A conditional-gradient-based single-loop augmented Lagrangian method for inequality constrained problems

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

ICOML 2026
July 2026

(Joint work with Xiaozhou Wang and Zev Woodstock)

Our problem

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{Minimize}} & f(\boldsymbol{x}) + h(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{g}(\boldsymbol{x}) := \begin{bmatrix} g_1(\boldsymbol{x}) & \cdots & g_m(\boldsymbol{x}) \end{bmatrix}^T \leq \mathbf{0}, \end{array}$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L_f-smooth for some $L_f > 0$.
- Each $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex and L_{g_i}-smooth for some $L_{g_i} > 0$.
- $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper closed convex with bounded dom h.

Our problem

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{Minimize}} & f(\boldsymbol{x}) + h(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{g}(\boldsymbol{x}) := \begin{bmatrix} g_1(\boldsymbol{x}) & \cdots & g_m(\boldsymbol{x}) \end{bmatrix}^T \leq \boldsymbol{0}, \end{array}$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L_f-smooth for some $L_f > 0$.
- Each $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex and L_{g_i}-smooth for some $L_{g_i} > 0$.
- $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper closed convex with bounded dom h. Moreover, for any $\boldsymbol{u} \in \mathbb{R}^n$, a minimizer of $\min_{\boldsymbol{v}\in\mathbb{R}^n}\{\langle \boldsymbol{v}, \boldsymbol{u}\rangle + h(\boldsymbol{v})\}$ exists and can be computed efficiently. (Efficient LMO)

Our problem

$$\begin{array}{ll}\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{Minimize}} & f(\boldsymbol{x}) + h(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{g}(\boldsymbol{x}) := \begin{bmatrix} g_1(\boldsymbol{x}) & \cdots & g_m(\boldsymbol{x}) \end{bmatrix}^T \leq \boldsymbol{0},\end{array}$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L_f-smooth for some $L_f > 0$.
- Each $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex and L_{g_i}-smooth for some $L_{g_i} > 0$.
- $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper closed convex with bounded dom h. Moreover, for any $\boldsymbol{u} \in \mathbb{R}^n$, a minimizer of $\min_{\boldsymbol{v}\in\mathbb{R}^n}\{\langle \boldsymbol{v}, \boldsymbol{u}\rangle + h(\boldsymbol{v})\}$ exists and can be computed efficiently. (Efficient LMO)
- A KKT point exists, i.e., there exists $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ such that

$$0 \in \nabla f(\boldsymbol{x}^*) + \partial h(\boldsymbol{x}^*) + \sum_{i=1}^{m} z_i^* \nabla g_i(\boldsymbol{x}^*),$$

$$\boldsymbol{g}(\boldsymbol{x}^*) \leq 0, \quad \boldsymbol{z}^* \geq 0, \quad \langle \boldsymbol{z}^*, \boldsymbol{g}(\boldsymbol{x}^*)\rangle = 0.$$

Motivation

Natural idea:

- Apply the Augmented Lagrangian (AL) framework, with subproblems approx. solved by Frank-Wolfe method.

Algorithm A conceptual AL method based on LMO

Step 1. Choose $(\boldsymbol{x}^0, \boldsymbol{z}^0) \in \operatorname{dom} h \times \mathbb{R}^m_+$, a positive increasing sequence $\{\lambda_k\}$ with $\lim_{k\to\infty} \lambda_k = \infty$.

Step 2. For $k = 0, \ldots,$

Apply FW method to approx. minimize $F_{\lambda_k}(\cdot, \boldsymbol{z}^k) + h(\cdot)$, where

$$F_{\lambda_k}(\boldsymbol{x}, \boldsymbol{z}^k) := f(\boldsymbol{x}) + \frac{\lambda_k}{2} \sum_{i=1}^{m} \left(\left[g_i(\boldsymbol{x}) + \frac{z_i^k}{\lambda_k} \right]_+^2 - \left(\frac{z_i^k}{\lambda_k} \right)^2 \right)$$

Set $\boldsymbol{z}^{k+1} = \max\{\boldsymbol{0}, \boldsymbol{z}^k + \lambda_k \boldsymbol{g}(\boldsymbol{x}^k)\}$.

Motivation

Natural idea:

- Apply the Augmented Lagrangian (AL) framework, with subproblems approx. solved by Frank-Wolfe method.

