Proximal algorithms for a class of nonconvex nonsmooth minimization problems involving piecewise smooth and min-weakly-convex functions

Jan Harold Alcantara
Institute of Statistical Sciences, Academia Sinica

A joint work with Ching-pei Lee

January 18, 2022

Outline

- Introduction
- Min-convex optimization
- Acceleration Methods
- Application to LCP
- Numerical Results

Problem formulation

We consider the problem

$$\min_{w \in \mathbb{E}} \; f(w) + g(w) - h(w)$$

where $f, g, h : \mathbb{E} \to (-\infty, +\infty]$ and $\mathbb{E}$ is a Euclidean space.

Problem formulation

We consider the problem

$$\min_{w\in\mathbb{E}} \; f(w) + g(w) - h(w)$$

where $f, g, h : \mathbb{E} \to (-\infty, +\infty]$ and $\mathbb{E}$ is a Euclidean space.

What is the "usual" setting[1] considered?

- f is convex and has L-Lipschitz continuous gradient.
- g is proper closed and convex.
- h is continuous convex.

[1] Wen, B. Chen, X. and Pong, T.K. A proximal difference-of-convex algorithm with extrapolation. *Computational Optimization and Applications*, 69:297–324, 2018.

Proximal difference-of-convex algorithm (pDCA)[1]

pDCA algorithm

$$w^{k+1} = \operatorname{prox}_{\lambda g}\left(w^k - \frac{1}{L}\nabla f(w^k) + \frac{1}{L}\xi^k\right)$$

where $\xi^k \in \partial h(w^k)$ and

$$\operatorname{prox}_{\lambda g}(w) := \operatorname*{arg\,min}_{z\in\mathbb{E}}\left\{g(z) + \frac{1}{2\lambda}\|z - w\|^2\right\}.$$

Proximal difference-of-convex algorithm (pDCA)[1]

pDCA algorithm

$$w^{k+1} = \operatorname{prox}_{\lambda g}\left(w^k - \frac{1}{L}\nabla f(w^k) + \frac{1}{L}\xi^k\right)$$

where $\xi^k \in \partial h(w^k)$ and

$$\operatorname{prox}_{\lambda g}(w) := \operatorname*{arg\,min}_{z\in\mathbb{E}} \left\{ g(z) + \frac{1}{2\lambda}\|z - w\|^2 \right\}.$$

Questions

Can we extend this to possibly nondifferentiable f?

How about to nonconvex functions f, g and h?

Outline

- Introduction
- Min-convex optimization
- Acceleration Methods
- Application to LCP
- Numerical Results

ρ-convex functions

Definition (ρ-convex function)

A function F is said to be ρ-convex if $F(w) - \frac{\rho}{2}\|w\|^2$ is a convex function.

F is said to be

- weakly convex if $\rho < 0$
- convex if $\rho \geq 0$
- strongly convex if $\rho > 0$.

min-ρ-convex functions

Definition (min-ρ-convex function)

We say that $g : \mathbb{E} \to (-\infty, +\infty]$ is a min-ρ-convex function if there exist an index set J with $|J| < \infty$, and ρ-convex, proper closed functions $g_j : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$, $j \in J$, such that

$$g(w) = \min_{j \in J} g_j(w), \quad \forall w \in \mathbb{E}.$$

min-ρ-convex functions

Definition (min-ρ-convex function)

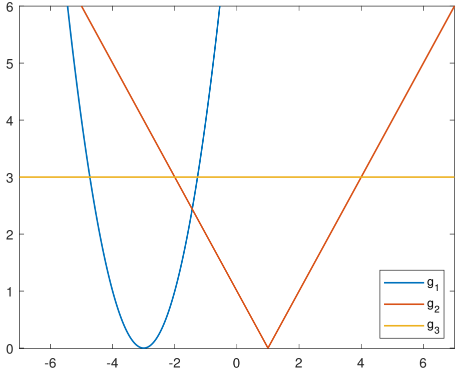
We say that $g : \mathbb{E} \to (-\infty, +\infty]$ is a min-ρ-convex function if there exist an index set J with $|J| < \infty$, and ρ-convex, proper closed functions $g_j : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$, $j \in J$, such that

