Linear Minimization versus Projections: Which is faster?

ICCOPT 2025, Los Angeles, CA, USA

Zev Woodstock
woodstzc[at]jmu.edu
James Madison University (JMU)

July 2025

Linear Minimization v. Projections: Which is faster?

1. Motivation
2. Results
3. Conclusion & more questions

Setting

Notation: $\mathcal{H}$ is a real Hilbert space with inner product, $\langle \cdot \mid \cdot \rangle$ and induced norm $\|\cdot\|$.

C is a nonempty compact convex subset of $\mathcal{H}$.

Consider two operations w.r.t. C: projection and linear minimization oracle

$$\mathrm{proj}_C(\boldsymbol{x}) = \operatorname*{Argmin}_{\boldsymbol{v}\in C} \|\boldsymbol{x} - \boldsymbol{v}\|^2 \qquad \mathrm{LMO}_C(\boldsymbol{x}) \in \operatorname*{Argmin}_{\boldsymbol{v}\in C} \langle \boldsymbol{x} \mid \boldsymbol{v}\rangle. \tag{1}$$

Let's race them.

image: Meta AI

. . . But why?

This will help us perform per-iteration complexity comparisons between two very large families of first-order algorithms: Projection methods and Frank-Wolfe, (AKA Conditional Gradient) methods.

... But why?

This will help us perform per-iteration complexity comparisons between two very large families of first-order algorithms: Projection methods and Frank-Wolfe, (AKA Conditional Gradient) methods.

Many works (e.g., [C. Combettes & Pokutta, 2021], [Dunn & Harshbarger, 1978], [Garber, Kaplan, & Sabach, 2021], ...) have established that (especially when $\dim \mathcal{H}$ is high), LMO is currently faster than proj on a variety of set classes C:

. . . But why?

This will help us perform per-iteration complexity comparisons between two very large families of first-order algorithms: Projection methods and Frank-Wolfe, (AKA Conditional Gradient) methods.

Many works (e.g., [C. Combettes & Pokutta, 2021], [Dunn & Harshbarger, 1978], [Garber, Kaplan, & Sabach, 2021], . . .) have established that (especially when $\dim \mathcal{H}$ is high), LMO is currently faster than proj on a variety of set classes C:

LMO-advantaged sets: nuclear norm ball, ℓ_1 ball, probability simplex, Birkhoff polytope, general LP, ...

... But why?

This will help us perform per-iteration complexity comparisons between two very large families of first-order algorithms: Projection methods and Frank-Wolfe, (AKA Conditional Gradient) methods.

Many works (e.g., [C. Combettes & Pokutta, 2021], [Dunn & Harshbarger, 1978], [Garber, Kaplan, & Sabach, 2021], ...) have established that (especially when $\dim \mathcal{H}$ is high), LMO is currently faster than proj on a variety of set classes C:

LMO-advantaged sets: nuclear norm ball, ℓ_1 ball, probability simplex, Birkhoff polytope, general LP, ...

Open question: Is there a compact convex C that is not "LMO-advantaged"?

Complexity / Definitions

For $\varepsilon \geqslant 0$, an ε-**approximate LMO of** x is a point $v \in C$ such that

$$0 \leqslant \langle v \mid x \rangle - \min_{c \in C} \langle c \mid x \rangle \leqslant \varepsilon.$$

At times, it will be convenient to use the set-valued notation

$$\mathrm{LMO}_C(x) = \operatorname*{Argmin}_{v \in C} \langle x \mid v \rangle \subset \mathcal{H}$$

Complexity / Definitions

For $\varepsilon \geqslant 0$, an **ε-approximate LMO of** x is a point $v \in C$ such that

$$0 \leqslant \langle v \mid x \rangle - \min_{c \in C} \langle c \mid x \rangle \leqslant \varepsilon.$$

At times, it will be convenient to use the set-valued notation

$$\mathrm{LMO}_C(x) = \operatorname*{Argmin}_{v \in C} \langle x \mid v \rangle \subset \mathcal{H}$$

Assumption 1: Suppose that projection and ε-approximate linear minimization can be performed over C using finitely many vector-arithmetic operations. Let P and $L(\varepsilon)$ respectively denote the smallest amount of operations required.*

[note] For most sets C, **we do not know** P and $L(\varepsilon)$

*: Black-box complexity model may be easier; article under revision.

