\frac{(\varepsilon^{(k)})^2}{4}\mathbf{1} \leq \mathbf{x}_i^{(k+1)} - \tilde{\mathbf{x}}_i^{(k)} + \nabla \mathbf{h}_i(\mathbf{x}_i^{(k+1)}) \boldsymbol{\lambda} \leq \frac{(\varepsilon^{(k)})^2}{4}\mathbf{1}, \\ -\left\langle \boldsymbol{\lambda}, \mathbf{h}_i(\mathbf{x}_i^{(k+1)}) \right\rangle \leq \frac{(\varepsilon^{(k)})^2}{4}, \end{cases}$$

at any subiteration where $\| \max\{\mathbf{h}_i(\mathbf{x}_i^{(k+1)}), 0\}\|_\infty \leq \frac{(\varepsilon^{(k)})^2}{4}$.

Error Bound for Subproblem

Lemma 1

Suppose that f is continuously differentiable with respect to each variable block over $\bigotimes_{i=1}^n \bar{\mathcal{S}}_i$, where $\bar{\mathcal{S}}_i := \{\mathbf{w}_i \in \mathbb{R}^{m_i} : \mathrm{dist}(\mathbf{w}_i, \mathcal{S}_i) \leq \bar{\varepsilon}\}$ for $i \in [n]$. For any $i \in [n]$, assume that $\mathcal{S}_i$ is convex compact and $\mathbf{h}_i$ fulfills one of the following:

(i) *$\mathbf{h}_i$ is a linear mapping;*

(ii) *$\mathbf{h}_i$ satisfies the Slater constraint qualification (CQ), i.e., $\mathbf{h}_i(\hat{\mathbf{x}}_i) < 0$ for some $\hat{\mathbf{x}}_i \in \mathbb{R}^{m_i}$, and the Hoffman-like bound*

$$\begin{aligned} &\mathrm{dist}(\mathbf{x}_i, \mathcal{S}_i) \leq \tilde{c}_i \|\max\{\mathbf{h}_i(\mathbf{x}_i), 0\}\|, \\ &\qquad \forall\, \mathbf{x}_i \in \tilde{\mathcal{S}}_i := \left\{ \mathbf{w}_i \in \mathbb{R}^{m_i} : \mathrm{dist}(\mathbf{w}_i, \bar{\mathcal{S}}_i) \leq \frac{\bar{\mathcal{M}}_i}{M_l} \right\} \end{aligned} \tag{3}$$

holds for some constant $\tilde{c}_i \geq 0$, where $\bar{\mathcal{M}}_i := \sup_{\mathbf{x} \in \bigotimes_{i=1}^n \bar{\mathcal{S}}_i} \|\nabla_i f(\mathbf{x})\|$.

Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I. *Then there exists a constant $\omega \geq 0$ such that $\|\mathbf{x}^{(k+1)} - \bar{\mathbf{x}}^{(k+1)}\| \leq \omega \varepsilon^{(k)}$ holds for any $k \geq 0$.*

When does the Hoffman-like bound hold?

[Mangasarian 1998; Bertsekas 1999]

- $\mathbf{h}_i$ is linear (Example 1).
- $\mathbf{h}_i$ satisfies an enhanced Slater CQ (Example 2).

$$\begin{cases} \exists\, \hat{\mathbf{x}}_i \in \mathbb{R}^{m_i}, \text{ s. t. } \mathbf{h}_i(\hat{\mathbf{x}}_i) < 0; \text{ and} \\ \exists\, \zeta \geq 0, \text{ s. t. } \dfrac{\|\mathbf{y}_i - \hat{\mathbf{x}}_i\| - \operatorname{dist}(\mathbf{y}_i, \mathcal{S}_i)}{\min_{j=1,\ldots,p_i}\{-h_{i,j}(\hat{\mathbf{x}}_i)\}} \leq \zeta, \ \forall\, \mathbf{y}_i \in \tilde{\mathcal{S}}_i. \end{cases}$$

Our inexact criteria are more implementable.

