A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems

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

International INFORMS 2018
June 2018
(Joint work with Tianxiang Liu and Akiko Takeda)

Motivating applications

Inducing simultaneous structures:

- Nonconvex fused regularized problems:

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|^2 + c_1\|x\|_1 + c_2\sum_{i=1}^{n-1}|x_{i+1}-x_i|^{\frac{1}{2}}.$$

- Simultaneous low rank and sparse matrix optimization problems:

$$\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{m\times n}} & \frac{1}{2}\|X-M\|_F^2 \\ \text{Subject to} & \mathrm{rank}(X)\le k,\ \|\mathrm{vec}(X)\|_0 \le s. \end{array}$$

Motivating applications

Inducing simultaneous structures:

- Nonconvex fused regularized problems:

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\|Ax - b\|^2 + c_1\|x\|_1 + c_2 \sum_{i=1}^{n-1} |x_{i+1} - x_i|^{\frac{1}{2}}.$$

- Simultaneous low rank and sparse matrix optimization problems:

$$\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & \frac{1}{2}\|X - M\|_F^2 \\ \text{Subject to} & \mathrm{rank}(X) \leq k, \ \|\mathrm{vec}(X)\|_0 \leq s. \end{array}$$

Other variants: $\frac{1}{2}\|P_\Omega(X - M)\|_F^2$, where Ω corresponds to known / observed entries.

General model

$$\min_{x\in\mathbb{R}^n} f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x),$$

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ is an L-smooth function;
- A_i, $i = 1, \ldots, m$, are linear maps;
- P_i, $i = 0, \ldots, m$, are nonnegative proper closed functions and are continuous in their domains.
- the sets $\mathrm{dom}\, P_i$, $i = 1, \ldots, m$, are closed, and

$$\mathrm{dom}\, P_0 \cap \bigcap_{i=1}^{m} A_i^{-1} \mathrm{dom}\, P_i \neq \emptyset.$$

General model

$$\min_{x \in \mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x),$$

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ is an L-smooth function;
- A_i, $i = 1, \ldots, m$, are linear maps;
- P_i, $i = 0, \ldots, m$, are nonnegative proper closed functions and are continuous in their domains.
- the sets $\mathrm{dom}\, P_i$, $i = 1, \ldots, m$, are closed, and

$$\mathrm{dom}\, P_0 \cap \bigcap_{i=1}^{m} A_i^{-1} \mathrm{dom}\, P_i \neq \emptyset.$$

- $f + P_0$ is level-bounded.

General model cont.

$$\min_{x \in \mathbb{R}^n} f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

Assumption cont.: An element of $\mathrm{prox}_{\lambda P_i}(x)$ is easy to compute for all $\lambda > 0$, $x \in \mathbb{R}^n$ and $i = 0, \ldots, m$, where

$$\mathrm{prox}_{\lambda P_i}(x) := \operatorname*{Arg\,min}_{y \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} \|y - x\|^2 + P_i(y) \right\}.$$

General model cont.

$$\min_{x\in\mathbb{R}^n} f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

Assumption cont.: An element of $\mathrm{prox}_{\lambda P_i}(x)$ is easy to compute for all $\lambda > 0$, $x \in \mathbb{R}^n$ and $i = 0, \ldots, m$, where

$$\mathrm{prox}_{\lambda P_i}(x) := \operatorname*{Arg\,min}_{y\in\mathbb{R}^n} \left\{ \frac{1}{2\lambda}\|y - x\|^2 + P_i(y) \right\}.$$

Idea: make use of variants of proximal gradient algorithm?

General model cont.

$$\min_{x \in \mathbb{R}^n} f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

Assumption cont.: An element of $\mathrm{prox}_{\lambda P_i}(x)$ is easy to compute for all $\lambda > 0$, $x \in \mathbb{R}^n$ and $i = 0, \ldots, m$, where

$$\mathrm{prox}_{\lambda P_i}(x) := \operatorname*{Arg\,min}_{y \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} \|y - x\|^2 + P_i(y) \right\}.$$

Idea: make use of variants of proximal gradient algorithm?