Algorithm A conceptual AL method based on LMO

Step 1. Choose $(\boldsymbol{x}^0, \boldsymbol{z}^0) \in \operatorname{dom} h \times \mathbb{R}^m_+$, a positive increasing sequence $\{\lambda_k\}$ with $\lim_{k\to\infty} \lambda_k = \infty$.

Step 2. For $k = 0, \ldots,$

Apply FW method to approx. minimize $F_{\lambda_k}(\cdot, \boldsymbol{z}^k) + h(\cdot)$, where

$$F_{\lambda_k}(\boldsymbol{x}, \boldsymbol{z}^k) := f(\boldsymbol{x}) + \frac{\lambda_k}{2} \sum_{i=1}^{m} \left(\left[g_i(\boldsymbol{x}) + \frac{z_i^k}{\lambda_k} \right]_+^2 - \left(\frac{z_i^k}{\lambda_k} \right)^2 \right)$$

Set $\boldsymbol{z}^{k+1} = \max\{\mathbf{0}, \boldsymbol{z}^k + \lambda_k \boldsymbol{g}(\boldsymbol{x}^k)\}$.

Complicated "Double loop" structure. ☹

Existing work

Existing Single-loop (i.e., 1 LMO per iteration) approaches:

- Most existing works focus on equality constraints, e.g., (Gidel, Pedregosa, Lacoste-Julien '18, Yurtsever, Fercoq, Cevher '19).

Existing work

Existing Single-loop (i.e., 1 LMO per iteration) approaches:

- Most existing works focus on equality constraints, e.g., (Gidel, Pedregosa, Lacoste-Julien '18, Yurtsever, Fercoq, Cevher '19).
- Two specialized algorithms that can achieve primal suboptimality and feasibility violation bounds of $O(1/\sqrt{k})$, which did not adopt the AL framework (Lan, Romeijn, Zhou '21, Asgari, Neely '24).

Existing work

Existing Single-loop (i.e., 1 LMO per iteration) approaches:

- Most existing works focus on equality constraints, e.g., (Gidel, Pedregosa, Lacoste-Julien '18, Yurtsever, Fercoq, Cevher '19).
- Two specialized algorithms that can achieve primal suboptimality and feasibility violation bounds of $O(1/\sqrt{k})$, which did not adopt the AL framework (Lan, Romeijn, Zhou '21, Asgari, Neely '24).

> Questions:
>
> - Develop an AL-based single-loop algorithm?

Existing work

Existing Single-loop (i.e., 1 LMO per iteration) approaches:

- Most existing works focus on equality constraints, e.g., (Gidel, Pedregosa, Lacoste-Julien '18, Yurtsever, Fercoq, Cevher '19).
- Two specialized algorithms that can achieve primal suboptimality and feasibility violation bounds of $O(1/\sqrt{k})$, which did not adopt the AL framework (Lan, Romeijn, Zhou '21, Asgari, Neely '24).

Questions:

- Develop an AL-based single-loop algorithm?
- Derive improved rates under uniform convexity?

Algorithmic framework

Algorithm A CG-based single-loop AL method

Step 1. Choose $(\boldsymbol{x}^0, \boldsymbol{z}^0) \in \operatorname{dom} h \times \mathbb{R}^m_+$, $\alpha_0 \in [0,1]$, a positive increasing sequence $\{\lambda_k\}$ with $\lim_{k\to\infty} \lambda_k = \infty$, and a positive summable sequence $\{\sigma_k\}$ with $\sigma_k \le \lambda_k$. Set $k = 0$.

Step 2. Compute

$$\boldsymbol{v}^k \in \operatorname*{Arg\,min}_{\boldsymbol{v}\in\mathbb{R}^n} \langle \nabla_{\boldsymbol{x}} F_{\lambda_k}(\boldsymbol{x}^k, \boldsymbol{z}^k), \boldsymbol{v}\rangle + h(\boldsymbol{v}),$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k(\boldsymbol{v}^k - \boldsymbol{x}^k),$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \sigma_k \max\left\{-\frac{\boldsymbol{z}^k}{\lambda_k}, \boldsymbol{g}(\boldsymbol{x}^{k+1})\right\}.$$

Step 3. Choose $\alpha_{k+1} \in [0,1]$ and update $k \leftarrow k+1$. Go to **Step 2**.

