High-precision linear minimization is no slower than projection

Zev Woodstock[1*]

[1*]Department of Mathematics and Statistics, James Madison University, MSC 1911, 60 Bluestone Dr., Harrisonburg, VA, 22807, USA.

Corresponding author(s). E-mail(s): woodstzc@jmu.edu;

Abstract

This note demonstrates that, for all compact convex sets, high-precision linear minimization can be performed via a single evaluation of the projection and a scalar-vector multiplication. In consequence, if ε-approximate linear minimization takes at least $L(\varepsilon)$ vector-arithmetic operations and projection requires P operations, then $\mathcal{O}(P) \geq \mathcal{O}(L(\varepsilon))$ is guaranteed. This concept is expounded with examples, an explicit error bound, and an exact linear minimization result for polyhedral sets.

Keywords: linear minimization, projection, Frank-Wolfe, complexity

MSC Classification: 90C30 , 03D15 , 47N10 , 52A07

1 Introduction

Notation 1. *$\mathcal{H}$ is a real Hilbert space with inner product $\langle \cdot \mid \cdot \rangle$ and induced norm $\|\cdot\|$. The set $C \subset \mathcal{H}$ is nonempty, convex, and compact. The* normal cone *and* indicator function *of C are denoted N_C and ι_C respectively (see, e.g., [2]). The* projection *operator onto C is denoted* $\mathrm{Proj}_C \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \mathrm{Argmin}_{c\in C}\|c - x\|$. *The set of linear minimization oracle points is* $\mathrm{LMO}_C \colon \mathcal{H} \to 2^{\mathcal{H}} \colon x \mapsto \mathrm{Argmin}_{c\in C}\langle c \mid x\rangle$. *For $\varepsilon \geq 0$, an* ε-approximate LMO of x *is a point $v \in C$ such that* $0 \leq \langle v \mid x\rangle - \min_{c\in C}\langle c \mid x\rangle \leq \varepsilon$.

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.* [1]

[1] P and $L(\varepsilon)$ exist as limit points of monotonic sequences in $\mathbb{N}$ bounded below.

In constrained first-order optimization, a guiding motivator for the development of Frank-Wolfe algorithms is the fact that their hallmark subroutine, the *linear minimization oracle* (i.e., a selection of the operator LMO_C), is currently faster than projection oracles on some sets arising in applications, particularly in high-dimensional settings [3, 4, 6]. While several works suggest that for specific sets, $P > L(0)$, this principle does not appear to have been definitively established for all compact convex sets. In fact, regardless of approximateness ($\varepsilon > 0$) or exactness ($\varepsilon = 0$), it appears there are no results pertaining to all compact convex sets that allow one to compare the computational complexity of linear minimization to that of projection.

The main contribution of this article is showing that, for any $\varepsilon > 0$, a high-precision ε-approximate LMO can be obtained via the use of one projection and a scalar-vector multiplication, yielding the complexity bound $\mathcal{O}(P) \geq \mathcal{O}(L(\varepsilon))$.[2] The error bound in Theorem 2 explicitly depends on the radius and boundedness of C. It is further demonstrated that, when C is also polyhedral, projection is no faster than exact linear minimization, i.e., the stronger inequality $\mathcal{O}(P) \geq \mathcal{O}(L(0))$ holds. The central approximation considered in this article comes from a geometric concept (see Figure 1) that is known (e.g., see [10]). Nonetheless, as far as the author is aware, the present error bound and complexity results appear to be new.

2 Relating linear minimization and projection

We begin with some basic facts; see, e.g., [2] for further background. Let x and z be points in a real Hilbert space $\mathcal{H}$. Then,

$$v \in \mathrm{LMO}_C(z) = \underset{c \in C}{\mathrm{Argmin}} \langle c \mid z \rangle \Leftrightarrow 0 \in \partial \left(\langle \cdot \mid z \rangle + \iota_C \right)(v) = z + N_C(v) \Leftrightarrow -z \in N_C(v)$$
$$\Leftrightarrow \begin{cases} v \in C \\ \sup\limits_{c \in C} \langle -z \mid c - v \rangle \leq 0. \end{cases} \tag{1}$$

Similarly, we have the following familiar identity

$$p = \mathrm{Proj}_C\, x \Leftrightarrow x - p \in N_C p \Leftrightarrow \begin{cases} p \in C \\ \sup\limits_{c \in C} \langle x - p \mid c - p \rangle \leq 0. \end{cases} \tag{2}$$

Proposition 1. *Let $C \subset \mathcal{H}$ be a nonempty compact convex set. Then, for every $x \in \mathcal{H}$,*

$$\mathrm{Proj}_C(x) \in \mathrm{LMO}_C(\mathrm{Proj}_C\, x - x). \tag{3}$$

Proof. By setting $z = \mathrm{Proj}_C\, x - x$ in (1), we see from (2) that $\mathrm{Proj}_C(x)$ is a solution of the characterization in (1). □

Remark 1. *Since linear minimization is not unique in general, evaluating $\mathrm{LMO}_C(\mathrm{Proj}_C\, x - x)$ may not yield $\mathrm{Proj}_C(x)$ for every numerical implementation of*

[2] For the sake of presentation, we suppress the dependence of P and L on C and the dimension of $\mathcal{H}$.