$$g(w) = \min_{j \in J} g_j(w), \quad \forall w \in \mathbb{E}.$$

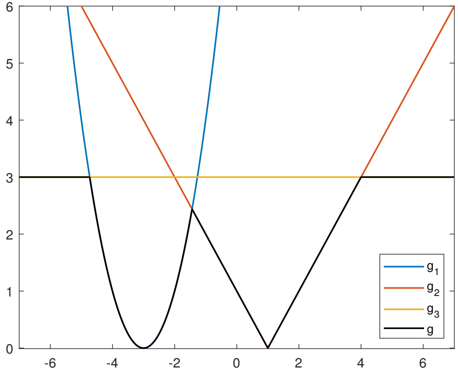
We call g

- min-weakly convex if $\rho < 0$
- min-convex if $\rho \geq 0$
- min-strongly convex if $\rho > 0$.

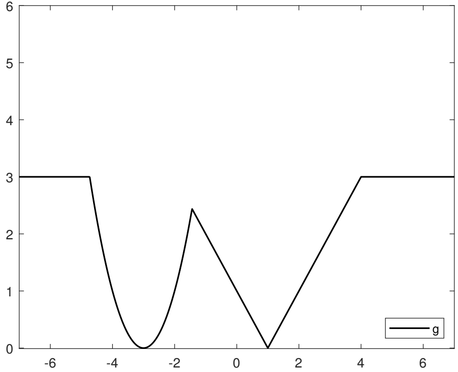
A min-convex function



A min-convex function



A min-convex function



$$\min_{w \in \mathbb{E}} f(w) + g(w) - h(w)$$

Assumption A

3 g is a min-ρ-convex function.

$$\min_{w \in \mathbb{E}} f(w) + g(w) - h(w)$$

Assumption A

1. The functions f, g and h are expressible as

$$f = \min_{i \in I} f_i, \quad g = \min_{j \in J} g_j, \quad \text{and} \quad h = \max_{m \in M} h_m,$$

where I, J and M are **finite** index sets.

3. g is a min-ρ-convex function.

$$\min_{w\in\mathbb{E}} f(w)+g(w)-h(w)$$

Assumption A

1. The functions f, g and h are expressible as

$$f=\min_{i\in I} f_i, \quad g=\min_{j\in J} g_j, \quad \text{and} \quad h=\max_{m\in M} h_m,$$

   where I, J and M are **finite** index sets.
2. $\forall i\in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. g is a min-ρ-convex function.

$$\min_{w \in \mathbb{E}} f(w) + g(w) - h(w)$$

Assumption A

1. The functions f, g and h are expressible as

$$f = \min_{i \in I} f_i, \quad g = \min_{j \in J} g_j, \quad \text{and} \quad h = \max_{m \in M} h_m,$$

   where I, J and M are **finite** index sets.
2. $\forall i \in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. g is a min-ρ-convex function.
4. $\forall m \in M$, h_m is a C^1 convex function on $\mathbb{E}$.

$$\min_{w \in \mathbb{E}} f(w) + g(w) - h(w)$$

Assumption A

1. The functions f, g and h are expressible as

$$f = \min_{i \in I} f_i, \quad g = \min_{j \in J} g_j, \quad \text{and} \quad h = \max_{m \in M} h_m,$$

   where I, J and M are **finite** index sets.
2. $\forall i \in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. g is a min-ρ-convex function.
4. $\forall m \in M$, h_m is a C^1 convex function on $\mathbb{E}$.
5. $\forall (i, j, m) \in I \times J \times M$, $f_i + g_j - h_m$ is coercive over $\mathbb{E}$.

Remarks

1. f is not necessarily smooth, but is piecewise smooth.

Remarks

1. f is not necessarily smooth, but is piecewise smooth.
2. g is not necessarily convex.

Remarks

1. f is not necessarily smooth, but is piecewise smooth.
2. g is not necessarily convex.
   $\mathrm{prox}_{\lambda g}$ is single-valued for any $w \in \mathbb{E}$ and $\lambda \in (0, \bar{\lambda})$ where

$$\bar{\lambda} = \begin{cases} -1/\rho & \text{if } \rho < 0, \\ +\infty & \text{if } \rho \geq 0. \end{cases}$$

Remarks

1. f is not necessarily smooth, but is piecewise smooth.
2. g is not necessarily convex.
   $\mathrm{prox}_{\lambda g}$ is single-valued for any $w \in \mathbb{E}$ and $\lambda \in (0, \bar{\lambda})$ where
   $$\bar{\lambda} = \begin{cases} -1/\rho & \text{if } \rho < 0, \\ +\infty & \text{if } \rho \geq 0. \end{cases}$$
3. h is a convex piecewise smooth function.

