

A smoothing moving balls approximation method

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

DOML 2026
July 2026
(Joint work with Nung-sing Sze and Jiefeng Xu)

Conic optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

where

- $\mathcal{K} \subset \mathbb{Y}$ is a closed convex **pointed** cone, and $\text{int } \mathcal{K} \neq \emptyset$;
- $f : \mathbb{X} \rightarrow \mathbb{R}$ and $G : \mathbb{X} \rightarrow \mathbb{Y}$ are L_f -smooth and L_G -smooth for some $L_f > 0$ and $L_G \geq 0$;
- $P_1 : \mathbb{X} \rightarrow \mathbb{R}$ and $P_2 : \mathbb{X} \rightarrow \mathbb{R}$ are convex;

Conic optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

where

- $\mathcal{K} \subset \mathbb{Y}$ is a closed convex **pointed** cone, and $\text{int } \mathcal{K} \neq \emptyset$;
- $f : \mathbb{X} \rightarrow \mathbb{R}$ and $G : \mathbb{X} \rightarrow \mathbb{Y}$ are L_f -smooth and L_G -smooth for some $L_f > 0$ and $L_G \geq 0$;
- $P_1 : \mathbb{X} \rightarrow \mathbb{R}$ and $P_2 : \mathbb{X} \rightarrow \mathbb{R}$ are convex;
- $\exists \hat{x} \in \mathbb{X}$ s.t. $G(\hat{x}) \in \text{int } \mathcal{K}$ and $\Omega_0 := \{x \in G^{-1}\mathcal{K} : \psi(x) \leq \psi(\hat{x})\}$ is bounded;
- It holds that

$$0 \in \text{int} (G(x) + DG(x)\mathbb{X} - \mathcal{K}) \quad \forall x \in \Omega_0.$$

This means the RCQ holds in Ω_0 .

Conic optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

where

- $\mathcal{K} \subset \mathbb{Y}$ is a closed convex **pointed** cone, and $\text{int } \mathcal{K} \neq \emptyset$;
- $f : \mathbb{X} \rightarrow \mathbb{R}$ and $G : \mathbb{X} \rightarrow \mathbb{Y}$ are L_f -smooth and L_G -smooth for some $L_f > 0$ and $L_G \geq 0$;
- $P_1 : \mathbb{X} \rightarrow \mathbb{R}$ and $P_2 : \mathbb{X} \rightarrow \mathbb{R}$ are convex;
- $\exists \hat{x} \in \mathbb{X}$ s.t. $G(\hat{x}) \in \text{int } \mathcal{K}$ and $\Omega_0 := \{x \in G^{-1}\mathcal{K} : \psi(x) \leq \psi(\hat{x})\}$ is bounded;
- It holds that

$$0 \in \text{int} (G(x) + DG(x)\mathbb{X} - \mathcal{K}) \quad \forall x \in \Omega_0.$$

This means the RCQ holds in Ω_0 .

Focus: Large-scale problems

Conic optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

where

- $\mathcal{K} \subset \mathbb{Y}$ is a closed convex **pointed** cone, and $\text{int } \mathcal{K} \neq \emptyset$;
- $f : \mathbb{X} \rightarrow \mathbb{R}$ and $G : \mathbb{X} \rightarrow \mathbb{Y}$ are L_f -smooth and L_G -smooth for some $L_f > 0$ and $L_G \geq 0$;
- $P_1 : \mathbb{X} \rightarrow \mathbb{R}$ and $P_2 : \mathbb{X} \rightarrow \mathbb{R}$ are convex;
- $\exists \hat{x} \in \mathbb{X}$ s.t. $G(\hat{x}) \in \text{int } \mathcal{K}$ and $\Omega_0 := \{x \in G^{-1}\mathcal{K} : \psi(x) \leq \psi(\hat{x})\}$ is bounded;
- It holds that

$$0 \in \text{int} (G(x) + DG(x)\mathbb{X} - \mathcal{K}) \quad \forall x \in \Omega_0.$$

This means the RCQ holds in Ω_0 .