Linear Minimization v. Projections: Which is faster?

Gameplan

Approximate $\mathsf{LMO}_C(x)$ using one evaluation of proj_C;

Gameplan

Approximate $\mathrm{LMO}_C(x)$ using one evaluation of proj_C; carefully manage the error.

Gameplan

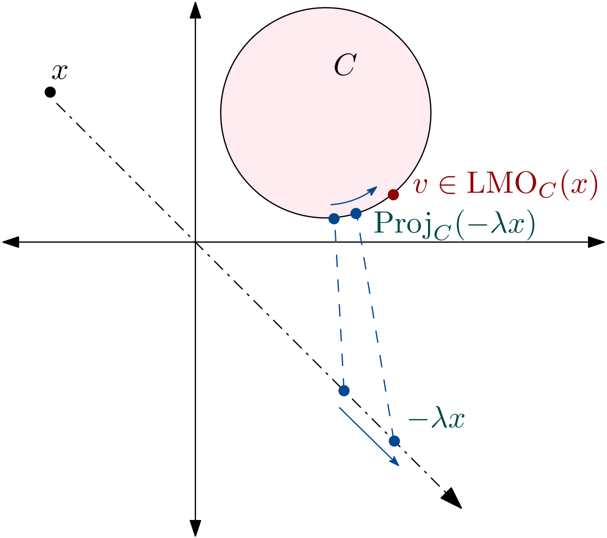
Approximate $\text{LMO}_C(x)$ using one evaluation of proj_C; carefully manage the error.

Two results:

1. For $\varepsilon > 0$
   "Optimal cost of ε-LMO" $\leqslant$ "Optimal cost of projection"

2. If C is polyhedral:
   "Optimal cost of exact LMO" $\leqslant$ "Optimal cost of projection"

Geometry



Geometric concept (similar to [Mortagy, Gupta, & Pokutta, 2023])

$$\mathrm{proj}_C(-\lambda x) \approx \mathrm{LMO}_C(x).$$

Q: What explicit λ is needed to guarantee $\mathrm{proj}_C(-\lambda x)$ is an ε-approximate LMO?

Proposition

Let $C \subset \mathcal{H}$ be a nonempty compact convex set. Then, for every $x \in \mathcal{H}$,

$$\operatorname{proj}_C x \in \mathrm{LMO}_C(\operatorname{proj}_C x - x). \tag{2}$$

[note]: Depending on your selection (single-valued implementation) of LMO_C, (2) might not hold with equality!

Proposition

Let $C \subset \mathcal{H}$ be a nonempty compact convex set. Then, for every $x \in \mathcal{H}$,

$$\operatorname{proj}_C x \in \operatorname{LMO}_C(\operatorname{proj}_C x - x). \tag{2}$$

Proof.

$$(\forall z \in \mathcal{H}) \quad v \in \operatorname{LMO}_C(z) = \underset{c \in C}{\operatorname{Argmin}} \langle z \mid c \rangle \Leftrightarrow \begin{cases} v \in C \\ \sup_{c \in C} \langle -z \mid c - v \rangle \leqslant 0. \end{cases} \tag{3}$$

$$p = \operatorname{proj}_C x \Leftrightarrow x - p \in N_C p \Leftrightarrow \begin{cases} p \in C \\ \sup_{c \in C} \langle x - p \mid c - p \rangle \leqslant 0. \end{cases} \tag{4}$$

Setting $z = \operatorname{proj}_C x - x$ in (3), we see from (4) that $\operatorname{proj}_C(x)$ solves (3). □

From the proposition,

$$\mathrm{proj}_C(-\lambda x) \in \underset{c \in C}{\mathrm{Argmin}} \ \langle c \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \mathrm{proj}_C(-\lambda x) \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \underset{c \in C}{\operatorname{Argmin}} \ \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \operatorname{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Dividing by λ, rearranging, and using norm inequalities