Outline

Assumptions on Input Data

Assumption 1

The objective function f is Lipschitz continuously differentiable with respect to each variable block over $\bigotimes_{i=1}^{n} \bar{\mathcal{S}}_i$, namely, for $i \in [n]$, there exists a modulus $L_i > 0$ such that $\|\nabla_i f(\mathbf{x}) - \nabla_i f(\mathbf{x}')\| \leq L_i \|\mathbf{x} - \mathbf{x}'\|$ for any $\mathbf{x}$, $\mathbf{x}' \in \bigotimes_{i=1}^{n} \bar{\mathcal{S}}_i$.

Assumption 2

For any $i \in [n]$, $\mathcal{S}_i$ is convex and compact, and one of the following holds for $\mathbf{h}_i$:

(a) $\mathbf{h}_i$ *is linear;*

(b) $\mathbf{h}_i$ *satisfies the Slater CQ and the Hoffman-like bound (3).*

Conditions on the Tolerance Sequence

Condition 1

(a) *The sequence* $\{\varepsilon^{(k)}\}$ *is nonnegative square summable.*

(b) *The sequence* $\{\varepsilon^{(k)}\}$ *is nonnegative summable and there exists a* $\bar{\theta} \in (0, 1)$ *such that* $\{(e^{(k)})^{\bar{\theta}}\}$ *is summable, where* $e^{(k)} := \sum_{t=k}^{\infty} (\varepsilon^{(t)})^2$ *for any* $k \geq 0$.

Remarks on Condition 1 (b)

- $\varepsilon^{(k)} = \frac{\bar{\varepsilon}}{(k+1)^{\ell}}$, $\ell > 1 \Rightarrow e^{(k)} = \mathcal{O}(k^{(2\ell-1)}) \Rightarrow \bar{\theta} \in \left(\frac{1}{2\ell-1}, 1\right)$.
- A more intuitive but restrictive alternative: $\sum_{k=1}^{\infty} k(\varepsilon^{(k)})^{2\bar{\theta}} < \infty$.

Sufficient Reduction over Surrogate Sequence

Let $L := \max_i L_i$ and $\Delta\mathbf{x}^{(k)} := \bar{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}$ for any $k \geq 0$.

Proposition 1

Suppose that Assumption 1 holds. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$. Then there exist positive sequences $\{C_0^{(k)}\}$ and $\{C_1^{(k)}\}$ such that, for any $k \geq 0$,*

$$f(\bar{\mathbf{x}}^{(k)}) - f(\bar{\mathbf{x}}^{(k+1)}) \geq C_0^{(k)} \|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\|^2 - C_1^{(k)} \|\Delta\mathbf{x}^{(k)}\|^2 - C_1^{(k+1)} \|\Delta\mathbf{x}^{(k+1)}\|^2.$$

Sufficient Reduction over Surrogate Sequence

Let $L := \max_i L_i$ and $\Delta\mathbf{x}^{(k)} := \bar{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}$ for any $k \geq 0$.

Proposition 1

Suppose that Assumption 1 holds. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$. Then there exist positive sequences $\{C_0^{(k)}\}$ and $\{C_1^{(k)}\}$ such that, for any $k \geq 0$,*

$$f(\bar{\mathbf{x}}^{(k)}) - f(\bar{\mathbf{x}}^{(k+1)}) \geq C_0^{(k)} \|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\|^2 - C_1^{(k)} \|\Delta\mathbf{x}^{(k)}\|^2 - C_1^{(k+1)} \|\Delta\mathbf{x}^{(k+1)}\|^2.$$

Proposition 2

Suppose that Assumptions 1 and 2 hold. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$ and $\{\varepsilon^{(k)}\}$ fulfills Condition 1 (a). Then*

(i) *the sequence $\left\{v^{(k)} := f(\bar{\mathbf{x}}^{(k)}) + u^{(k)} + u^{(k+1)}\right\}$ is well defined, where*

$$u^{(k)} := \sum_{t=k}^{\infty} C_1^{(t)} \|\Delta\mathbf{x}^{(t)}\|^2;$$

Sufficient Reduction over Surrogate Sequence

Let $L := \max_i L_i$ and $\Delta\mathbf{x}^{(k)} := \bar{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}$ for any $k \geq 0$.