Focus: Large-scale problems — First-order methods

Conic optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

where

- $\mathcal{K} \subset \mathbb{Y}$ is a closed convex **pointed** cone, and $\text{int } \mathcal{K} \neq \emptyset$;
- $f : \mathbb{X} \rightarrow \mathbb{R}$ and $G : \mathbb{X} \rightarrow \mathbb{Y}$ are **L_f -smooth** and **L_G -smooth** for some $L_f > 0$ and $L_G \geq 0$;
- $P_1 : \mathbb{X} \rightarrow \mathbb{R}$ and $P_2 : \mathbb{X} \rightarrow \mathbb{R}$ are convex;
- $\exists \hat{x} \in \mathbb{X}$ s.t. $G(\hat{x}) \in \text{int } \mathcal{K}$ and $\Omega_0 := \{x \in G^{-1}\mathcal{K} : \psi(x) \leq \psi(\hat{x})\}$ is bounded;
- It holds that

$$0 \in \text{int} (G(x) + DG(x)\mathbb{X} - \mathcal{K}) \quad \forall x \in \Omega_0.$$

This means the RCQ holds in Ω_0 .

Focus: Large-scale problems — First-order methods — we are interested in **feasible** methods.

Moving balls approximation method

When $\mathcal{K} = \mathbb{R}_-^m$: In this case, $G(x) \in \mathcal{K}$ reduces to $g_i(x) \leq 0$ for all i .

Moving balls approximation method

When $\mathcal{K} = \mathbb{R}_-^m$: In this case, $G(x) \in \mathcal{K}$ reduces to $g_i(x) \leq 0$ for all i .

- Moving balls approximation (MBA) method (Auslender, Shefi, Teboulle '10) when $P_1 = P_2 = 0$: Let $x^0 = \hat{x}$. For $k = 0, 1, \dots$

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

Moving balls approximation method

When $\mathcal{K} = \mathbb{R}^m_-$: In this case, $G(x) \in \mathcal{K}$ reduces to $g_i(x) \leq 0$ for all i .

- Moving balls approximation (MBA) method (Auslender, Shefi, Teboulle '10) when $P_1 = P_2 = 0$: Let $x^0 = \hat{x}$. For $k = 0, 1, \dots$

$$x^{k+1} = \arg \min \{ \langle \nabla f(x^k), x - x^k \rangle + (L_f/2) \|x - x^k\|^2 \}$$
$$\text{s.t. } g_i(x^k) + \langle \nabla g_i(x^k), x - x^k \rangle + (L_{g_i}/2) \|x - x^k\|^2 \leq 0 \quad \forall i.$$

- Convergence analyzed under various assumptions (Auslender, Shefi, Teboulle '10, Bolte, Pauwels '16, Bolte, Chen, Pauwels '20, Yu, P. Lu '21), assuming exactly solved subproblems.
- Recent works Boob, Deng, Lan '25, Liu, Pan, Bi '25 analyzed convergence under inexact criteria for subproblem solvers.

Moving balls approximation method

When $\mathcal{K} = \mathbb{R}^m$: In this case, $G(x) \in \mathcal{K}$ reduces to $g_i(x) \leq 0$ for all i .

- Moving balls approximation (MBA) method (Auslender, Shefi, Teboulle '10) when $P_1 = P_2 = 0$: Let $x^0 = \hat{x}$. For $k = 0, 1, \dots$

$$x^{k+1} = \arg \min \{ \langle \nabla f(x^k), x - x^k \rangle + (L_f/2) \|x - x^k\|^2 \}$$
$$\text{s.t. } g_i(x^k) + \langle \nabla g_i(x^k), x - x^k \rangle + (L_{g_i}/2) \|x - x^k\|^2 \leq 0 \quad \forall i.$$

- Convergence analyzed under various assumptions (Auslender, Shefi, Teboulle '10, Bolte, Pauwels '16, Bolte, Chen, Pauwels '20, Yu, P. Lu '21), assuming exactly solved subproblems.
- Recent works Boob, Deng, Lan '25, Liu, Pan, Bi '25 analyzed convergence under inexact criteria for subproblem solvers.