Not trivial! The proximal mapping of $x \mapsto P_0(x) + \sum_{i=1}^{m} P_i(A_i x)$ is in general difficult to compute.

Existing approaches

$$\min_{x \in \mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

Existing approaches

$$\min_{x\in\mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):

Existing approaches

$$\min_{x\in\mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):
  - ⋆ Introducing extra variables $y_i = A_i x$ and λ_i.
  - ⋆ Each iteration involves the proximal mapping of P_i and a multiplier update.

Existing approaches

$$\min_{x \in \mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):
  - ⋆ Introducing extra variables $y_i = A_i x$ and λ_i.
  - ⋆ Each iteration involves the proximal mapping of P_i and a multiplier update.
  - ⋆ Convergence not guaranteed when $m > 0$ and for general linear maps.

Existing approaches

$$\min_{x \in \mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):
  - ⋆ Introducing extra variables $y_i = A_i x$ and λ_i.
  - ⋆ Each iteration involves the proximal mapping of P_i and a multiplier update.
  - ⋆ Convergence not guaranteed when $m > 0$ and for general linear maps.
- Proximal averaging (Yu, Zheng '15):

Existing approaches

$$\min_{x\in\mathbb{R}^n}\; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):
  - ⋆ Introducing extra variables $y_i = A_i x$ and λ_i.
  - ⋆ Each iteration involves the proximal mapping of P_i and a multiplier update.
  - ⋆ Convergence not guaranteed when $m > 0$ and for general linear maps.
- Proximal averaging (Yu, Zheng '15):
  All P_i have to be Lipschitz continuous, $A_i = I$. Essentially **only** solving a smooth approximation.

Existing approaches

$$\min_{x \in \mathbb{R}^n} \; f(x) + P_0(x) + \sum_{i=1}^{m} P_i(A_i x).$$

When P_i's are possibly nonconvex:

- Alternating direction method of multipliers (Hong et al. '16):
  - ⋆ Introducing extra variables $y_i = A_i x$ and λ_i.
  - ⋆ Each iteration involves the proximal mapping of P_i and a multiplier update.
  - ⋆ Convergence not guaranteed when $m > 0$ and for general linear maps.
- Proximal averaging (Yu, Zheng '15):
  All P_i have to be Lipschitz continuous, $A_i = I$. Essentially **only** solving a smooth approximation.

Question: How to develop an approach with convergence guarantee based solely on computing proximal mappings of P_i and ∇f?

Key ideas I

When P_i's are all convex:

- For each $\lambda > 0$, the Moreau envelope

$$e_\lambda P_i(x) := \inf_{y\in\mathbb{R}^n}\left\{\frac{1}{2\lambda}\|y-x\|^2 + P_i(y)\right\}$$

is convex and **smooth**, with $\nabla e_\lambda P_i(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda P_i}(x))$, and

$$\|\nabla e_\lambda P_i(x) - \nabla e_\lambda P_i(y)\| \le \lambda^{-1}\|x-y\|$$

for all $x, y \in \mathbb{R}^n$.

Key ideas I

When P_i's are all convex:

- For each $\lambda > 0$, the Moreau envelope
$$e_\lambda P_i(x) := \inf_{y \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} \|y - x\|^2 + P_i(y) \right\}$$
is convex and **smooth**, with $\nabla e_\lambda P_i(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda P_i}(x))$, and
$$\|\nabla e_\lambda P_i(x) - \nabla e_\lambda P_i(y)\| \le \lambda^{-1} \|x - y\|$$
for all $x, y \in \mathbb{R}^n$.
- The function $f(x) + \sum_{i=1}^m e_\lambda P_i(A_i x) + P_0(x)$ can be minimized by variants of the proximal gradient algorithm efficiently.
- Nesterov's smoothing technique (Nesterov '05); the basis of popular software TFOCS (Becker, Candès, Grant '11).