From the proposition,

$$\mathrm{proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c \in C} \langle c \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \mathrm{proj}_C(-\lambda x) \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Dividing by λ, rearranging, and using norm inequalities

$$\langle \mathrm{proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle \leqslant \lambda^{-1}\left(\langle v \mid \mathrm{proj}_C(-\lambda x) \rangle - \| \mathrm{proj}_C(-\lambda x) \|^2\right) \tag{7}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \underset{c \in C}{\operatorname{Argmin}} \ \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Dividing by λ, rearranging, and using norm inequalities

$$0 \leqslant \langle \operatorname{proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle \leqslant \lambda^{-1} \left(\langle v \mid \operatorname{proj}_C(-\lambda x) \rangle - \| \operatorname{proj}_C(-\lambda x) \|^2 \right) \tag{7}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c \in C} \ \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Dividing by λ, rearranging, and using norm inequalities

$$\begin{aligned} 0 \leqslant \langle \operatorname{proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle &\leqslant \lambda^{-1}(\langle v \mid \operatorname{proj}_C(-\lambda x) \rangle - \| \operatorname{proj}_C(-\lambda x)\|^2) && (7) \\ &\leqslant \lambda^{-1}(\|v\| \| \operatorname{proj}_C(-\lambda x)\| - \| \operatorname{proj}_C(-\lambda x)\|^2) && (8) \\ &= \lambda^{-1}\| \operatorname{proj}_C(-\lambda x)\|(\|v\| - \| \operatorname{proj}_C(-\lambda x)\|) && (*) \end{aligned}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c \in C} \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Dividing by λ, rearranging, and using norm inequalities

$$\begin{aligned} 0 \leqslant \langle \operatorname{proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle &\leqslant \lambda^{-1}(\langle v \mid \operatorname{proj}_C(-\lambda x) \rangle - \| \operatorname{proj}_C(-\lambda x)\|^2) && (7) \\ &\leqslant \lambda^{-1}(\|v\| \| \operatorname{proj}_C(-\lambda x)\| - \| \operatorname{proj}_C(-\lambda x)\|^2) && (8) \\ &= \lambda^{-1}\| \operatorname{proj}_C(-\lambda x)\|(\|v\| - \| \operatorname{proj}_C(-\lambda x)\|) && (*) \\ &\leqslant \lambda^{-1}\| \operatorname{proj}_C(-\lambda x)\| \|v - \operatorname{proj}_C(-\lambda x)\| \end{aligned}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \underset{c\in C}{\operatorname{Argmin}} \ \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle. \tag{6}$$

Set $\delta_C := \sup_{(c_1,c_2)\in C^2} \|c_1 - c_2\| \geqslant 0$ and $\mu_C := \sup_{c\in C} \|c\| \geqslant 0$.

$$\begin{aligned} 0 \leqslant \langle \operatorname{proj}_C(-\lambda x) \mid x\rangle - \langle v \mid x\rangle &\leqslant \lambda^{-1}(\langle v \mid \operatorname{proj}_C(-\lambda x)\rangle - \|\operatorname{proj}_C(-\lambda x)\|^2) && (7)\\ &\leqslant \lambda^{-1}(\|v\|\|\operatorname{proj}_C(-\lambda x)\| - \|\operatorname{proj}_C(-\lambda x)\|^2) && (8)\\ &= \lambda^{-1}\|\operatorname{proj}_C(-\lambda x)\|(\|v\| - \|\operatorname{proj}_C(-\lambda x)\|) && (*)\\ &\leqslant \lambda^{-1}\|\operatorname{proj}_C(-\lambda x)\|\|v - \operatorname{proj}_C(-\lambda x)\| \end{aligned}$$

From the proposition,

$$\operatorname{proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c\in C} \ \langle c \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \operatorname{proj}_C(-\lambda x) \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle \leqslant \langle v \mid \operatorname{proj}_C(-\lambda x) + \lambda x\rangle. \tag{6}$$

Set $\delta_C := \sup_{(c_1,c_2)\in C^2} \|c_1 - c_2\| \geqslant 0$ and $\mu_C := \sup_{c\in C} \|c\| \geqslant 0$.

$$\begin{aligned}
0 \leqslant \langle \operatorname{proj}_C(-\lambda x) \mid x\rangle - \langle v \mid x\rangle &\leqslant \lambda^{-1}(\langle v \mid \operatorname{proj}_C(-\lambda x)\rangle - \|\operatorname{proj}_C(-\lambda x)\|^2) && (7)\\
&\leqslant \lambda^{-1}(\|v\|\|\operatorname{proj}_C(-\lambda x)\| - \|\operatorname{proj}_C(-\lambda x)\|^2) && (8)\\
&= \lambda^{-1}\|\operatorname{proj}_C(-\lambda x)\|(\|v\| - \|\operatorname{proj}_C(-\lambda x)\|) && (*)\\
&\leqslant \lambda^{-1}\|\operatorname{proj}_C(-\lambda x)\|\|v - \operatorname{proj}_C(-\lambda x)\|\\
&\leqslant \lambda^{-1}\delta_C\mu_C
\end{aligned}$$