Motivating questions:

- Design an MBA variant that admits easy subproblems?
- Design an MBA variant for general $\mathcal{K}$?

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{array}$$

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{array}$$

Ideas:

- Apply MBA?

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{aligned} & \underset{x \in \mathbb{X}}{\text{Minimize}} && f(x) + P_1(x) - P_2(x) \\ & \text{Subject to} && \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{aligned}$$

Ideas:

- Apply MBA? $\sigma_{\mathcal{B}}$ is nonsmooth ☹.

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{array}$$

Ideas:

- Apply MBA? $\sigma_{\mathcal{B}}$ is nonsmooth ☹. **Smoothing?**

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{array}$$

Ideas:

- Apply MBA? $\sigma_{\mathcal{B}}$ is nonsmooth ☹. **Smoothing?**
- Subproblems involve only **one inequality constraint** — intuitively, are easy to solve ☺.

MBA for conic programs

Reformulation: Since $\mathcal{K}$ is a closed convex **pointed** cone with $\text{int } \mathcal{K} \neq \emptyset$, $\mathcal{K}^\circ$ admits a compact base $\mathcal{B}$.

Recall that a convex set $\mathcal{D}$ is called a base of a cone $\mathcal{Q}$ if $0 \notin \text{cl}(\mathcal{D})$ and $\mathcal{Q} = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$.

Then we have

- $y \in \mathcal{K} \iff \sigma_{\mathcal{B}}(y) \leq 0$;
- $y \in \text{int } \mathcal{K} \iff \sigma_{\mathcal{B}}(y) < 0$.

Our problem is equivalently reformulated as

$$\begin{aligned} & \underset{x \in \mathbb{X}}{\text{Minimize}} && f(x) + P_1(x) - P_2(x) \\ & \text{Subject to} && \sigma_{\mathcal{B}}(G(x)) \leq 0. \end{aligned}$$

Ideas:

- Apply MBA? $\sigma_{\mathcal{B}}$ is nonsmooth ☹. **Smoothing?**
- Subproblems involve only **one inequality constraint** — intuitively, are easy to solve ☺.

Questions: How to smooth? How to **evolve** the smoothing?

Smoothing: Assumptions

Definition: (Adapted from Beck, Teboulle '12)

We say that a convex function $h: \mathbb{Y} \rightarrow \mathbb{R}$ is $(\alpha_1, \alpha_2, \alpha_3)$ -majorizingly smoothable for some $\alpha_1 \geq 0$, $\alpha_2 > 0$ and $\alpha_3 > 0$, if $\exists$ a family $\{h_\mu\}_{\mu>0}$, referred to as a **majorizing smoothing approximation** (MSA) of h , such that:

- (i) h_μ is convex, differentiable, and ∇h_μ is $(\alpha_1 + \alpha_2/\mu)$ -Lipschitz continuous, for every $\mu > 0$;
- (ii) it holds that

$$h(y) \leq h_\mu(y) \leq h(y) + \mu\alpha_3 \quad \forall y \in \mathbb{Y}, \mu > 0.$$

Smoothing: Assumptions

Definition: (Adapted from Beck, Teboulle '12)

We say that a convex function $h: \mathbb{Y} \rightarrow \mathbb{R}$ is $(\alpha_1, \alpha_2, \alpha_3)$ -majorizingly smoothable for some $\alpha_1 \geq 0$, $\alpha_2 > 0$ and $\alpha_3 > 0$, if $\exists$ a family $\{h_\mu\}_{\mu>0}$, referred to as a **majorizing smoothing approximation** (MSA) of h , such that:

- (i) h_μ is convex, differentiable, and ∇h_μ is $(\alpha_1 + \alpha_2/\mu)$ -Lipschitz continuous, for every $\mu > 0$;
- (ii) it holds that

$$h(y) \leq h_\mu(y) \leq h(y) + \mu\alpha_3 \quad \forall y \in \mathbb{Y}, \mu > 0.$$