Key ideas I

When P_i's are all convex:

- For each $\lambda > 0$, the Moreau envelope

$$e_\lambda P_i(x) := \inf_{y \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} \|y - x\|^2 + P_i(y) \right\}$$

  is convex and **smooth**, with $\nabla e_\lambda P_i(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda P_i}(x))$, and

$$\|\nabla e_\lambda P_i(x) - \nabla e_\lambda P_i(y)\| \le \lambda^{-1} \|x - y\|$$

  for all $x, y \in \mathbb{R}^n$.
- The function $f(x) + \sum_{i=1}^m e_\lambda P_i(A_i x) + P_0(x)$ can be minimized by variants of the proximal gradient algorithm efficiently.
- Nesterov's smoothing technique (Nesterov '05); the basis of popular software TFOCS (Becker, Candès, Grant '11).

Question: Using Moreau envelope for nonconvex P_i?

Key ideas II

When P_i's are possibly nonconvex:

- For each $\lambda > 0$, the Moreau envelope is in general **not smooth**,

Key ideas II

When P_i's are possibly nonconvex:

- For each $\lambda > 0$, the Moreau envelope is in general **not smooth**, but it is a difference-of-convex (DC) function:

$$e_\lambda P_i(x) = \frac{1}{2\lambda}\|x\|^2 - \underbrace{\sup_{y \in \operatorname{dom} P_i} \left\{ \frac{1}{\lambda}\langle x, y\rangle - \frac{1}{2\lambda}\|y\|^2 - P_i(y) \right\}}_{h_i(x)}.$$

Key ideas II

When P_i's are possibly nonconvex:

- For each $\lambda > 0$, the Moreau envelope is in general **not smooth**, but it is a difference-of-convex (DC) function:

$$e_\lambda P_i(x) = \frac{1}{2\lambda}\|x\|^2 - \underbrace{\sup_{y\in \operatorname{dom} P_i}\left\{\frac{1}{\lambda}\langle x, y\rangle - \frac{1}{2\lambda}\|y\|^2 - P_i(y)\right\}}_{h_i(x)}.$$

  Moreover,

$$\frac{1}{\lambda}\operatorname{prox}_{\lambda P_i}(x) \subseteq \partial h_i(x).$$

- The function $f(x) + \sum_{i=1}^m e_\lambda P_i(A_i x) + P_0(x)$ can be minimized by variants of DC/majorization-based algorithm efficiently.

Algorithm: subproblem

To minimize $F_\lambda(x) := f(x) + \sum_{i=1}^m e_\lambda P_i(A_i x) + P_0(x)$:

Algorithm 1: NPG$_{\rm major}$

Step 0. Input $x^0 \in \operatorname{dom} P_0$, $L_{\max} \geq L_{\min} > 0$, $\tau > 1$, $c > 0$ and an integer $M \geq 0$. Set $t = 0$.

Step 1. Choose any $L_t^0 \in [L_{\min}, L_{\max}]$ and set $L_t = L_t^0$.

1a) Pick $u \in \operatorname{prox}_{L_t^{-1} P_0}\left(x^t - \frac{1}{L_t}\left[\nabla f(x^t) + \frac{1}{\lambda}\sum_{i=1}^m A_i^*(A_i x^t - \operatorname{prox}_{\lambda P_i}(A_i x^t))\right]\right)$.

1b) Go to to **Step 2)** if

$$F_\lambda(u) \leq \max_{[t-M]_+ \leq i \leq t} F_\lambda(x^i) - \frac{c}{2}\|u - x^t\|^2.$$

Else, set $L_t \leftarrow \tau L_t$ and go to **Step 1a)**.

Step 2. Set $\bar{L}_t = L_t$, $x^{t+1} = u$, $t = t + 1$. Go to **Step 1**.