Proposition 1

Suppose that Assumption 1 holds. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$. Then there exist positive sequences $\{C_0^{(k)}\}$ and $\{C_1^{(k)}\}$ such that, for any $k \geq 0$,*

$$f(\bar{\mathbf{x}}^{(k)}) - f(\bar{\mathbf{x}}^{(k+1)}) \geq C_0^{(k)}\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\|^2 - C_1^{(k)}\|\Delta\mathbf{x}^{(k)}\|^2 - C_1^{(k+1)}\|\Delta\mathbf{x}^{(k+1)}\|^2.$$

Proposition 2

Suppose that Assumptions 1 and 2 hold. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$ and $\{\varepsilon^{(k)}\}$ fulfills Condition 1 (a). Then*

(i) *the sequence $\left\{v^{(k)} := f(\bar{\mathbf{x}}^{(k)}) + u^{(k)} + u^{(k+1)}\right\}$ is well defined, where*

$$u^{(k)} := \sum_{t=k}^{\infty} C_1^{(t)}\|\Delta\mathbf{x}^{(t)}\|^2;$$

(ii) *for any $k \geq 0$, $v^{(k)} - v^{(k+1)} \geq C_0^{(k)}\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \geq 0$;*

(iii) *the sequence $\{v^{(k)}\}$ converges monotonically to some $\overline{F}$.*

Proposition 3

Suppose that Assumption 1 holds. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$. Then there exist an $\bar{M} > 0$ such that, for any $k \geq 0$,*

$$\operatorname{dist}(0, \partial F(\bar{\mathbf{x}}^{(k+1)})) \leq \bar{M} \left(\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\| + \|\Delta \mathbf{x}^{(k+1)}\| \right),$$

where $\partial F(\bar{\mathbf{x}}^{(k+1)})$ refers to the Fréchet subdifferential of F at $\bar{\mathbf{x}}^{(k+1)}$.

- Stationary point of F: any $\mathbf{x}$ satisfying $0 \in \partial F(\mathbf{x})$.

Proposition 4

Suppose that Assumptions 1 and 2 hold. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$ and $\{\varepsilon^{(k)}\}$ fulfills Condition 1 (a). Then $\{\mathbf{x}^{(k)}\}$ has at least one limit point and each of the limit point is a stationary point of F.*

$\{\varepsilon^{(k)}\} \in \ell^2 \Rightarrow$ **Stationarity of any limit point**

Definition 1 ([Attouch-Bolte 2009])

Let $G : \mathbb{E} \to (-\infty, \infty]$ be proper closed, where $\mathbb{E}$ is an Euclidean space. The function G is said to have the Łojasiewicz property at some stationary point $\bar{\mathbf{x}}$ if there exist $c > 0$, $\theta \in [0, 1)$, and $\eta > 0$ such that

$$|G(\mathbf{x}) - G(\bar{\mathbf{x}})|^{\theta} \leq c \cdot \operatorname{dist}(0, \partial G(\mathbf{x})), \ \forall\, \mathbf{x} \in B_{\eta}(\bar{\mathbf{x}}) := \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \eta\}.$$

We call θ the Łojasiewicz exponent of G at $\bar{\mathbf{x}}$.

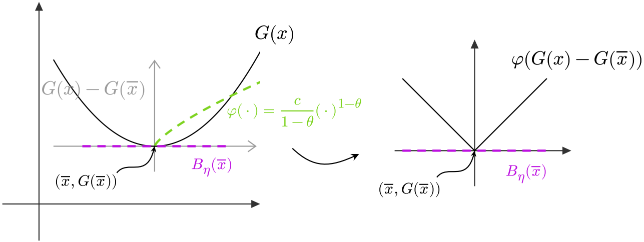

Fig. 1. The illustration of the Łojasiewicz property (drawing with Mathcha).

Definition 1 ([Attouch-Bolte 2009])

Let $G : \mathbb{E} \to (-\infty, \infty]$ be proper closed, where $\mathbb{E}$ is an Euclidean space. The function G is said to have the Łojasiewicz property at some stationary point $\bar{\mathbf{x}}$ if there exist $c > 0$, $\theta \in [0, 1)$, and $\eta > 0$ such that

$$|G(\mathbf{x}) - G(\bar{\mathbf{x}})|^{\theta} \leq c \cdot \mathrm{dist}(0, \partial G(\mathbf{x})), \; \forall \, \mathbf{x} \in B_{\eta}(\bar{\mathbf{x}}) := \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \eta\}.$$

We call θ the Łojasiewicz exponent of G at $\bar{\mathbf{x}}$.

Examples: real-analytic functions, convex functions satisfying certain growth conditions, semialgebraic functions, etc.