From the proposition,

$$\mathrm{proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c \in C} \ \langle c \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{5}$$

So, for any $v \in \mathrm{LMO}_C(x)$,

$$\langle \mathrm{proj}_C(-\lambda x) \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle \leqslant \langle v \mid \mathrm{proj}_C(-\lambda x) + \lambda x \rangle. \tag{6}$$

Set $\delta_C := \sup_{(c_1,c_2) \in C^2} \|c_1 - c_2\| \geqslant 0$ and $\mu_C := \sup_{c \in C} \|c\| \geqslant 0$.

$$\begin{aligned} 0 \leqslant \langle \mathrm{proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle &\leqslant \lambda^{-1}(\langle v \mid \mathrm{proj}_C(-\lambda x)\rangle - \|\mathrm{proj}_C(-\lambda x)\|^2) && (7)\\ &\leqslant \lambda^{-1}(\|v\| \|\mathrm{proj}_C(-\lambda x)\| - \|\mathrm{proj}_C(-\lambda x)\|^2) && (8)\\ &= \lambda^{-1}\|\mathrm{proj}_C(-\lambda x)\|(\|v\| - \|\mathrm{proj}_C(-\lambda x)\|) && (*)\\ &\leqslant \lambda^{-1}\|\mathrm{proj}_C(-\lambda x)\|\|v\| \\ &\leqslant \lambda^{-1}\mu_C^2 \end{aligned}$$

Theorem (Projection as Approximate LMO: Explicit error bound; W. 2025)

Let $x \in \mathcal{H}$ and let C be a nonempty, compact, and convex subset of $\mathcal{H}$ with diameter $\delta_C := \sup_{(c_1,c_2)\in C^2} \|c_1 - c_2\| \geqslant 0$ and bound $\mu_C := \sup_{c\in C} \|c\| \geqslant 0$. Then, for every $\lambda > 0$ and every $v \in \mathrm{LMO}_C(x)$,

$$0 \leqslant \langle \mathrm{proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C} \langle c \mid x\rangle \leqslant \frac{\|\mathrm{proj}_C(-\lambda x)\|}{\lambda}\Big(\|v\| - \|\mathrm{proj}_C(-\lambda x)\|\Big). \qquad (*)$$

In consequence, we have $\|\mathrm{proj}_C(-\lambda x)\| \leqslant \|v\|$ and for every $\varepsilon > 0$,

$$\lambda \geqslant \frac{\min\{\delta_C\mu_C, \mu_C^2\}}{\varepsilon} \quad \Rightarrow \quad 0 \leqslant \langle \mathrm{proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C}\langle c \mid x\rangle \leqslant \varepsilon. \qquad (9)$$

Theorem (Projection as Approximate LMO: Explicit error bound; W. 2025)

Let $x \in \mathcal{H}$ and let C be a nonempty, compact, and convex subset of $\mathcal{H}$ with diameter $\delta_C := \sup_{(c_1,c_2)\in C^2} \|c_1 - c_2\| \geqslant 0$ and bound $\mu_C := \sup_{c\in C} \|c\| \geqslant 0$. Then, for every $\lambda > 0$ and every $v \in \mathrm{LMO}_C(x)$,

$$0 \leqslant \langle \mathrm{proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C} \langle c \mid x\rangle \leqslant \frac{\|\mathrm{proj}_C(-\lambda x)\|}{\lambda}\Big(\|v\| - \|\mathrm{proj}_C(-\lambda x)\|\Big). \qquad (*)$$

In consequence, we have $\|\mathrm{proj}_C(-\lambda x)\| \leqslant \|v\|$ and for every $\varepsilon > 0$,

$$\lambda \geqslant \frac{\min\{\delta_C\mu_C, \mu_C^2\}}{\varepsilon} \quad \Rightarrow \quad 0 \leqslant \langle \mathrm{proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C}\langle c \mid x\rangle \leqslant \varepsilon. \qquad (9)$$

If $\mathrm{proj}_C(-\lambda^* x) \in \mathrm{LMO}_C(x)$, then it is the minimal-norm element of $\mathrm{LMO}_C(x)$.