How do we solve such a problem under these assumptions?

Some notations . . .

How do we solve such a problem under these assumptions?

Some notations . . .

1. Given $S \subseteq \mathbb{E}$, $\operatorname{prox}_{\lambda g}(S) := \bigcup_{w \in S} \operatorname{prox}_{\lambda g}(w)$.

How do we solve such a problem under these assumptions?

Some notations ...

1. Given $S \subseteq \mathbb{E}$, $\operatorname{prox}_{\lambda g}(S) := \bigcup_{w \in S} \operatorname{prox}_{\lambda g}(w)$.
2. We denote $f' : \mathbb{E} \rightrightarrows \mathbb{E}$ is defined by

$$f'(w) := \{\nabla f_i(w) : i \in I \text{ such that } f(w) = f_i(w)\},$$

How do we solve such a problem under these assumptions?

Some notations . . .

1. Given $S \subseteq \mathbb{E}$, $\mathrm{prox}_{\lambda g}(S) := \bigcup_{w \in S} \mathrm{prox}_{\lambda g}(w)$.
2. We denote $f' : \mathbb{E} \rightrightarrows \mathbb{E}$ is defined by

$$f'(w) := \{\nabla f_i(w) : i \in I \text{ such that } f(w) = f_i(w)\},$$

and similarly, $h' : \mathbb{E} \rightrightarrows \mathbb{E}$ is given by

$$h'(w) := \{\nabla h_m(w) : m \in M \text{ such that } h(w) = h_m(w)\}.$$

Proximal difference-of-min-convex algorithm (PDMC)

PDMC algorithm (A. & Lee, 2022)

$$w^{k+1} \in \operatorname{prox}_{\lambda g}\left(w^k - \lambda f'(w^k) + \lambda h'(w^k)\right), \qquad \text{(PDMC)}$$

where $\lambda \in (0, \bar{\lambda}) \cap (0, 1/L]$, and $L := \max_{i \in I} L_i$

Proximal difference-of-min-convex algorithm (PDMC)

PDMC algorithm (A. & Lee, 2022)

$$w^{k+1} \in \operatorname{prox}_{\lambda g}\left(w^k - \lambda f'(w^k) + \lambda h'(w^k)\right), \qquad \text{(PDMC)}$$

where $\lambda \in (0, \bar{\lambda}) \cap (0, 1/L]$, and $L := \max_{i \in I} L_i$

What can we say about the convergence of this algorithm?

Global convergence to critical points

Theorem (A. & Lee, 2022)

Let $\{w^k\}$ *be any sequence generated by* (PDMC) *with* $\lambda \in (0, \min\{\bar{\lambda}, 1/L\})$.
If Assumption A holds, then $\{w^k\}$ *is bounded and its accumulation points are critical points*[2] *of* $f + g - h$.

[2] We say that w is a critical point if $0 \in \partial f(w) + \partial g(w) - \partial h(w)$.

Special cases

Define $T^{\lambda} : \mathbb{E} \rightrightarrows \mathbb{E}$ by

$$T^{\lambda}(w) := \operatorname{prox}_{\lambda g}\left(w - \lambda f'(w) + \lambda h'(w)\right)$$

Special cases

Define $T^{\lambda}: \mathbb{E} \rightrightarrows \mathbb{E}$ by

$$T^{\lambda}(w) := \mathrm{prox}_{\lambda g}\left(w - \lambda f'(w) + \lambda h'(w)\right)$$

Full convergence

If w^* is an accumulation point and T^{λ} is single-valued at w^*, then $w^k \to w^*$ under any of the following conditions:

1. each $Id - \lambda \nabla f_i$ and ∇h_m are nonexpansive and g_j is ρ-convex with $\rho \geq 1$, or
2. each $Id - \lambda \nabla f_i$ is nonexpansive, $h \equiv 0$ and g_j is ρ-convex with $\rho \geq 0$,

with local linear rate if $\rho > 1$ and $\rho > 0$, respectively.