Properties of $\text{NPG}_{\text{major}}$

Theorem 1. (Liu, P., Takeda '18)
Let $\{x^t\}$ be the sequence generated by $\text{NPG}_{\text{major}}$. Then

1. $F_\lambda(x^t) \leq F_\lambda(x^0)$ for all $t \geq 0$.
2. $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$.

Properties of $\mathrm{NPG}_{\mathrm{major}}$

Theorem 1. (Liu, P., Takeda '18)
Let $\{x^t\}$ be the sequence generated by $\mathrm{NPG}_{\mathrm{major}}$. Then

1. $F_\lambda(x^t) \leq F_\lambda(x^0)$ for all $t \geq 0$.
2. $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$.
3. It holds that

$$\lim_{t\to\infty} \mathrm{dist}\left(0, \nabla f(x^t) + \partial P_0(x^{t+1}) + \sum_{i=1}^{m} \frac{A_i^*(A_i x^t - \mathrm{prox}_{\lambda P_i}(A_i x^t))}{\lambda}\right) = 0.$$

Successive DC approximation method

Algorithm 2: SDCAM

Step 0. Pick $\epsilon_\nu \downarrow 0$ and $\lambda_\nu \downarrow 0$. Set $\nu = 0$. Pick an $x^0 \in \mathrm{dom}\, P_0$ and

$$x^{\text{feas}} \in \mathrm{dom}\, P_0 \cap \bigcap_{i=1}^{m} A_i^{-1} \mathrm{dom}\, P_i.$$

Step 1. If $F_{\lambda_\nu}(x^\nu) < F_{\lambda_\nu}(x^{\text{feas}})$, set $x^{\nu,0} = x^\nu$. Else, set $x^{\nu,0} = x^{\text{feas}}$.

Step 2. Apply $\text{NPG}_{\text{major}}$ to $F_{\lambda_\nu}(x)$ starting at $x^{\nu,0}$. Terminate at x^{ν,l_ν} when $\|x^{\nu,l_\nu+1} - x^{\nu,l_\nu}\| \le \epsilon_\nu$, $F_{\lambda_\nu}(x^{\nu,l_\nu}) \le F_{\lambda_\nu}(x^{\nu,0})$, and

$$\mathrm{dist}\left(0, \nabla f(x^{\nu,l_\nu}) + \partial P_0(x^{\nu,l_\nu+1}) + \sum_{i=1}^m \frac{1}{\lambda_\nu} A_i^*[A_i x^{\nu,l_\nu} - \mathrm{prox}_{\lambda_\nu P_i}(A_i x^{\nu,l_\nu})]\right) \le \epsilon_\nu.$$

Step 3. Update $x^{\nu+1} = x^{\nu,l_\nu}$ and $\nu = \nu + 1$. Go to **Step 1**.

Convergence of SDCAM

Theorem 2. (Liu, P., Takeda '18)
Let $\{x^t\}$ be the sequence generated by SDCAM. Then $\{x^t\}$ is bounded. Let x^* be an accumulation point of this sequence. Then:

(i) It holds that $x^* \in \mathrm{dom}\, P_0 \cap \bigcap_{i=1}^m A_i^{-1}\mathrm{dom}\, P_i$.

(ii) Suppose the following condition holds:

$$y_0 + \sum_{i=1}^m A_i^* y_i = 0 \;\&\; y_0 \in \partial^\infty P_0(x^*), \;\; y_i \in \partial^\infty P_i(A_i x^*), \; \forall\, i = 1, \ldots, m$$

$$\Longrightarrow \;\; y_i = 0 \;\; \forall\, i = 0, \ldots, m.$$

Then

$$0 \in \nabla f(x^*) + \partial P_0(x^*) + \sum_{i=1}^m A_i^* \partial P_i(A_i x^*).$$

Convergence of SDCAM

Theorem 2. (Liu, P., Takeda '18)
Let $\{x^t\}$ be the sequence generated by SDCAM. Then $\{x^t\}$ is bounded. Let x^* be an accumulation point of this sequence. Then:

(i) It holds that $x^* \in \mathrm{dom}\, P_0 \cap \bigcap_{i=1}^m A_i^{-1}\mathrm{dom}\, P_i$.