Corollary (Projection is no faster than approximate LMO)

Let $\varepsilon > 0$ and suppose that Assumption 1 holds. Then $P + 1 \geqslant L(\varepsilon)$. In consequence, if $P \geqslant 1$, we also have

$$\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$$

.

Proof.

$L(\varepsilon)$ is bounded above by the cost of evaluating $\text{proj}_C(-\lambda x)$ which is $P + 1$. □

Corollary (Projection is no faster than approximate LMO)

Let $\varepsilon > 0$ and suppose that Assumption 1 holds. Then $P + 1 \geqslant L(\varepsilon)$. In consequence, if $P \geqslant 1$, we also have

$$\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$$

.

Proof.

$L(\varepsilon)$ is bounded above by the cost of evaluating $\mathrm{proj}_C(-\lambda x)$ which is $P + 1$. □

Drawbacks:

- For some sets, $\varepsilon \searrow 0$ means $\lambda \nearrow +\infty$, so this result cannot be used to compare exact LMO to exact projection in general.
- What about comparing exact LMO to exact projection?

What about exact LMO?

Is there a finite λ^* such that $\text{proj}_C(-\lambda^* x) \in \text{LMO}_C(x)$?

What about exact LMO?

Is there a finite λ^* such that $\text{proj}_C(-\lambda^* x) \in \text{LMO}_C(x)$?

Proposition (Projection is no faster than exact LMO on polyhedral sets; W. 2025)

Let $x \in \mathbb{R}^n =: \mathcal{H}$ and suppose that $C \subset \mathcal{H}$ is compact, convex, and polyhedral. Then there exists a finite value $\lambda^* \geqslant 0$ such that $\text{proj}_C(-\lambda^* x) \in \text{LMO}_C(x)$. Further, if Assumption 1 holds, then $P + 1 \geqslant L(0)$; if $P \geqslant 1$, then $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.

What about exact LMO?

Is there a finite λ^* such that $\mathrm{proj}_C(-\lambda^* x) \in \mathrm{LMO}_C(x)$?

Proposition (Projection is no faster than exact LMO on polyhedral sets; W. 2025)

Let $x \in \mathbb{R}^n =: \mathcal{H}$ and suppose that $C \subset \mathcal{H}$ is compact, convex, and polyhedral. Then there exists a finite value $\lambda^* \geqslant 0$ such that $\mathrm{proj}_C(-\lambda^* x) \in \mathrm{LMO}_C(x)$. Further, if Assumption 1 holds, then $P + 1 \geqslant L(0)$; if $P \geqslant 1$, then $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.

Proof idea: Partial dualization + strong duality argument, à la [Geoffrion, 1971] (and [Theorem 11.5, Güler, 2010]): there exists $\lambda^* > 0$ such that (w/ $\nu = \min_{v \in C} \langle v \mid x \rangle$)

$$\mathrm{proj}_{\mathrm{LMO}_C(x)}(\mathbf{0}) = \underset{\substack{z \in C \\ \langle z | x \rangle \leqslant \nu}}{\mathrm{minimize}} \ \frac{1}{2}\|z\|^2 = \underset{z \in C}{\mathrm{Argmin}} \ \frac{1}{2}\|z\|^2 + \lambda^*(\langle x \mid z \rangle - \nu)$$

$$= \underset{z \in C}{\mathrm{Argmin}} \ \frac{1}{2}\|-\lambda^* x - z\|^2 = \mathrm{proj}_C(-\lambda^* x)$$

What about exact LMO?

Is there a finite λ^* such that $\mathrm{proj}_C(-\lambda^* x) \in \mathrm{LMO}_C(x)$?

Proposition (Projection is no faster than exact LMO on polyhedral sets; W. 2025)

Let $x \in \mathbb{R}^n =: \mathcal{H}$ and suppose that $C \subset \mathcal{H}$ is compact, convex, and polyhedral. Then there exists a finite value $\lambda^* \geqslant 0$ such that $\mathrm{proj}_C(-\lambda^* x) \in \mathrm{LMO}_C(x)$. Further, if Assumption 1 holds, then $P + 1 \geqslant L(0)$; if $P \geqslant 1$, then $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.