Local linear rate also holds when

3 $h \equiv 0$, $g_j = \delta_{R_j}$ and each $Id - \lambda \nabla f_i$ is a contraction over R_j, where each R_j is a convex set[4].

[4] In this case, g_j is a ρ-convex function with $\rho = 0$.

Local linear rate also holds when

3. $h \equiv 0$, $g_j = \delta_{R_j}$ and each $Id - \lambda \nabla f_i$ is a contraction over R_j, where each R_j is a convex set[4].

Remark

1. For case 2, PDMC reduces to a generalized forward-backward algorithm.
2. For case 3, PDMC simplifies to a generalized projected subgradient algorithm.

[4] In this case, g_j is a ρ-convex function with $\rho = 0$.

Outline

- Introduction
- Min-convex optimization
- Acceleration Methods
- Application to LCP
- Numerical Results

Acceleration method 1: Extrapolation

We do extrapolation if consecutive iterates activate the same piece of $f + g - h$.

$$\chi_k := \begin{cases} 1 & \text{if } w^k \;\&\; w^{k-1} \text{ activate the same piece and } k \geq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

Algorithm 1: Accelerated proximal difference-of-min-convex algorithm

Let $\phi = f + g - h$. Choose $\sigma > 0$, $\lambda \in (0, 1/L] \cap (0, \bar{\lambda})$, and $w^0 \in \mathbb{E}$.

Set $w^{-1} = w^0$ and $k = 0$.

Step 1. Set $z^k = w^k + t_k \chi_k p^k$, where $p^k = w^k - w^{k-1}$, $t_k \geq 0$ satisfies

$$\phi(z^k) \leq \phi(w^k) - \frac{\sigma}{2} t_k^2 \chi_k^2 \|p^k\|^2, \tag{2}$$

and χ_k is given by (1).

Step 2. Set $w^{k+1} \in T^\lambda(z^k)$, $k = k + 1$, and go back to Step 1.

Acceleration method 2: Component identification

Algorithm 2: Proximal difference-of-min-convex algorithm with component identification

Choose $w^0 \in \mathbb{E}$, $N \in \mathbb{N}$. Set Unchanged $= 0$, $k = 0$.

Step 1. Set Unchanged $= \chi_k(\text{Unchanged} + 1)$

Step 2. Compute w^{k+1} according to the following rule:

2.1 If Unchanged $< N$: set $w^{k+1} \in T^\lambda(w^k)$. Terminate if $w^{k+1} \in \text{Fix}(T^\lambda)$; otherwise, set $k = k+1$ and go back to Step 1.

2.2 If Unchanged $= N$: pick (i, j, m) activated by w^k, and solve

$$w^{k+1} \in \arg\min_{z \in \mathbb{E}} f_i(z) + g_j(z) - h_m(z). \tag{3}$$

Terminate if $w^{k+1} \in \text{Fix}(T^\lambda)$; otherwise, set Unchanged $= -1$, $w^{k+1} = w^k$, $k = k+1$, and go back to Step 1.

Outline

Application: Linear complementarity problem

- Consider the linear complementarity problem (LCP): Find $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad Mx - d \geq 0, \quad \text{and} \quad \langle x, Mx - d \rangle = 0, \tag{LCP}$$

where $M \in \mathbb{R}^{n \times n}$ and $d \in \mathbb{R}^n$.

Application: Linear complementarity problem

- Consider the linear complementarity problem (LCP): Find $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad Mx - d \geq 0, \quad \text{and} \quad \langle x, Mx - d \rangle = 0, \tag{LCP}$$

where $M \in \mathbb{R}^{n \times n}$ and $d \in \mathbb{R}^n$.

- Let $y := Mx - d$. Then (LCP) is equivalent to

$$\begin{cases} x \geq 0, \quad y \geq 0, \quad \langle x, y \rangle = 0 \\ Mx - y = d. \end{cases}$$

Application: Linear complementarity problem

- Consider the linear complementarity problem (LCP): Find $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad Mx - d \geq 0, \quad \text{and} \quad \langle x, Mx - d \rangle = 0, \qquad \text{(LCP)}$$

where $M \in \mathbb{R}^{n \times n}$ and $d \in \mathbb{R}^n$.