Here, $F_\lambda(\boldsymbol{x}, \boldsymbol{z}) := f(\boldsymbol{x}) + \frac{\lambda}{2}\sum_{i=1}^m \left(\left[g_i(\boldsymbol{x}) + \frac{z_i}{\lambda}\right]_+^2 - \left(\frac{z_i}{\lambda}\right)^2\right)$.

General convergence result

Define $\mathcal{L}_\lambda(\boldsymbol{x},\boldsymbol{z}) := f(\boldsymbol{x}) + \frac{\lambda}{2}\sum_{i=1}^m \left(\left[g_i(\boldsymbol{x}) + \frac{z_i}{\lambda}\right]_+^2 - \left(\frac{z_i}{\lambda}\right)^2 \right) + h(\boldsymbol{x})$.

Theorem 1. (Wang, P., Woodstock '26)
Let $\{(\boldsymbol{x}^k,\boldsymbol{z}^k)\}$ be generated by Algorithm 1 and define

$$T_k := [\mathcal{L}_{\lambda_k}(\boldsymbol{x}^k,\boldsymbol{z}^k) - L^*]_+ \quad \forall k \geq 0.$$

Then it holds that

$$\max\{\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|, |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|\} = \mathcal{O}(\max\{T_k, 1/\lambda_k\}),$$

where L^* is the optimal value. Consequently, if

$$\lim_{k\to\infty} T_k = 0,$$

then

$$\lim_{k\to\infty} \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\| = \lim_{k\to\infty} |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*| = 0.$$

General convergence result

Define $\mathcal{L}_\lambda(\boldsymbol{x},\boldsymbol{z}) := f(\boldsymbol{x}) + \frac{\lambda}{2}\sum_{i=1}^m \left(\left[g_i(\boldsymbol{x}) + \frac{z_i}{\lambda}\right]_+^2 - \left(\frac{z_i}{\lambda}\right)^2 \right) + h(\boldsymbol{x})$.

Theorem 1. (Wang, P., Woodstock '26)
Let $\{(\boldsymbol{x}^k,\boldsymbol{z}^k)\}$ be generated by Algorithm 1 and define

$$T_k := [\mathcal{L}_{\lambda_k}(\boldsymbol{x}^k,\boldsymbol{z}^k) - L^*]_+ \quad \forall k \geq 0.$$

Then it holds that

$$\max\{\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|, |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|\} = \mathcal{O}(\max\{T_k, 1/\lambda_k\}),$$

where L^* is the optimal value. Consequently, if

$$\lim_{k\to\infty} T_k = 0,$$

then

$$\lim_{k\to\infty} \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\| = \lim_{k\to\infty} |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*| = 0.$$

Question: How can we guarantee $\lim_{k\to\infty} T_k = 0$?

General convergence result

Define $\mathcal{L}_\lambda(\boldsymbol{x}, \boldsymbol{z}) := f(\boldsymbol{x}) + \frac{\lambda}{2} \sum_{i=1}^m \left(\left[g_i(\boldsymbol{x}) + \frac{z_i}{\lambda} \right]_+^2 - \left(\frac{z_i}{\lambda} \right)^2 \right) + h(\boldsymbol{x})$.

Theorem 1. (Wang, P., Woodstock '26)
Let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1 and define

$$T_k := [\mathcal{L}_{\lambda_k}(\boldsymbol{x}^k, \boldsymbol{z}^k) - L^*]_+ \quad \forall k \geq 0.$$

Then it holds that

$$\max\{\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|, |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|\} = \mathcal{O}(\max\{T_k, 1/\lambda_k\}),$$

where L^* is the optimal value. Consequently, if

$$\lim_{k \to \infty} T_k = 0,$$

then

$$\lim_{k \to \infty} \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\| = \lim_{k \to \infty} |f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*| = 0.$$

Question: How can we guarantee $\lim_{k \to \infty} T_k = 0$? Suitably choose α_k, λ_k and σ_k.

Open-loop stepsize

Open-loop stepsize assumptions:

(i) $\{\lambda_k\}$ is increasing, $\lambda_k = \Theta(k^\tau)$, $\lambda_{k+1} - \lambda_k = \mathcal{O}(k^{-(1-\tau)})$ for some $\tau \in (0, 1)$.

(ii) $\{\alpha_k\} \subset (0, 1]$ is nonincreasing, $\alpha_k = \Theta(k^{-p})$ for some $p \in (\tau, 1)$.

(iii) $\{\sigma_k\}$ satisfies $\sigma_k \leq \lambda_k$ and $\sigma_k = \Theta(k^{-(1+\gamma)})$ for some $\gamma > 0$.