(ii) Suppose the following condition holds:

$$y_0 + \sum_{i=1}^m A_i^* y_i = 0 \;\&\; y_0 \in \partial^\infty P_0(x^*), \quad y_i \in \partial^\infty P_i(A_i x^*), \;\forall\, i = 1, \ldots, m$$
$$\Longrightarrow \; y_i = 0 \;\; \forall\, i = 0, \ldots, m.$$

Then
$$0 \in \nabla f(x^*) + \partial P_0(x^*) + \sum_{i=1}^m A_i^* \partial P_i(A_i x^*).$$

Remark: The condition in (ii) holds if all $A_i = I$ and all except one P_i are locally Lipschitz.

Simulations: fused regularization

$$\min_{x\in\mathbb{R}^n} \; \frac{1}{2}\|x-b\|^2 + c_1\|x\|_1 + c_2\sum_{i=1}^{n-1}|x_{i+1}-x_i|^{\frac{1}{2}}.$$

- b is noisy measurement of a sparse piecewise constant signal.
- Set $\lambda_\nu = 0.1^{\nu+1}$ in SDCAM; terminate when $\lambda_\nu < 10^{-9}$.
- $\mathrm{NPG}_{\mathrm{major}}$ for subproblems is terminated when successive changes are small.

Simulations: fused regularization

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|x-b\|^2 + c_1\|x\|_1 + c_2\sum_{i=1}^{n-1}|x_{i+1}-x_i|^{\frac{1}{2}}.$$

- b is noisy measurement of a sparse piecewise constant signal.
- Set $\lambda_\nu = 0.1^{\nu+1}$ in SDCAM; terminate when $\lambda_\nu < 10^{-9}$.
- $\mathrm{NPG_{major}}$ for subproblems is terminated when successive changes are small.
- Compare with NPG for a smooth approximation based on $(s^2+\lambda_\nu^2)^{\frac{1}{4}} \approx |s|^{\frac{1}{2}}$. (sNPG) $x^{\rm feas}$ is not used.
Terminate when $\lambda_\nu < 10^{-8}$.

Simulations: fused regularization

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\|x - b\|^2 + c_1\|x\|_1 + c_2 \sum_{i=1}^{n-1} |x_{i+1} - x_i|^{\frac{1}{2}}.$$

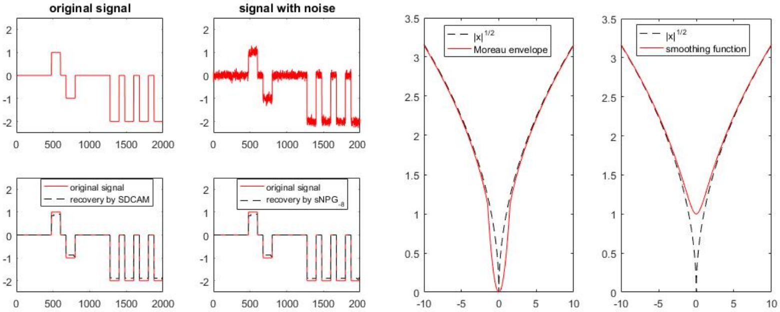
- b is noisy measurement of a sparse piecewise constant signal.
- Set $\lambda_\nu = 0.1^{\nu+1}$ in SDCAM; terminate when $\lambda_\nu < 10^{-9}$.
- $\mathrm{NPG}_{\mathrm{major}}$ for subproblems is terminated when successive changes are small.
- Compare with NPG for a smooth approximation based on $(s^2 + \lambda_\nu^2)^{\frac{1}{4}} \approx |s|^{\frac{1}{2}}$. (sNPG) x^{feas} is not used.
  Terminate when $\lambda_\nu < 10^{-8}$.
- All codes are run in Matlab R2016a on a 64-bit PC with an Intel(R) Core(TM) i7-6700 CPU (3.41GHz) and 32GB of RAM.