Assumption for smoothing:

The σ_B admits an MSA $\{h_\mu\}_{\mu>0}$ with parameters $(\alpha_1, \alpha_2, \alpha_3)$, and $\exists \alpha_4 > 0$ s.t.

$$h_{\mu_1}(y) \leq h_{\mu_0}(y) - \alpha_4(\mu_0 - \mu_1) \quad \forall y \in \mathbb{Y}, \mu_0 > \mu_1 > 0.$$

Smoothing: Examples

Examples of MSA: (Beck, Teboulle '12)

$\mathcal{K}$	$\sigma_{\mathcal{B}}(y)$	$h_{\mu}(y)$	$(\alpha_1, \alpha_2, \alpha_3)$
$\mathbb{R}_+^m$	$\max_{1 \leq i \leq m} y_i$	$\mu \log \left(\sum_{i=1}^m e^{y_i/\mu} \right)$	$(0, 1, \log(m))$
$\mathcal{S}_+^m$	$\lambda_{\max}(y)$	$\mu \log \left(\sum_{i=1}^m e^{\lambda_i(y)/\mu} \right)$	$(0, 1, \log(m))$

Smoothing: Examples

Examples of MSA: (Beck, Teboulle '12)

$\mathcal{K}$	$\sigma_{\mathcal{B}}(y)$	$h_{\mu}(y)$	$(\alpha_1, \alpha_2, \alpha_3)$
$\mathbb{R}_-^m$	$\max_{1 \leq i \leq m} y_i$	$\mu \log \left(\sum_{i=1}^m e^{y_i/\mu} \right)$	$(0, 1, \log(m))$
$\mathcal{S}_-^m$	$\lambda_{\max}(y)$	$\mu \log \left(\sum_{i=1}^m e^{\lambda_i(y)/\mu} \right)$	$(0, 1, \log(m))$

Proposition 1: (Xu, P., Sze '25)

Let $\{\bar{h}_{\mu}\}$ be a *nondecreasing* MSA (i.e., $\bar{h}_{\mu_1} \leq \bar{h}_{\mu_0}$ pointwise whenever $\mu_0 \geq \mu_1 > 0$). Then $h_{\mu} := \bar{h}_{\mu} + \alpha_4 \mu$ is an MSA with parameters $(\alpha_1, \alpha_2, \alpha_3 + \alpha_4)$ and satisfies the **assumption for smoothing**.

Smoothing: Examples

Examples of MSA: (Beck, Teboulle '12)

$\mathcal{K}$	$\sigma_{\mathcal{B}}(y)$	$h_{\mu}(y)$	$(\alpha_1, \alpha_2, \alpha_3)$
$\mathbb{R}_-^m$	$\max_{1 \leq i \leq m} y_i$	$\mu \log \left(\sum_{i=1}^m e^{y_i/\mu} \right)$	$(0, 1, \log(m))$
$\mathcal{S}_-^m$	$\lambda_{\max}(y)$	$\mu \log \left(\sum_{i=1}^m e^{\lambda_i(y)/\mu} \right)$	$(0, 1, \log(m))$

Proposition 1: (Xu, P., Sze '25)

Let $\{\bar{h}_{\mu}\}$ be a *nondecreasing* MSA (i.e., $\bar{h}_{\mu_1} \leq \bar{h}_{\mu_0}$ pointwise whenever $\mu_0 \geq \mu_1 > 0$). Then $h_{\mu} := \bar{h}_{\mu} + \alpha_4 \mu$ is an MSA with parameters $(\alpha_1, \alpha_2, \alpha_3 + \alpha_4)$ and satisfies the **assumption for smoothing**.

Examples:

- $h_{\mu}(y) := \mu \log \left(\sum_{i=1}^m e^{y_i/\mu} \right) + \mu$ for $\mathbb{R}_-^m$.
- $h_{\mu}(y) := \mu \log \left(\sum_{i=1}^m e^{\lambda_i(y)/\mu} \right) + \mu$ for $\mathcal{S}_-^m$.

sMBA

Define $g_\mu := h_\mu \circ G$, where $\{h_\mu\}$ satisfies **assumption for smoothing**.