Open-loop stepsize

Open-loop stepsize assumptions:

(i) $\{\lambda_k\}$ is increasing, $\lambda_k = \Theta(k^\tau)$, $\lambda_{k+1} - \lambda_k = \mathcal{O}(k^{-(1-\tau)})$ for some $\tau \in (0,1)$.

(ii) $\{\alpha_k\} \subset (0,1]$ is nonincreasing, $\alpha_k = \Theta(k^{-p})$ for some $p \in (\tau, 1)$.

(iii) $\{\sigma_k\}$ satisfies $\sigma_k \le \lambda_k$ and $\sigma_k = \Theta(k^{-(1+\gamma)})$ for some $\gamma > 0$.

Theorem 2. (Wang, P., Woodstock '26)
Under the Open-loop stepsize assumptions, let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1 and L^* be the optimal value. Then

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(\max\{k^{-(p-\tau)}, k^{-\tau}\}).$$

If, in addition, our problem has a Slater point, i.e., $\exists \hat{\boldsymbol{x}} \in \operatorname{dom} h$ with $\boldsymbol{g}(\boldsymbol{x}) < \boldsymbol{0}$, then there exists a subsequence of $\{(\boldsymbol{x}^k, [\lambda_k \boldsymbol{g}(\boldsymbol{x}^k) + \boldsymbol{z}^k]_+)\}$ such that all of its cluster points are KKT points.

Short stepsize

Short stepsize assumptions:

(i) $\{\lambda_k\}$ is increasing, $\lambda_k = \Theta(k^\tau)$ and $\lambda_{k+1} - \lambda_k = \mathcal{O}(k^{-(1-\tau)})$ for some $\tau \in (0,1)$.

(ii) The α_k is the short stepsize, i.e.,

$$\alpha_k = \begin{cases} 0 & \text{if } G_{\boldsymbol{z}^k,\lambda_k}(\boldsymbol{x}^k) = \|\boldsymbol{v}^k - \boldsymbol{x}^k\| = 0, \\ \min\left\{1, \frac{G_{\boldsymbol{z}^k,\lambda_k}(\boldsymbol{x}^k)}{(L_\Psi(\boldsymbol{x}^k,\boldsymbol{z}^k)+L_f)\|\boldsymbol{v}^k-\boldsymbol{x}^k\|^2}\right\} & \text{else,} \end{cases}$$

where

$$G_{\boldsymbol{z},\lambda}(\boldsymbol{x}) := \langle \nabla_{\boldsymbol{x}} F_\lambda(\boldsymbol{x},\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{v}^+\rangle + h(\boldsymbol{x}) - h(\boldsymbol{v}^+),$$

where $\boldsymbol{v}^+ \in \operatorname{Arg\,min}_{\boldsymbol{v}} \langle \nabla_{\boldsymbol{x}} F_\lambda(\boldsymbol{x},\boldsymbol{z}), \boldsymbol{v}\rangle + h(\boldsymbol{v})$, and $L_\Psi(\boldsymbol{x},\boldsymbol{z}) + L_f$ is a Lipschitz constant for $\nabla_{\boldsymbol{x}} F_\lambda(\boldsymbol{x},\boldsymbol{z})$ on $\operatorname{dom} h$.

(iii) $\{\sigma_k\}$ satisfies $\sigma_k \le \lambda_k$ and $\sigma_k = \Theta(k^{-(1+\gamma)})$ for some $\gamma > 0$.

Recall that $F_\lambda(\boldsymbol{x},\boldsymbol{z}) := f(\boldsymbol{x}) + \frac{\lambda}{2}\sum_{i=1}^m \left([g_i(\boldsymbol{x}) + \frac{z_i}{\lambda}]_+^2 - (\frac{z_i}{\lambda})^2 \right)$.

Short stepsize cont.

Theorem 3. (Wang, P., Woodstock '26)
Under the Short stepsize assumptions, let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1 and L^* be the optimal value. Then

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(\max\{k^{-1/2}, k^{-(1-\tau)}, k^{-\tau}\}).$$

If, in addition, our problem has a Slater point, i.e., $\exists \hat{\boldsymbol{x}} \in \operatorname{dom} h$ with $\boldsymbol{g}(\boldsymbol{x}) < \boldsymbol{0}$, then there exists a subsequence of $\{(\boldsymbol{x}^k, [\lambda_k \boldsymbol{g}(\boldsymbol{x}^k) + \boldsymbol{z}^k]_+)\}$ such that all of its cluster points are KKT points.