LMO_C *[1]. However, for a particular selection of the operator* LMO_C*, (3) is guaranteed to hold with equality instead of inclusion.*

For $\lambda > 0$ sufficiently large, one can approximate an element of $\mathrm{LMO}_C(x)$ with

$$\mathrm{Proj}_C(-\lambda x). \tag{4}$$

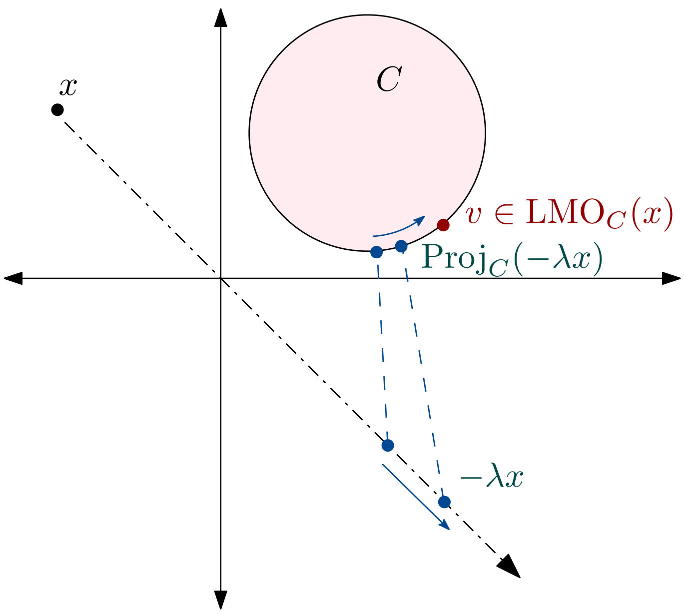

Fig. 1 As λ gets larger, $\mathrm{Proj}_C(-\lambda x)$ approaches $\mathrm{LMO}_C(x)$ for a shifted ℓ_2 ball.

Selecting the parameter λ can be guided as follows.

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

$$0 \leq \langle \mathrm{Proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C} \langle c \mid x\rangle \leq \frac{\|\mathrm{Proj}_C(-\lambda x)\|}{\lambda}\Big(\|v\| - \|\mathrm{Proj}_C(-\lambda x)\|\Big). \tag{5}$$

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

$$\lambda \geq \frac{\min\left\{\delta_C\mu_C, \mu_C^2\right\}}{\varepsilon} \quad \Rightarrow \quad 0 \leq \langle \mathrm{Proj}_C(-\lambda x) \mid x\rangle - \min_{c\in C}\langle c \mid x\rangle \leq \varepsilon. \tag{6}$$

Proof. Let $v \in \mathrm{LMO}_C(x)$. By Proposition 1, and the definition of the LMO,

$$\mathrm{Proj}_C(-\lambda x) \in \operatorname*{Argmin}_{c\in C} \langle c \mid \mathrm{Proj}_C(-\lambda x) + \lambda x\rangle, \tag{7}$$

so, for all $c \in C$, $\langle \operatorname{Proj}_C(-\lambda x) \mid \operatorname{Proj}_C(-\lambda x) + \lambda x \rangle \leq \langle c \mid \operatorname{Proj}_C(-\lambda x) + \lambda x \rangle$. In particular,

$$\langle \operatorname{Proj}_C(-\lambda x) \mid \operatorname{Proj}_C(-\lambda x) + \lambda x \rangle \leq \langle v \mid \operatorname{Proj}_C(-\lambda x) + \lambda x \rangle. \tag{8}$$

Dividing by λ and rearranging, then proceeding with standard norm inequalities yields

$$\begin{aligned} \langle \operatorname{Proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle &\leq \lambda^{-1}\big(\langle v \mid \operatorname{Proj}_C(-\lambda x)\rangle - \|\operatorname{Proj}_C(-\lambda x)\|^2\big) & (9)\\ &\leq \lambda^{-1}\big(\|v\|\|\operatorname{Proj}_C(-\lambda x)\| - \|\operatorname{Proj}_C(-\lambda x)\|^2\big) & (10)\\ &= \lambda^{-1}\|\operatorname{Proj}_C(-\lambda x)\|\big(\|v\| - \|\operatorname{Proj}_C(-\lambda x)\|\big) & (11)\\ &\leq \lambda^{-1}\|\operatorname{Proj}_C(-\lambda x)\|\|v - \operatorname{Proj}_C(-\lambda x)\|. & (12) \end{aligned}$$

Since v is optimal, we can also lower bound (9) by 0; hence (5) follows from (11). An immediate consequence of (5) is that $\|\operatorname{Proj}_C(-\lambda x)\| \leq \|v\|$. Proceeding with the bounds in (12), we have

$$0 \leq \langle \operatorname{Proj}_C(-\lambda x) \mid x \rangle - \langle v \mid x \rangle \leq \lambda^{-1}\delta_C\mu_C. \tag{13}$$