Assumption 3

The Łojasiewicz property holds for F at each stationary point.

Theorem 1

Suppose that Assumptions 1, 2, and 3 hold. Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by PALM-I*, where $M_u \geq M_l > L$ and $\{\varepsilon^{(k)}\}$ fulfills Condition 1 (b). Then $\{\mathbf{x}^{(k)}\}$ converges to a stationary point of F.*

Łojasiewicz property + $\{\varepsilon^{(k)}\}$, $\{(e^{(k)})^{\bar{\theta}}\} \in \ell^1 \Rightarrow$ **Iterate convergence**

$$\{\mathbf{x}^{(k)}\} \xleftarrow{\text{sufficient reduction}} \{F^{(k)}\} \xleftarrow{\text{Łojasiewicz property}} \{\partial F^{(k)}\} \xleftarrow{\text{relative error bound}} \{\mathbf{x}^{(k)}\}$$

Why Extra Conditions on $\{\varepsilon^{(k)}\}$ are Required?

Let $\bar{\mathbf{x}}$ be the unique limit point.

Summability of $\{\varepsilon^{(k)}\}$

$$\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \sum_{t=k}^{\infty} \|\bar{\mathbf{x}}^{(t+1)} - \mathbf{x}^{(t)}\| + \sum_{t=k}^{\infty} \|\Delta \mathbf{x}^{(t+1)}\|.$$

Summability of $\{(e^{(k)})^{\bar{\theta}}\}$

- $v^{(k)} - v^{(k+1)} \geq C_0^{(k)} \|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\|^2$, $v^{(k)} = f(\bar{\mathbf{x}}^{(k)}) + u^{(k)} + u^{(k+1)}$.
- $\mathrm{dist}(0, \partial F(\bar{\mathbf{x}}^{(k+1)})) \leq \bar{M} \left(\|\bar{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}\| + \|\Delta \mathbf{x}^{(k+1)}\| \right)$.
- $\mathrm{dist}(0, \partial F(\bar{\mathbf{x}}^{(k+1)})) \xleftrightarrow{\text{Łojasiewicz property}} \left| F(\bar{\mathbf{x}}^{(k+1)}) - \overline{F} \right|^{\theta}$.

Outline

Theorem 2

Suppose that the assumptions in Theorem 1 hold. Let $\bar{\mathbf{x}}$ be the unique limit point of the sequence $\{\mathbf{x}^{(k)}\}$ generated by PALM-I*, where $M_u \geq M_l > L$. Assume that there exists a $K \in \mathbb{N}$ such that $v^{(K)} = \overline{F}$.*

(i) *If $\varepsilon^{(k)} = \bar{\varepsilon}\tilde{\rho}^k$ for any $k \geq 0$, where $\tilde{\rho} \in (0,1)$, then $\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \mathcal{O}(\tilde{\rho}^k)$ for any $k \geq K$.*

(ii) *If $\varepsilon^{(k)} = \frac{\bar{\varepsilon}}{(k+1)^\ell}$ for any $k \geq 0$, where $\ell > 1$, then $\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \mathcal{O}\left(k^{-(\ell-1)}\right)$ for any $k \geq K$.*

Exponentially decreasing errors

Theorem 3

Suppose that the assumptions in Theorem 1 hold with $\varepsilon^{(k)} = \bar{\varepsilon}\tilde{\rho}^k$ *for any* $k \geq 0$, *where* $\tilde{\rho} \in (0,1)$. *Let* $\bar{\mathbf{x}}$ *be the unique limit point of the sequence* $\{\mathbf{x}^{(k)}\}$ *generated by* PALM-I, *where* $M_u \geq M_l > L$, *and* $\theta \in [0,1)$ *be the Łojasiewicz exponent of* F *at* $\bar{\mathbf{x}}$. *Assume that* $v^{(k)} > \overline{F}$ *for any* $k \geq 0$.

(i) *If* $\theta = 0$, *then there exists a* $\rho_1 \in (0,1)$ *such that* $\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \mathcal{O}(\rho_1^k)$ *for all sufficiently large* k.

(ii) *If* $\theta \in (0, \frac{1}{2}]$, *then there exists a* $\rho_2 \in (0,1)$ *such that* $\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \mathcal{O}(\rho_2^k)$ *for all sufficiently large* k.