Short stepsize cont.

Theorem 3. (Wang, P., Woodstock '26)
Under the Short stepsize assumptions, let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1 and L^* be the optimal value. Then

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(\max\{k^{-1/2}, k^{-(1-\tau)}, k^{-\tau}\}).$$

If, in addition, our problem has a Slater point, i.e., $\exists \hat{\boldsymbol{x}} \in \operatorname{dom} h$ with $\boldsymbol{g}(\boldsymbol{x}) < \boldsymbol{0}$, then there exists a subsequence of $\{(\boldsymbol{x}^k, [\lambda_k \boldsymbol{g}(\boldsymbol{x}^k) + \boldsymbol{z}^k]_+)\}$ such that all of its cluster points are KKT points.

Remark: Taking $\tau = 0.5$, we see that Algorithm 1 achieves the rate

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(1/\sqrt{k}).$$

Uniform convexity

Uniform convexity assumptions:

(i) Function h is the indicator function of a set $\mathcal{C}$ that is compact and (ν, q)-uniformly convex for some $q \geq 2$ and $\nu > 0$.

(ν, q)-uniformly convexity means $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$, $t \in [0, 1]$ and $\|\boldsymbol{z}\| \leq 1$, it holds that $t\boldsymbol{x} + (1 - t)\boldsymbol{y} + t(1 - t)\nu\|\boldsymbol{x} - \boldsymbol{y}\|^q \boldsymbol{z} \in \mathcal{C}$.

(ii) For some $\epsilon > 0$, it holds that

$$\zeta := \inf\left\{\left\|\nabla f(\boldsymbol{x}) + \textstyle\sum_{i=1}^m z_i \nabla g_i(\boldsymbol{x})\right\| : (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{A}_\epsilon, \boldsymbol{x} \in \mathcal{C}\right\} > 0,$$

where

$$\mathcal{A}_\epsilon := \{(\boldsymbol{x}, \boldsymbol{z}) \in \boldsymbol{g}^{-1}((-\infty, \epsilon]^m) \times \mathbb{R}_+^m : \forall i, \text{ if } g_i(\boldsymbol{x}) < -\epsilon, \text{ then } z_i = 0\}.$$

Uniform convexity

Uniform convexity assumptions:

(i) Function h is the indicator function of a set $\mathcal{C}$ that is compact and (ν, q)-uniformly convex for some $q \geq 2$ and $\nu > 0$.

(ν, q)-uniformly convexity means $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$, $t \in [0, 1]$ and $\|\boldsymbol{z}\| \leq 1$, it holds that $t\boldsymbol{x} + (1-t)\boldsymbol{y} + t(1-t)\nu\|\boldsymbol{x} - \boldsymbol{y}\|^q \boldsymbol{z} \in \mathcal{C}$.

(ii) For some $\epsilon > 0$, it holds that

$$\zeta := \inf\left\{\left\|\nabla f(\boldsymbol{x}) + \textstyle\sum_{i=1}^m z_i \nabla g_i(\boldsymbol{x})\right\| : (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{A}_\epsilon, \boldsymbol{x} \in \mathcal{C}\right\} > 0,$$

where

$$\mathcal{A}_\epsilon := \{(\boldsymbol{x}, \boldsymbol{z}) \in \boldsymbol{g}^{-1}((-\infty, \epsilon]^m) \times \mathbb{R}_+^m : \forall i, \text{ if } g_i(\boldsymbol{x}) < -\epsilon, \text{ then } z_i = 0\}.$$

Proposition 1. (Wang, P., Woodstock '26)
Suppose Uniform convexity assumption (i) holds. Suppose that $\exists \epsilon > 0$ and $\tilde{\boldsymbol{x}}$ such that $g_i(\tilde{\boldsymbol{x}}) < -\epsilon$ for all i, and it holds that

$$\operatorname*{Arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n}\{f(\boldsymbol{x}) : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{\eta}\} \cap \mathcal{C} = \emptyset \quad \forall \boldsymbol{\eta} \in [-\epsilon, \epsilon]^m.$$

Then Uniform convexity assumption (ii) holds with the above ϵ.