Algorithm sMBA (Xu, P., Sze '25)

Require. $\tau_1 > 0, \tau_2 > 0, \hat{L} \geq \check{L} > 0, x^0 = \hat{x} \in G^{-1}(\text{int}(\mathcal{K}))$.

Step 1. Find $\mu_0 > 0$ with $g_{\mu_0}(x^0) < 0$. Set $k = 0$.

Step 2. Choose $\xi^k \in \partial P_2(x^k)$ and $L_f^{k,0}, L_g^{k,0} \in [\check{L}, \hat{L}]$. Let $i = j = 0$.

Step 3. Set $L_f^{k,i} = 2^i L_f^{k,0}, L_g^{k,j} = 2^j L_g^{k,0}$. Solve the following for the **unique solution** $x^{k,i,j}$ and a **Lagrange multiplier** $\lambda_{k,i,j} \geq 0$:

$$\begin{aligned} \min_{x \in \mathbb{X}} \quad & P_1(x) + \langle \nabla f(x^k) - \xi^k, x - x^k \rangle + \frac{L_f^{k,i}}{2} \|x - x^k\|^2 \\ \text{s.t.} \quad & g_{\mu_k}(x^k) + \langle \nabla g_{\mu_k}(x^k), x - x^k \rangle + \frac{L_g^{k,j}}{2\mu_k} \|x - x^k\|^2 \leq 0. \end{aligned}$$

Step 4. If $\psi(x^{k,i,j}) \leq \psi(x^k) - \frac{\tau_1 \mu_k + \tau_2 \lambda_{k,i,j}}{2\mu_k} \|x^{k,i,j} - x^k\|^2$ and $g_{\mu_k}(x^{k,i,j}) \leq 0$, go to **Step 5**; else if $g_{\mu_k}(x^{k,i,j}) > 0$, let $j \leftarrow j + 1$, go to **Step 3**; else let $i \leftarrow i + 1$ and $j \leftarrow j + 1$, go to **Step 3**.

Step 5. Let $x^{k+1} = x^{k,i,j}, \lambda_{k+1} = \lambda_{k,i,j}$. Choose $\mu_{k+1} \in (0, \mu_k)$. Update $k \leftarrow k + 1$ and go to **Step 2**.

Well-definedness & choice of $\{\mu_k\}$

Theorem 1. (Xu, P., Sze '25)

The sMBA is well-defined. Specifically, the μ_0 in **Step 1** can be found, the subproblems are well-defined, and there exists ι (independent of k) such that the backtracking steps are performed at most ι times per iteration.

Well-definedness & choice of $\{\mu_k\}$

Theorem 1. (Xu, P., Sze '25)

The sMBA is well-defined. Specifically, the μ_0 in **Step 1** can be found, the subproblems are well-defined, and there exists ι (independent of k) such that the backtracking steps are performed at most ι times per iteration.

Idea: One can show that $x^k \in \Omega_0$ and $g_{\mu_k}(x^k) < 0$ for all k , thanks to RCQ and the fact that $\mu_{k+1} \in (0, \mu_k)$.

Well-definedness & choice of $\{\mu_k\}$

Theorem 1. (Xu, P., Sze '25)

The sMBA is well-defined. Specifically, the μ_0 in **Step 1** can be found, the subproblems are well-defined, and there exists ι (independent of k) such that the backtracking steps are performed at most ι times per iteration.

Idea: One can show that $x^k \in \Omega_0$ and $g_{\mu_k}(x^k) < 0$ for all k , thanks to RCQ and the fact that $\mu_{k+1} \in (0, \mu_k)$.

Assumption for μ_k :

The sequence $\{\mu_k\}$ is positive, decreasing, and satisfies

$$\lim_{k \rightarrow \infty} \mu_k = 0 \text{ and } S_K := \sum_{k=\lceil K/2 \rceil}^K \mu_k \rightarrow \infty.$$

Example: $\mu_k := \mu_0(k+1)^{-r}$ satisfies assumption for μ_k when $r \in (0, 1)$.