(iii) *If* $\theta \in (\frac{1}{2}, 1)$, *then* $\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$ *for all sufficiently large* k.

No finite termination for PALM-I when $\theta = 0$

Sublinearly decreasing errors

Theorem 4

Suppose that the assumptions in Theorem 1 hold with $\varepsilon^{(k)} = \frac{\bar{\varepsilon}}{(k+1)^{\ell}}$ *for any* $k \geq 0$, *where* $\ell > 1$. *Let* $\bar{\mathbf{x}}$ *be the unique limit point of the sequence* $\{\mathbf{x}^{(k)}\}$ *generated by* PALM-I, *where* $M_u \geq M_l > L$, *and* $\theta \in [0, 1)$ *be the Łojasiewicz exponent of* F *at* $\bar{\mathbf{x}}$. *Assume that* $v^{(k)} > \overline{F}$ *for any* $k \geq 0$. *Then, for all sufficiently large* k,

$$\|\mathbf{x}^{(k)} - \bar{\mathbf{x}}\| \leq \begin{cases} \mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right), & \text{if } \ell \in \left[\frac{\theta}{2\theta-1}, \infty\right) \text{ and } \theta \in \left(\frac{1}{2}, 1\right); \\ \mathcal{O}\left(k^{-(\ell-1)}\right), & \text{otherwise.} \end{cases}$$

Asymptotic Convergence Rates of PALM-I (Cont.)

Sublinearly decreasing errors

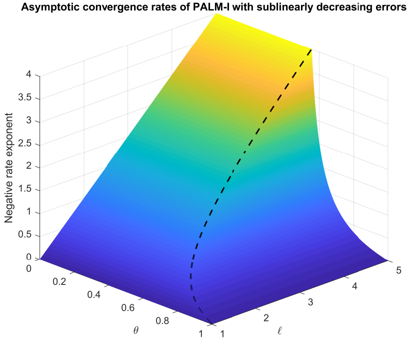

Fig. 2. The surface of the asymptotic convergence rates of PALM-I with sublinearly decreasing errors. The dash line denotes the critical boundary $\ell = \frac{\theta}{2\theta-1}$.

Continuous connection with the known rates

A Review of Asymptotic Rates of PALM

Table 1: Asymptotic convergence rates of PALM under different settings.

θ	$\varepsilon^{(k)}$	Extra assumptions	Rates	References
0	0	-	Finite termination	[Xu-Yin 2013] [Bolte et al. 2014]
	$\tilde{\rho}^k$	$\tilde{\rho} \in (0,1)$	$\mathcal{O}(\rho_1^k)$, where $\rho_1 \in (0,1)$	This work (Theorem 3)
	$\frac{1}{(k+1)^\ell}$	$\ell \in (1,\infty)$	$\mathcal{O}\left(k^{-(\ell-1)}\right)$	This work (Theorem 4)
$(0, \frac{1}{2}]$	0	-	$\mathcal{O}(\rho^k)$, where $\rho \in (0,1)$	[Xu-Yin 2013] [Bolte et al. 2014]
	$\tilde{\rho}^k$	$\tilde{\rho} \in (0,1)$	$\mathcal{O}(\rho_2^k)$, where $\rho_2 \in (0,1)$	This work (Theorem 3)
	$\frac{1}{(k+1)^\ell}$	$\ell \in (1,\infty)$	$\mathcal{O}\left(k^{-(\ell-1)}\right)$	This work (Theorem 4)
$(\frac{1}{2}, 1)$	0	-	$\mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$	[Xu-Yin 2013] [Bolte et al. 2014]
	$\tilde{\rho}^k$	$\tilde{\rho} \in (0,1)$	$\mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$	This work (Theorem 3)
	$\frac{1}{(k+1)^\ell}$	$\ell \in (1,\infty)$	$\mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$ if $\ell \geq \frac{\theta}{2\theta-1}$ $\mathcal{O}\left(k^{-(\ell-1)}\right)$ if $\ell < \frac{\theta}{2\theta-1}$	This work (Theorem 4)