Simulations: fused regularization

Table: Results for SDCAM and sNPG, $c_1 = c_2 = \sigma\sqrt{n}/40$.

n	iter		CPU		fval	
	SDCAM	sNPG	SDCAM	sNPG	SDCAM	sNPG
2000	27796	23968	5.7	9.1	1.7728e2	1.7729e2
4000	41686	42336	16.9	28.2	4.9592e2	4.9593e2
6000	45573	46124	25.5	39.5	8.4943e2	8.4939e2
8000	49089	39759	34.5	42.9	1.3216e3	1.3215e3
10000	45320	48645	45.2	64.6	1.6587e3	1.6586e3

Simulations: fused regularization



Simulations: low rank and sparse matrix

$$\begin{array}{ll} \min_{X \in \mathbb{R}^{m\times n}} & \frac{1}{2}\|X - M\|_F^2 \\ \text{Subject to} & \mathrm{rank}(X) \le k, \ \|\mathrm{vec}(X)\|_0 \le s. \end{array}$$

- $M = M_1 M_2 + \sigma\Delta$, where $M_1 \in \mathbb{R}^{m\times k}$, $M_2 \in \mathbb{R}^{k\times n}$, and $m/10$ random rows of M_1 are zero.

Simulations: low rank and sparse matrix

$$\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m\times n}} & \frac{1}{2}\|X - M\|_F^2 \\ \text{Subject to} & \mathrm{rank}(X) \leq k, \ \|\mathrm{vec}(X)\|_0 \leq s. \end{array}$$

- $M = M_1 M_2 + \sigma\Delta$, where $M_1 \in \mathbb{R}^{m\times k}$, $M_2 \in \mathbb{R}^{k\times n}$, and $m/10$ random rows of M_1 are zero.
- Apply SDCAM with $P_0 = \delta_{\mathrm{rank}(\cdot)\leq k}$ (SDCAM$_r$), or with $P_0 = \delta_{\|\mathrm{vec}(\cdot)\|_0 \leq s}$ (SDCAM$_s$).
- NPG$_{\mathrm{major}}$ for subproblems is terminated when successive changes are small.
- Terminate (SDCAM$_r$) when distance to being s-sparse is small.
- Terminate (SDCAM$_s$) when distance to having rank at most k is small.

Simulations: low rank and sparse matrix

Table: Comparison of SDCAM_r and SDCAM_s, $k = 10$, $s = 0.1mn$, $n = 500$.

σ	m	CPU		vio	
		SDCAM_r	SDCAM_s	SDCAM_r	SDCAM_s
0.005	1000	4.7	378.1	4.7569e-4	1.0515e-4
	2000	4.0	647.0	6.7084e-4	1.5247e-4
	3000	6.0	862.8	8.2038e-4	1.8857e-4
0.010	1000	379.3	529.2	9.4347e-5	2.1032e-4
	2000	653.6	912.6	1.3412e-4	3.0580e-4
	3000	969.5	1080.6	1.6434e-4	3.7701e-4
0.020	1000	413.7	769.2	1.8985e-4	4.2222e-4
	2000	675.5	1251.3	2.6849e-4	6.1136e-4
	3000	1003.5	2043.0	3.2804e-4	7.5510e-4

Conclusion

- We make use of the fact that Moreau envelopes are difference-of-convex (DC) to construct a sequence of "DC" subproblems.
- These subproblems can be solved by variants of DC algorithm.
- Convergence to stationary points of the original problem is established under mild assumptions.

Reference:

- T. Liu, T. K. Pong and A. Takeda. *A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems*. Available at *https://arxiv.org/abs/1710.05778*.

Thanks for coming!