Uniform convexity: Open-loop stepsize

Theorem 4. (Wang, P., Woodstock '26)
Suppose that the Open-loop stepsize assumptions and Uniform convexity assumptions hold. Let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1. Then it holds that

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\}$$
$$= \begin{cases} \mathcal{O}(k^{-\tau}) & \text{if } \mu = 0, \\ \mathcal{O}(\max\{\frac{1}{k^{(1-\mu)\omega+(p-\tau)}}, \frac{1}{k^\tau}\}) & \text{if } \mu \in (0,1), \end{cases}$$

where $\omega := \min\{p-\tau, 1-p+\tau\}$ with $0 < \tau < p < 1$ and $\mu := 1 - 2/q$.

Uniform convexity: Open-loop stepsize

Theorem 4. (Wang, P., Woodstock '26)
Suppose that the Open-loop stepsize assumptions and Uniform convexity assumptions hold. Let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1. Then it holds that

$$\max\{|f(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) - L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\}$$
$$= \begin{cases} \mathcal{O}(k^{-\tau}) & \text{if } \mu = 0, \\ \mathcal{O}(\max\{\frac{1}{k^{(1-\mu)\omega+(p-\tau)}}, \frac{1}{k^\tau}\}) & \text{if } \mu \in (0,1), \end{cases}$$

where $\omega := \min\{p-\tau, 1-p+\tau\}$ with $0 < \tau < p < 1$ and $\mu := 1 - 2/q$.

Remark:

- The above rates are no worse than the previously proved rate of $\mathcal{O}(\max\{k^{-(p-\tau)}, k^{-\tau}\})$.
- When $\mathcal{C}$ is strongly convex, the above rate can be made arbitrarily close to $\mathcal{O}(1/k)$ by choosing $p > \tau$ with $\tau \in (0,1)$ being arbitrarily close to 1.

Uniform convexity: Short stepsize

Theorem 5. (Wang, P., Woodstock '26)
Suppose that the Short stepsize assumptions and Uniform convexity assumptions hold. Let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1. Then it holds that

$$\max\{|f(\boldsymbol{x}^k)+h(\boldsymbol{x}^k)-L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(\max\{1/k^{\frac{1}{1+\mu}}, 1/k^{\frac{2-2\tau}{1+\mu}}, 1/k^{\tau}\}),$$

where $\mu := 1 - 2/q \in [0, 1)$.

Uniform convexity: Short stepsize

Theorem 5. (Wang, P., Woodstock '26)
Suppose that the Short stepsize assumptions and Uniform convexity assumptions hold. Let $\{(\boldsymbol{x}^k, \boldsymbol{z}^k)\}$ be generated by Algorithm 1. Then it holds that

$$\max\{|f(\boldsymbol{x}^k)+h(\boldsymbol{x}^k)-L^*|, \|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|\} = \mathcal{O}(\max\{1/k^{\frac{1}{1+\mu}}, 1/k^{\frac{2-2\tau}{1+\mu}}, 1/k^{\tau}\}),$$

where $\mu := 1 - 2/q \in [0, 1)$.

Remark:

- The above rates are no worse than the previously proved rate of $\mathcal{O}(\max\{k^{-1/2}, k^{-(1-\tau)}, k^{-\tau}\})$.
- When $\mathcal{C}$ is strongly convex, the best rate is $\mathcal{O}(k^{-2/3})$ by choosing $\tau = 2/3$.

Preliminary numerics

Consider random instances of

$$\begin{array}{rl} \underset{\boldsymbol{x}\in\mathcal{C}}{\text{Minimize}} & f(\boldsymbol{x}) := \langle \boldsymbol{x}-\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})\rangle \\ \text{Subject to} & g_i(\boldsymbol{x}) := \langle \boldsymbol{x}, \boldsymbol{Q}_i\boldsymbol{x}\rangle + \langle \boldsymbol{x}, \boldsymbol{r}_i\rangle + d_i \le 0, \quad i = 1, 2, \end{array}$$

where

- $\mathcal{C} := \{\boldsymbol{x} \in \mathbb{R}^{500\times 500} : \boldsymbol{x}\mathbf{1} = \mathbf{1},\ \boldsymbol{x}^T\mathbf{1} = \mathbf{1},\ \boldsymbol{x} \ge \mathbf{0}\}$ is the Birkhoff polytope.
- $\boldsymbol{A} \in \mathcal{S}^{500}$ and $\boldsymbol{Q}_i \in \mathcal{S}^{500}$ are positive definite.
- d_i is chosen to contain the barycenter $\boldsymbol{c}$ of $\mathcal{C}$ (s.t. $g_i(\boldsymbol{c}) < 0$) while ruling out one of its corners.
- $\boldsymbol{b} \notin \mathcal{C}$ is obtained by extrapolating (out of $\mathcal{C}$) from $\boldsymbol{c}$ towards a ruled-out corner.