Outline

Optimization with Linear Constraints

$$\begin{aligned}\min_{X} \quad & \sum_{i=2}^{N} \langle X_i, \mathrm{Diag}(\boldsymbol{\varrho}) C \rangle + \sum_{i=2}^{N} \sum_{j>i} \left(\langle X_i, \mathrm{Diag}(\boldsymbol{\varrho}) X_j C \rangle + \beta \langle X_i, X_j \rangle \right), \\ \text{s. t.} \quad & X_i \in \mathcal{S} := \{W \in \mathbb{R}^{K \times K} : W\mathbf{1} = \mathbf{1},\ W^\top \boldsymbol{\varrho} = \boldsymbol{\varrho},\ \mathrm{Tr}(W) = 0,\ W \geq 0\},\ \ \forall\, i.\end{aligned}$$

- Electronic structure calculation. [H. et al. 2023; H.-Liu 2023]
- $\beta \in \mathbb{R}_+$: penalty parameter, $N \in \mathbb{N}$: # electrons, $K \in \mathbb{N}$: # finite elements.
- $\mathcal{T} := \{e_k\}_{k=1}^{K}$: mesh discretizing a bounded domain $\Omega \subseteq \mathbb{R}^d$ ($d \in \{1, 2, 3\}$).
- $\{X_i\}_{i=2}^{N} \subseteq \mathbb{R}^{K \times K}$: transport plans (variables).
- $C \in \mathbb{R}^{K \times K}$: Coulomb cost matrix, defined as

$$C_{ij} := \begin{cases} \|\mathbf{d}_i - \mathbf{d}_j\|^{-1}, & \text{if } i \neq j; \\ 0, & \text{otherwise,} \end{cases}$$

  where $\{\mathbf{d}_k\}_{k=1}^{K} \subseteq \mathbb{R}^d$ are the barycenters of elements $\{e_k\}_{k=1}^{K}$.
- $\boldsymbol{\varrho} := [\varrho_1, \ldots, \varrho_K]^\top \in \mathbb{R}^K$: discretized density vector, defined as

$$\varrho_k := \frac{1}{N} \int_{e_k} \rho(\mathbf{r})\, \mathrm{d}\mathbf{r},\ \ k \in [K],$$

  where $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is the single-particle density.

Experimental Settings

Problem data generation

- $d = 1$, $\beta = 1$, $\Omega = [-1, 1]$, $N = 3$, $K = 36$ (# variables $= 2592$).
- $\rho(x) \propto \exp(-x^2/\sqrt{\pi})$, for any $x \in \mathbb{R}$.
- Equi-mass discretization $\Rightarrow$ all the entries in $\boldsymbol{\varrho}$ are identical.
- Closed-form optimal solution $X^\star$. [Colombo et al. 2015]

Running environment

- CPU: Intel Xeon Gold 6242R CPU @ 3.10GHz.
- RAM: 510GB.
- Operating system: Ubuntu 20.04.
- Software: MATLAB R2018b.

Experimental Settings (Cont.)

Algorithms in comparison – PALM-E, PALM-I

- Proximal parameters: $\sigma_i^{(k)} \equiv \sigma = 10^{-2}, i = 2, \ldots, N, k \geq 0.$
- Subsolver: semismooth Newton-CG (**SSNCG**). [Li et al. 2020]
- Tolerance sequence for SSNCG:
  - PALM-E: $\{\varepsilon^{(k)} \equiv 10^{-7}\}$.
  - PALM-I: $\{\varepsilon^{(k)} = \max\{\frac{10^{-1}}{(k+1)^{\ell}}, 10^{-7}\}\}$ with $\ell = 0.75$.
- Stopping criterion: relative KKT violation[1)] is smaller than 10^{-6}.

Metrics: KKT violation, CPU time in seconds (s), objective value.

[1)] Based upon Assumption 2, the stationarity can be characterized by the KKT conditions.

Numerical Results with Random Initialization

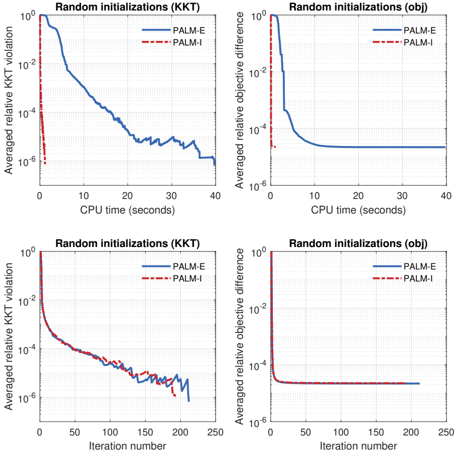

Fig. 3. Averaged relative KKT violation and objective difference with random initialization.