Complexity results I

Recall that

$$\begin{array}{ll} \text{Minimize} & f(x) + P_1(x) - P_2(x) \\ & x \in \mathbb{X} \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Theorem 2. (Xu, P., Sze '25)

Assume the **Assumption for smoothing** and **Assumption for μ_k** . Let $\{x^k\}$ and $\{\lambda_{k+1}\}$ be generated by sMBA. Then $G(x^k) \in \mathcal{K}$ for all k and

$$v^{k+1} := \lambda_{k+1} \nabla h_{\mu_k}(G(x^{k+1})) \in \mathcal{K}^\circ \quad \forall k.$$

Moreover, for all $K \geq 0$, there exists $\hat{k} \in [\lceil K/2 \rceil, K]$ such that

$$\begin{aligned} \text{dist} \left(0, \partial P_1(x^{\hat{k}+1}) - \partial P_2(x^{\hat{k}}) + \nabla f(x^{\hat{k}+1}) + DG(x^{\hat{k}+1})^* v^{\hat{k}+1} \right) &= O(S_K^{-1/2}), \\ - \langle G(x^{\hat{k}+1}), v^{\hat{k}+1} \rangle &= O(\mu_{\lceil K/2 \rceil}), \quad \|x^{\hat{k}+1} - x^{\hat{k}}\| = O(\mu_{\lceil K/2 \rceil}^{1/2} S_K^{-1/2}) \end{aligned}$$

Complexity results I

Recall that

$$\begin{array}{ll} \text{Minimize} & f(x) + P_1(x) - P_2(x) \\ & x \in \mathbb{X} \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Theorem 2. (Xu, P., Sze '25)

Assume the **Assumption for smoothing** and **Assumption for μ_k** . Let $\{x^k\}$ and $\{\lambda_{k+1}\}$ be generated by sMBA. Then $G(x^k) \in \mathcal{K}$ for all k and

$$v^{k+1} := \lambda_{k+1} \nabla h_{\mu_k}(G(x^{k+1})) \in \mathcal{K}^\circ \quad \forall k.$$

Moreover, for all $K \geq 0$, there exists $\hat{k} \in [\lceil K/2 \rceil, K]$ such that

$$\begin{aligned} \text{dist} \left(0, \partial P_1(x^{\hat{k}+1}) - \partial P_2(x^{\hat{k}}) + \nabla f(x^{\hat{k}+1}) + DG(x^{\hat{k}+1})^* v^{\hat{k}+1} \right) &= O(S_K^{-1/2}), \\ -\langle G(x^{\hat{k}+1}), v^{\hat{k}+1} \rangle &= O(\mu_{\lceil K/2 \rceil}), \quad \|x^{\hat{k}+1} - x^{\hat{k}}\| = O(\mu_{\lceil K/2 \rceil}^{1/2} S_K^{-1/2}) \end{aligned}$$

Note: When $\mu_k = \mu_0(k+1)^{-r}$ for some $r \in (0, 1)$, then $S_K = \Omega(K^{1-r})$. Hence, $S_K^{-1/2} = O(K^{-\frac{1-r}{2}})$, $\mu_{\lceil K/2 \rceil} = O(K^{-r})$, $\mu_{\lceil K/2 \rceil}^{1/2} S_K^{-1/2} = O(K^{-1/2})$.

Convergence under convexity

Recall that

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Convexity assumption:

The function f is convex, $P_2 = 0$ and G is $(-\mathcal{K})$ -convex.

Convergence under convexity

Recall that

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Convexity assumption:

The function f is convex, $P_2 = 0$ and G is $(-\mathcal{K})$ -convex.