Proof idea: Partial dualization + strong duality argument, à la [Geoffrion, 1971] (and [Theorem 11.5, Güler, 2010]): there exists $\lambda^* > 0$ such that (w/ $\nu = \min_{v \in C} \langle v \mid x \rangle$)

$$\mathrm{LMO}_C(x) \ni \mathrm{proj}_{\mathrm{LMO}_C(x)}(\mathbf{0}) = \underset{\substack{z \in C \\ \langle z | x \rangle \leqslant \nu}}{\mathrm{minimize}} \ \frac{1}{2}\|z\|^2 = \underset{z \in C}{\mathrm{Argmin}} \ \frac{1}{2}\|z\|^2 + \lambda^*(\langle x \mid z \rangle - \nu)$$

$$= \underset{z \in C}{\mathrm{Argmin}} \ \frac{1}{2}\| - \lambda^* x - z\|^2 = \mathrm{proj}_C(-\lambda^* x)$$

Linear Minimization v. Projections: Which is faster?

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)
- Polyhedral sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)
- Polyhedral sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.
- Current evidence suggests $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$ for nuclear norm ball, ~~ℓ_1 ball, probability simplex, Birkhoff polytope, general LP~~

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)
- Polyhedral sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.
- Current evidence suggests $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$ for nuclear norm ball, ~~ℓ_1 ball, probability simplex, Birkhoff polytope, general LP~~

Open question:

Does there exist **any** nonempty compact convex set such that $P < L(0)$?

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)
- Polyhedral sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.
- Current evidence suggests $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$ for nuclear norm ball, ~~ℓ_1 ball, probability simplex, Birkhoff polytope, general LP~~

Open question:

> Does there exist **any** nonempty compact convex set such that $P < L(0)$?

My guess: “no” if $\dim \mathcal{H}$ is finite

Recap

- All compact convex sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(\varepsilon))$ for $\varepsilon > 0$ (This cannot yield a result on comparison with $L(0)$ in general – some sets require $\lambda \nearrow +\infty$)
- Polyhedral sets satisfy $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$.
- Current evidence suggests $\mathcal{O}(P) \geqslant \mathcal{O}(L(0))$ for nuclear norm ball, ~~ℓ_1 ball, probability simplex, Birkhoff polytope, general LP~~

Open question:

Does there exist **any** nonempty compact convex set such that $P < L(0)$?

My guess: "no" if $\dim \mathcal{H}$ is finite

Contact: woodstzc[at]jmu.edu

Preprint (Currently under revision): arXiv:2501.18454

Thank you for your attention!

References

M. Besançon, M. Carderera, and S. Pokutta, FrankWolfe. jl: A high-performance and flexible toolbox for Frank–Wolfe algorithms and conditional gradients
INFORMS J. Comput., vol. 34 (5), pp. 2611–2620, 2022.

C. W. Combettes and S. Pokutta, Complexity of linear minimization and projection on some sets
Oper. Res. Lett., vol. 49, no. 4, pp. 565–571, 2021

J. Dunn and S. Harshbarger, Conditional gradient algorithms with open loop step size rules
J. Math. Anal. Appl., vol. 62, pp. 432–444, 1978.

R. M. Freund and P. Grigas, New analysis and results for the Frank-Wolfe method
Math. Program., Ser. A, vol. 155, pp. 199–230, 2016.

D. Garber, A. Kaplan, and S. Sabach, Improved complexities of conditional gradient-type methods with applications to robust matrix recovery problems
Math. Program., vol. 186, pp. 185–208, 2021.

A. M. Geoffrion, Duality in nonlinear programming: A simplified applications-oriented development
SIAM Rev., vol. 13 (1) pp. 1–37, 1971.

References

O. Güler *Foundations of Optimization*,
Springer, New York, 2010.

M. Jaggi, Revisiting Frank-Wolfe: Projection-free sparse convex optimization
Proc. 30th International Conference on Machine Learning, in PMLR, vol. 28 (1), pp. 427–435, 2013.

H. Mortagy, S. Gupta, and S. Pokutta, Walking in the shadow: A new perspective on descent directions for constrained minimization
Proc. 34th Conference on Neural Information Processing Systems, vol. 33, pp. 12873–12883, 2020.

A. Silveti-Falls, C. Molinari and J. Fadili, Inexact and stochastic generalized conditional gradient with augmented Lagrangian and proximal step
J. Nonsmooth Anal. Optim., vol. 2, 2021.