- 100 initial points randomly sampled from standard uniform distribution.
- Average CPU time:
  - PALM-E: 15.97s.
  - PALM-I: 0.46s.
- Solutions with similar qualities.

Numerical Results with Good Initialization

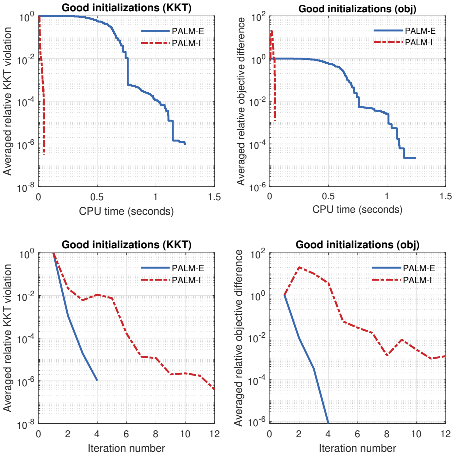

Fig. 4. Averaged relative KKT violation and objective difference with good initialization.

- 100 initial points generated by random perturbation around $X^\star$ (absolute entrywise deviation $\leq 10^{-3}$).
- Average CPU time:
  - PALM-E: 0.85s.
  - PALM-I: 0.03s.
- $\left|f(X^{\text{PALM-I}}) - f(X^\star)\right| \leq 10^{-5}$ on $\geq 75\%$ samples.

Optimization with Nonlinear Constraints

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2} \langle \mathbf{x}, A\mathbf{x} \rangle + \langle \mathbf{b}, \mathbf{x} \rangle, \\ \text{s. t.} \quad & \frac{1}{2} \langle \mathbf{x}_i, B_i \mathbf{x}_i \rangle + \langle \mathbf{c}_i, \mathbf{x}_i \rangle \leq 1, \ \forall\, i \in [n]. \end{aligned}$$

- 3D anisotropic friction contact, topology optimization, polynomial optimization, sensor network localization, ...
  [Kučera 2008; He et al. 2010; Luo-Zhang 2010; Li et al. 2012; Liu et al. 2020; ...]
- $m_i = m \in \mathbb{N}$, $\mathbf{x}_i \in \mathbb{R}^m$, for any $i \in [n]$.
- $A \in \mathbb{S}^{mn}$, $\{B_i\}_{i=1}^n \subseteq \mathbb{S}^m_{++}$.
- $\mathbf{b} \in \mathbb{R}^{mn}$, $\{\mathbf{c}_i\}_{i=1}^n \subseteq \mathbb{R}^m$.

Experimental Settings

Problem data generation

- $n = 5$, $m = 500$ (# variables $= 2500$).
- A, $\mathbf{b}$: randomly sampled from standard normal distribution.
- For any $i \in [n]$, $B_i = (b_{i,jk})$, where [Dai 2006; Jia et al. 2017]

$$b_{i,jk} = \begin{cases} 10^{\frac{j-1}{m-1}\mathrm{ncond}_i}, & \text{if } j = k; \\ 0, & \text{otherwise}, \end{cases}$$

  and $\{\mathrm{ncond}_i\}_{i=1}^n = \{3.00, 3.25, 3.50, 3.75, 4.00\}$.
- For any $i \in [n]$, $\mathbf{c}_i = 0 \Rightarrow$ all ellipsoids are concentric.

Running environment

- CPU: Intel Xeon Gold 6242R CPU @ 3.10GHz.
- RAM: 510GB.
- Operating system: Ubuntu 20.04.
- Software: MATLAB R2018b.

Experimental Settings (Cont.)