Theorem 3. (Xu, P., Sze '25)

Assume the **Assumption for smoothing**, **Assumption for μ_k** and **Convexity assumption**. Let $\{x^k\}$ be generated by sMBA. Then

$$\psi(x^{K+1}) - \psi^* = O\left(\frac{\max\{1, \sum_{k=\lceil K/2 \rceil}^K \mu_k^2\}}{\sum_{k=\lceil K/2 \rceil}^K \mu_k}\right),$$

where ψ^* is the **optimal value**.

Convergence under convexity

Recall that

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Convexity assumption:

The function f is convex, $P_2 = 0$ and G is $(-\mathcal{K})$ -convex.

Theorem 3. (Xu, P., Sze '25)

Assume the **Assumption for smoothing**, **Assumption for μ_k** and **Convexity assumption**. Let $\{x^k\}$ be generated by sMBA. Then

$$\psi(x^{K+1}) - \psi^* = O\left(\frac{\max\{1, \sum_{k=\lceil K/2 \rceil}^K \mu_k^2\}}{\sum_{k=\lceil K/2 \rceil}^K \mu_k}\right),$$

where ψ^* is the **optimal value**. If we assume further that $\sum_{k=0}^{\infty} \mu_k^2 < \infty$, then $\{x^k\}$ **converges to** some optimal solution.

Convergence under convexity

Recall that

$$\begin{array}{ll} \underset{x \in \mathbb{X}}{\text{Minimize}} & \psi(x) := f(x) + P_1(x) - P_2(x) \\ \text{Subject to} & G(x) \in \mathcal{K}, \end{array}$$

Convexity assumption:

The function f is convex, $P_2 = 0$ and G is $(-\mathcal{K})$ -convex.

Theorem 3. (Xu, P., Sze '25)

Assume the **Assumption for smoothing**, **Assumption for μ_k** and **Convexity assumption**. Let $\{x^k\}$ be generated by sMBA. Then

$$\psi(x^{K+1}) - \psi^* = O\left(\frac{\max\{1, \sum_{k=\lceil K/2 \rceil}^K \mu_k^2\}}{\sum_{k=\lceil K/2 \rceil}^K \mu_k}\right),$$

where ψ^* is the **optimal value**. If we assume further that $\sum_{k=0}^{\infty} \mu_k^2 < \infty$, then $\{x^k\}$ **converges to** some optimal solution.

Note: There are explicit estimates for the constants in the big O .

Local rate via error bounds

Theorem 4. (Xu, P., Sze '25)

Assume the **Assumption for smoothing**, **Assumption for μ_k** and **Convexity assumption**. Assume that $\mu_k = \Theta((k+1)^{-\bar{r}})$ for some $\bar{r} \in (0.5, 1)$, and that $\exists \kappa > 0$, $\theta \in (0, 1)$, $\epsilon_0 > 0$ and $\epsilon_1 \in (0, 1)$ s.t.

$$\text{dist}(x, \Omega^*) \leq \kappa(\psi(x) - \psi^*)^{1-\theta}$$

for all $x \in \Omega_0$ with $\text{dist}(x, \Omega^*) \leq \epsilon_0$ and $\psi(x) \leq \psi^* + \epsilon_1$.¹

Let $\{x^k\}$ be generated by sMBA. Then $x^* := \lim_{k \rightarrow \infty} x^k$ exists, and $\exists k_1 \in \mathbb{N}_0$ and $\kappa_1 > 0$ s.t.

$$\|x^k - x^*\| \leq \kappa_1 \cdot (k+1)^{-s} \quad \forall k \geq k_1,$$

where

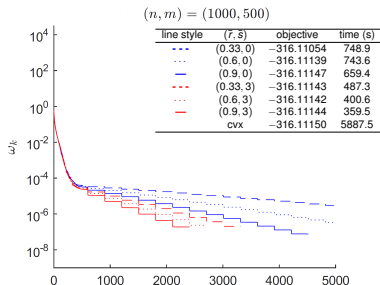
$$s := s(\bar{r}) := \begin{cases} \bar{r} - \frac{1}{2} & \text{if } \theta \in (0, \frac{1}{2}], \\ \min \left\{ \bar{r} - \frac{1}{2}, \frac{(1-\bar{r})(1-\theta)}{2\theta-1} \right\} & \text{if } \theta \in (\frac{1}{2}, 1). \end{cases}$$

¹Here, Ω^* and ψ^* be the **solution set** and the **optimal value**, respectively.