Preliminary numerics

Consider random instances of

$$\begin{array}{rl} \underset{\boldsymbol{x}\in\mathcal{C}}{\text{Minimize}} & f(\boldsymbol{x}) := \langle \boldsymbol{x}-\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})\rangle \\ \text{Subject to} & g_i(\boldsymbol{x}) := \langle \boldsymbol{x}, \boldsymbol{Q}_i\boldsymbol{x}\rangle + \langle \boldsymbol{x}, \boldsymbol{r}_i\rangle + d_i \leq 0, \quad i = 1, 2, \end{array}$$

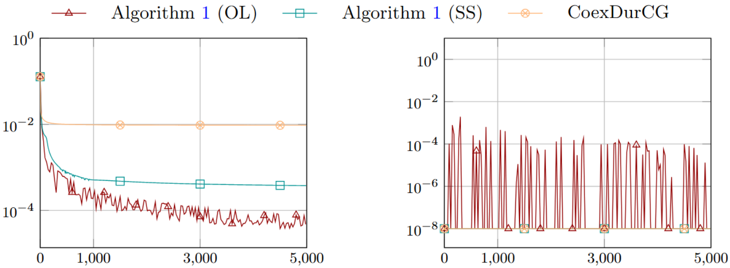
where

- $\mathcal{C} := \{\boldsymbol{x} \in \mathbb{R}^{500\times 500} : \boldsymbol{x}\mathbf{1} = \mathbf{1},\ \boldsymbol{x}^T\mathbf{1} = \mathbf{1},\ \boldsymbol{x} \geq \mathbf{0}\}$ is the Birkhoff polytope.
- $\boldsymbol{A} \in \mathcal{S}^{500}$ and $\boldsymbol{Q}_i \in \mathcal{S}^{500}$ are positive definite.
- d_i is chosen to contain the barycenter $\boldsymbol{c}$ of $\mathcal{C}$ (s.t. $g_i(\boldsymbol{c}) < 0$) while ruling out one of its corners.
- $\boldsymbol{b} \notin \mathcal{C}$ is obtained by extrapolating (out of $\mathcal{C}$) from $\boldsymbol{c}$ towards a ruled-out corner.

We compare our algorithm with open-loop stepsize (OL), short stepsize (SS) (suitably scaled for the first 1000 iterations), and the CoexDurCG in Lan, Romeijn, Zhou '21.
We use Gurobi as the benchmark.

Preliminary numerics cont.

$$\mathrm{val}_k := \frac{|f(\boldsymbol{x}^k) - f^*_{\mathrm{Gurobi}}|}{\max\{f^*_{\mathrm{Gurobi}}, 1\}}, \quad \mathrm{feas}_k := \max\left\{ \frac{\|[\boldsymbol{g}(\boldsymbol{x}^k)]_+\|_\infty}{\max\{\|\boldsymbol{g}(\boldsymbol{0})\|_\infty, 1\}}, 10^{-8} \right\}.$$



(a) val_k over iterations

(b) feas_k over iterations

Algorithms	iterations	objective	feasibility	val_{5000}	time (sec)
Algorithm 1 (OL)	5000	123414.30	0	4.75×10^{-5}	1260.93
Algorithm 1 (SS)	5000	123454.93	0	3.77×10^{-4}	852.82
CoexDurCG	5000	124580.37	0	9.45×10^{-3}	2835.03
Gurobi	18	123408.44	6.87×10^{-13}	0	5108.93

Conclusion

- An AL-based single-loop algorithm was developed for minimizing the sum of a smooth convex function and an LMO-friendly convex function over smooth convex inequality constraints: each iteration requires one call to the LMO.
- Iteration complexity results were obtained for both open-loop stepsize and short stepsize schedules: the latter matches the state-of-the-art complexity result.
- Improved complexity results were established under uniform convexity assumption.

Reference:

- X. Wang, T. K. Pong and Z. Woodstock.
  A conditional-gradient-based single-loop augmented Lagrangian method for inequality constrained optimization.
  Available at https://arxiv.org/abs/2605.22539.

Thanks for coming! ☺