- Let $y := Mx - d$. Then (LCP) is equivalent to

$$\begin{cases} x \geq 0, \quad y \geq 0, \quad \langle x, y \rangle = 0 \\ Mx - y = d. \end{cases}$$

We denote $w := (x, y)$.

Feasibility reformulation of LCP

$$\text{Find } w \in S_1 \cap S_2$$

where

$$S_1 = \{w \in \mathbb{R}^{2n} : Tw = d\} \quad \text{where } T := [M \ \ -I_n]$$
$$S_2 = \{w \in \mathbb{R}^{2n} : w_j \geq 0, \ w_{n+j} \geq 0, \text{and } w_j w_{n+j} = 0 \ \forall j \in [n]\}.$$

Feasibility reformulation of LCP

$$\text{Find } w \in S_1 \cap S_2$$

where

$$S_1 = \{w \in \mathbb{R}^{2n} : Tw = d\} \quad \text{where } T := [M \;\; -I_n]$$
$$S_2 = \{w \in \mathbb{R}^{2n} : w_j \geq 0,\; w_{n+j} \geq 0, \text{and } w_j w_{n+j} = 0 \; \forall j \in [n]\}.$$

1. S_1 is an affine set, and therefore convex.
2. S_2 is nonconvex, but can be expressed as a finite union of closed convex sets (called a union convex set[5])

[5] Dao, M.N. and Tam, M.K.. Union averaged operators with applications to proximal algorithms for min-convex functions. *J. Optim. Theory Appl.*, 181:61–94, 2019.

Example

Let $n = 1$ so that

$$S_2 = \{(w_1, w_2) : w_1, w_2 \geq 0 \quad \text{and} \quad w_1 w_2 = 0\}.$$

Example

Let $n = 1$ so that

$$S_2 = \{(w_1, w_2) : w_1, w_2 \geq 0 \quad \text{and} \quad w_1 w_2 = 0\}.$$

Then $S_1 = R_1 \cup R_2$ where

$$R_1 = \{(a, 0) : a \geq 0\}$$
$$R_2 = \{(0, b) : b \geq 0\}$$

From feasibility reformulation to optimization problem

From feasibility reformulation to optimization problem

The following are equivalent:

1. $w \in S_1 \cap S_2$
2. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\operatorname{dist}(w, S_2)^2 = 0$
3. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \delta_{S_2}(w) = 0.$

From feasibility reformulation to optimization problem

The following are equivalent:

1. $w \in S_1 \cap S_2$
2. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\operatorname{dist}(w, S_2)^2 = 0$
3. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \delta_{S_2}(w) = 0.$

Merit functions

$$f + g - h$$

1. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\|w\|^2 - \left(\frac{1}{2}\|w\|^2 - \frac{1}{2}\operatorname{dist}(w, S_2)^2\right)$
2. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\operatorname{dist}(w, S_2)^2 - 0$
3. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \delta_{S_2}(w) - 0$

From feasibility reformulation to optimization problem

The following are equivalent:

1. $w \in S_1 \cap S_2$
2. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\operatorname{dist}(w, S_2)^2 = 0$
3. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \delta_{S_2}(w) = 0.$

Merit

Do these functions f, g and h satisfy Assumption A?

1. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\|w\|^2 - \left(\frac{1}{2}\|w\|^2 - \frac{1}{2}\operatorname{dist}(w, S_2)^2\right)$
2. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \frac{1}{2}\operatorname{dist}(w, S_2)^2 - 0$
3. $\frac{1}{2}\operatorname{dist}(w, S_1)^2 + \delta_{S_2}(w) - 0$

Recall...

$$\min_{w \in \mathbb{E}} f(w) + g(w) - h(w)$$

Assumptions A1-A4

1. $f = \min_{i \in I} f_i$, $g = \min_{j \in J} g_j$, and $h = \max_{m \in M} h_m$, where $|I|, |J|, |M| < \infty$
2. $\forall i \in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. $\forall j \in J$, g_j is a ρ-convex function.
4. $\forall m \in M$, h_m is a C^1 convex function on $\mathbb{E}$.