Numerical results

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimize}} && \sum_{i=1}^n \left(\frac{1}{4} d_i x_i^4 + \frac{1}{3} c_i |x_i|^3 \right) + \frac{1}{2} x^\top Q x + b^\top x + \sum_{i=1}^n |x_i| \\ & \text{Subject to} && -A_0 - \sum_{i=1}^n x_i A_i \in \mathcal{S}_-^m, \end{aligned}$$

for randomly generated $c \in \mathbb{R}_+^n$, $d \in \mathbb{R}_+^n$, $Q \in \mathcal{S}_+^n$, $b \in \text{Range}(Q)$, $A_i \in \mathcal{S}_+^m$, $i = 1, \dots, n$ and $A_0 \succ 0$.



Here, $h_\mu(y) := \mu \log \left(\sum_{i=1}^m e^{\lambda_i(y)/\mu} \right) + 10^{-5} \mu$; $\omega_k := (\psi(x^k) - \psi^{\text{cvx}}) / \max\{1, |\psi^{\text{cvx}}|\}$;
 $\mu_k := \mu_0 (\bar{k} + 1)^{-\bar{r}_k} (\log(\bar{k} + 3))^{-s_k}$ with $\bar{k} := 301k_2 + \frac{k_1}{3001}$ where $k = 301k_2 + k_1$,
 $k_1 \leq 300$; $s_j := \min\{1, \frac{j}{5000}\} \bar{s}$; $r_j := 0.01 + \min\{1, \frac{j}{5000}\} (\bar{r} - 0.01)$.

Numerical results cont.

Termination criterion:

$$\max \left\{ \frac{\sqrt{\tau_1 \mu_k + \lambda_{k+1} \tau_2}}{\mu_k} \|x^{k+1} - x^k\|, -\langle G(x^{k+1}), v^{k+1} \rangle \right\} \leq \epsilon \max\{1, \|x^{k+1}\|\},$$

where $v^{k+1} := \lambda_{k+1} \nabla h_{\mu_k}(G(x^{k+1}))$.

Table: Comparisons between sMBA and CVX, averaged over 30 random instances for each (m, n) , with $(\bar{r}, \bar{s}) = (0.9, 3)$. The time for CVX includes the time for problem automatic reformulation.

m	n	sMBA ($\epsilon = 10^{-5}$) iter. / objective / time (s)	sMBA ($\epsilon = 10^{-7}$) iter. / objective / time (s)	CVX objective / time (s)
100	100	667.7 / -25.501867 / 0.6	1675 / -25.506259 / 1.5	-25.506288 / 6.2
	500	841.6 / -134.76308 / 2.4	2573 / -134.76856 / 6.8	-134.76879 / 57.0
	1000	1009 / -272.22067 / 5.7	2577 / -272.22625 / 14.1	-272.22794 / 178.7
500	100	972.5 / -29.639329 / 34.4	2260 / -29.640247 / 82.4	-29.640248 / 301.6
	500	692.8 / -139.35545 / 71.9	2265 / -139.36169 / 200.2	-139.36169 / 2000.2
	1000	798.9 / -289.82629 / 143.9	2196 / -289.83148 / 350.3	-289.83151 / 5943.3

Conclusion

- An MBA variant is developed for a class of conic programs, based on **smoothing** the support function of the base of the dual cone.
- Each subproblem involves only **one** single inequality constraint, and only **one** MBA step is applied before the smoothing evolves.
- **Iteration complexity** and **(local) convergence rate** are derived.

Reference:

- J. Xu, T. K. Pong and N.-s. Sze.
A smoothing moving balls approximation method for a class of conic-constrained difference-of-convex optimization problems.
Available at <https://arxiv.org/abs/2505.12314>.

Thanks for coming! ☺