Algorithms in comparison – PALM-E, PALM-F, PALM-I

- Proximal parameters: $\sigma_i^{(k)} \equiv \sigma = 1$, $i \in [n]$, $k \geq 0$.
- Subsolvers: [Dai 2006; Jia et al. 2017]
  - PALM-F: hybrid projection algorithm (**HP**).
  - PALM-E/I: self-adaptive alternating direction method of multiplier (**S-ADMM**).
- Inexact criterion for HP: [Hua-Yamashita 2016]

$$\|\mathbf{x}_i^{(k+1)} - \tilde{\mathbf{x}}_i^{(k)} + \lambda_{i,\text{HP}}^{(k+1)}(B_i \mathbf{x}_i^{(k+1)} + \mathbf{c}_i)\| \leq \frac{\eta}{2}\|\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}\|_\infty,$$

  where $\eta = 0.99\sigma$, $\lambda_{i,\text{HP}}^{(k+1)}$ can be determined by least squares.
- Tolerance sequence for S-ADMM:
  - PALM-E: $\{\varepsilon^{(k)} \equiv 10^{-6}\}$.
  - PALM-I: $\{\varepsilon^{(k)} = \max\{\frac{10^{-1}}{(k+1)^\ell}, 10^{-6}\}\}$ with $\ell = 0.75$.
- Stopping criterion: relative KKT violation[2)] is smaller than 10^{-5}.

Metrics: KKT violation, CPU time in seconds (s), objective value.

2) Based upon Assumption 2, the stationarity can also be characterized by the KKT conditions.

Numerical Results

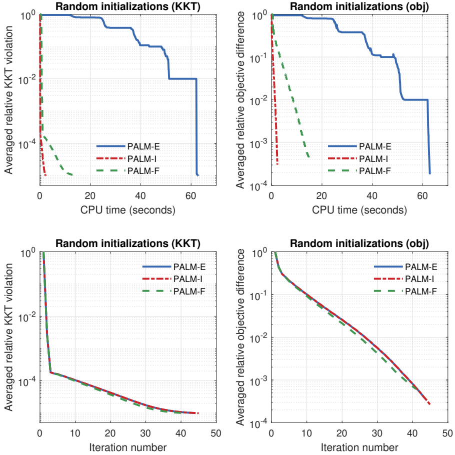

Fig. 5. Averaged relative KKT violation and objective difference with random initialization.

- 100 initial points randomly sampled from standard normal distribution.
- Average CPU time:
  - PALM-E: 30.15s.
  - PALM-F: 9.26s.
  - PALM-I: 1.74s.
- $\left|f(\mathbf{x}^{\text{PALM-I}}) - f(\mathbf{x}^{\text{PALM-E}})\right| \ll \left|f(\mathbf{x}^{\text{PALM-F}}) - f(\mathbf{x}^{\text{PALM-E}})\right|$.
- $f(\mathbf{x}^{\text{PALM-I}}) < f(\mathbf{x}^{\text{PALM-F}})$ on 95% samples.

Outline

Conclusions and Future Work

Conclusions

- Indispensability of infeasible subsolvers in PALM.
- Better performances of PALM-I compared with PALM-E and PALM-F.
- Shortage and limitations of existing works on PALM-I.
- Convergence properties of PALM-I with implementable inexact criteria.
  - Tools: surrogate sequence + computable error bound + Łojasiewicz property.
  - $\{\varepsilon^{(k)}\} \in \ell^2 \Rightarrow$ stationarity of any limit point.
  - $\{\varepsilon^{(k)}\}, \{(e^{(k)})^{\bar{\theta}}\} \in \ell^1 \Rightarrow$ iterate convergence to stationary point.
  - Asymptotic rates with exponentially and sublinearly decreasing errors.

Future work

- Weaker assumptions & Adaptive inexact criteria.
- Accelerated versions (inertia, momentum, ...).
- Nonsmoothness & Nonconvexity.

References

- **H.** and X. Liu. The convergence properties of infeasible inexact proximal alternating linearized minimization. *Sci. China Math.*, 2023, DOI: 10.1007/s11425-022-2074-7. (***Main reference***)
- **H.**, H. Chen, and X. Liu. A global optimization approach for multimarginal optimal transport problems with Coulomb cost. *SIAM J. Sci. Comput.*, 2023, 45(3): A1214–A1238.
- **H.** and X. Liu. The exactness of the ℓ_1 penalty function for a class of mathematical programs with generalized complementarity constraints. *Fundam. Res.*, 2023, DOI: 10.1016/j.fmre.2023.04.006.
- J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, 2014, 146(1): 459–494.
- P. Frankel, G. Garrigos, and J. Peypouquet. Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates. *J. Optim. Theory Appl.*, 2015, 165(3): 874–900.

Thanks for your attention!

Email: ykhu@lsec.cc.ac.cn