Illustration: Merit Function 2

$$\underbrace{0.5\,\mathrm{dist}(w, S_1)^2}_{f(w)} + \underbrace{0.5\,\mathrm{dist}(w, S_2)^2}_{g(w)} - \underbrace{0}_{h(w)}$$

Illustration: Merit Function 2

$$\underbrace{0.5\,\mathrm{dist}(w,S_1)^2}_{f(w)} + \underbrace{0.5\,\mathrm{dist}(w,S_2)^2}_{g(w)} - \underbrace{0}_{h(w)}$$

1. Clearly, f and h satisfy Assumption A2 and A4.

Illustration: Merit Function 2

$$\underbrace{0.5\,\mathrm{dist}(w, S_1)^2}_{f(w)} + \underbrace{0.5\,\mathrm{dist}(w, S_2)^2}_{g(w)} - \underbrace{0}_{h(w)}$$

1. Clearly, f and h satisfy Assumption A2 and A4.
2. Since S_2 is a union convex set, then

$$S_2 = \bigcup_{j \in J} R_j.$$

Illustration: Merit Function 2

$$\underbrace{0.5\,\mathrm{dist}(w,S_1)^2}_{f(w)} + \underbrace{0.5\,\mathrm{dist}(w,S_2)^2}_{g(w)} - \underbrace{0}_{h(w)}$$

1. Clearly, f and h satisfy Assumption A2 and A4.
2. Since S_2 is a union convex set, then

$$S_2 = \bigcup_{j\in J} R_j.$$

Thus,

$$g(w) = \frac{1}{2}\mathrm{dist}(w,S_2)^2 = \min_{j\in J}\frac{1}{2}\mathrm{dist}(w,R_j)^2 =: \min_{j\in J} g_j(w).$$

where each g_j is convex. A3 is satisfied!

(Complete) Assumption A

1. $f = \min_{i \in I} f_i$, $g = \min_{j \in J} g_j$, and $h = \max_{m \in M} h_m$, where $|I|, |J|, |M| < \infty$ where I, J and M are finite index sets.
2. $\forall i \in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. $\forall j \in J$, g_j is a ρ-convex function.
4. $\forall m \in M$, h_m is a C^1 convex function on $\mathbb{E}$.
5. For all $(i, j, m) \in I \times J \times M$, the function $f_i + g_j - h_m$ is coercive over $\mathbb{E}$.

(Complete) Assumption A

1. $f = \min_{i \in I} f_i$, $g = \min_{j \in J} g_j$, and $h = \max_{m \in M} h_m$, where $|I|, |J|, |M| < \infty$ where I, J and M are finite index sets.
2. $\forall i \in I$, f_i has L_i-Lipschitz continuous gradient on $\mathbb{E}$.
3. $\forall j \in J$, g_j is a ρ-convex function.
4. $\forall m \in M$, h_m is a C^1 convex function on $\mathbb{E}$.
5. For all $(i, j, m) \in I \times J \times M$, the function $f_i + g_j - h_m$ is coercive over $\mathbb{E}$.

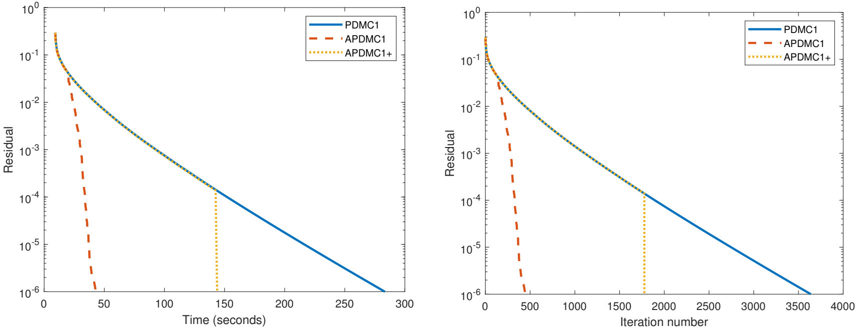
Remark

For the LCP, Assumption A5 holds when M is a P-matrix (A. & Lee, 2022).

Outline

- Introduction
- Min-convex optimization
- Acceleration Methods
- Application to LCP
- Numerical Results

Merit Function 1



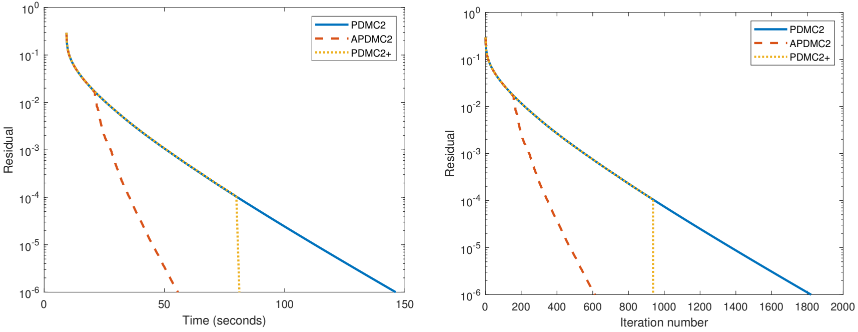
(a) Time

(b) Iterations

Figure: Non-accelerated and accelerated PDMC for Merit Function 1 for solving a standard LCP.[7]

[7]Kanzow, C. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4):851–868, 1996.

Merit Function 2

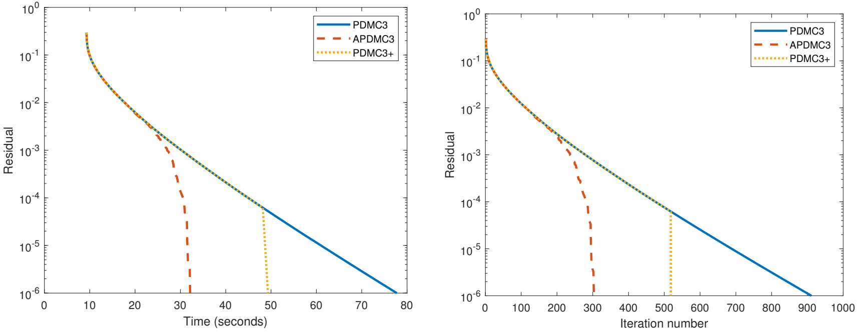


(a) Time (b) Iterations

Figure: Non-accelerated and accelerated PDMC for Merit Function 2 for solving a standard LCP.[7]

[7] Kanzow, C. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4):851–868, 1996.

Merit Function 3



(a) Time

(b) Iterations

Figure: Non-accelerated and accelerated PDMC for Merit Function 3 for solving a standard LCP.[7]

[7] Kanzow, C. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4):851–868, 1996.

Thank you for listening!

Some references

- Jan Harold Alcantara & Ching-pei Lee. Global convergence and acceleration of fixed point iterations of union upper semicontinuous operators: proximal algorithms, alternating and averaged nonconvex projections, and linear complementarity problems, 2022.
- Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. The Linear Complementarity Prob- lem. Academic Press, New York, NY, 1992.
- Minh N. Dao and Matthew K. Tam. Union averaged operators with applications to proximal algorithms for min-convex functions. *J. Optim. Theory Appl.*, 181:61–94, 2019.
- Christian Kanzow. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4):851–868, 1996.
- R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317 of Grundlehren der MathematischenWissenschaften. Springer, Berlin, 1998.
- Bo Wen, Xiaojun Chen, and Ting Kei Pong. A proximal difference-of-convex algorithm with extrapolation. *Computational Optimization and Applications*, 69:297–324, 2018.

Critical points

For any function F, its *subdifferential*[2] at w is

$$\partial F(w) :=$$

$$\limsup_{\bar{w}\to w, F(\bar{w})\to F(w)} \left(\hat{\partial} F(\bar{w}) := \{v : v \in \mathbb{E}, h(z) \geq h(w) + \langle v, z - w\rangle + o(\|z-w\|)\}\right),$$

Definition[3]

We say that w is a critical point of $f + g - h$ if

$$0 \in \partial f(w) + \partial g(w) - \partial h(w).$$

[2] Rockafellar, R.T. and Wets, R.J. *Variational Analysis*, volume 317 of Grundlehren der Mathematischen Wissenschaften. Springer, Berlin, 1998.

[3] Coincides with the definition of critical point of Wen et. al. in the "usual" setting.