On the other hand, dropping the negative term in (11) implies $0 \leq \langle \operatorname{Proj}_C(-\lambda x) \mid x\rangle - \langle v \mid x \rangle, \leq \lambda^{-1}\mu_C^2$. Hence, in either case of $\lambda \geq \varepsilon^{-1}\min\{\delta_C\mu_C, \mu_C^2\}$, (6) holds. In particular, $\operatorname{Proj}_C(-\lambda x)$ is an ε-approximate LMO. □

Corollary 1 (Projection is no faster than approximate LMO)**.** *Let $\varepsilon > 0$ and suppose that Assumption 1 holds. Then $P + 1 \geq L(\varepsilon)$. In consequence, if $P \geq 1$, we also have $\mathcal{O}(P) \geq \mathcal{O}(L(\varepsilon))$.*

Proof. From Theorem 2, $L(\varepsilon)$ is bounded above by the cost of evaluating $\operatorname{Proj}_C(-\lambda x)$, which is $P + 1$. □

Remark 2. *The type of bound provided by Theorem 2 is consistent with many algorithms that allow for inexact linear minimization, e.g., [4, 5, 9, 11].*

In general, Theorem 2 requires $\lambda \to \infty$ to drive the error bound to zero (as demonstrated, e.g., for a shifted ℓ_2 ball). However, for some sets (e.g., the ℓ_∞ ball), exact minimization is achieved for finite λ, i.e., $\operatorname{Proj}_C(-\lambda x) \in \operatorname{LMO}_C(x)$.[3] As will be seen in Proposition 3, exact linear minimization for finite λ can be achieved more generally.

Proposition 3 (Projection is no faster than exact LMO on polyhedral sets)**.** *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 $\lambda^* \geq 0$ such that $\operatorname{Proj}_C(-\lambda^* x) \in \operatorname{LMO}_C(x)$. Further, if Assumption 1 holds, then $P + 1 \geq L(0)$; if $P \geq 1$, then $\mathcal{O}(P) \geq \mathcal{O}(L(0))$.*

Proof. Without loss of generality $C = \big\{x \in \mathcal{H} \;\big|\; Ax \preceq b\big\}$ where $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^m$. Let $\nu = \min_{c\in C}\langle c \mid x\rangle$ and consider the problem of computing $\operatorname{Proj}_{\operatorname{LMO}_C(x)}(\mathbf{0})$,

[3] Interactive graph demonstrations in 2-D for the shifted ℓ_2 ball: https://www.desmos.com/calculator/ntpe1pncpu and ℓ_∞ ball: https://www.desmos.com/calculator/qk3tnqskgw.

i.e., the minimal-norm element of $\mathrm{LMO}_C(x)$:

$$\underset{\substack{z\in\mathcal{H}\\ Az\preceq b\\ \langle z|x\rangle\leq\nu}}{\text{minimize}}\ \frac{1}{2}\|z\|^2. \tag{14}$$

By [8, Theorem 11.15], strong duality holds and hence the perturbation function $\boldsymbol{p}\colon \mathbb{R}^{m+1}\to[-\infty,+\infty] : y\mapsto \inf_{z\in\mathcal{H}}\{f(x)\mid Az-b\preceq(y_i)_{i=1}^m;\ \ \langle z\mid x\rangle-\nu\leq y_{m+1}\}$ is *stable* in the sense of [7], i.e., $\boldsymbol{p}(\mathbf{0})$ is finite and there exists $M>0$ such that [7, pp. 8]

$$(\forall\xi>0)\quad \frac{\boldsymbol{p}(\mathbf{0})-\boldsymbol{p}(\xi,\ldots,\xi)}{\xi}\leq M. \tag{15}$$

With an eye towards using [7, Theorem 3], we will use (15) to show that the partially-dualized perturbation function $p\colon\mathbb{R}\to[-\infty,+\infty] : \xi\mapsto\inf_{z\in\mathcal{H};Az-b\preceq 0}\{f(x)\mid\langle x\mid z\rangle-\nu\leq\xi\}$ is also stable: this follows from the fact that $p(0)=\boldsymbol{p}(\mathbf{0})$ and, for all $\xi>0$, $p(\xi)\geq\boldsymbol{p}(\xi,\ldots,\xi)$, so

$$\frac{p(0)-p(\xi)}{\xi}=\frac{\boldsymbol{p}(\mathbf{0})-p(\xi)}{\xi}\leq\frac{\boldsymbol{p}(\mathbf{0})-\boldsymbol{p}(\xi,\ldots,\xi))}{\xi}\leq M, \tag{16}$$

as claimed. By [7, Theorem 3], there exists $\lambda^*\geq 0$ such that

$$\mathrm{Proj}_{\mathrm{LMO}_C(x)}(\mathbf{0})=\underset{\substack{z\in\mathcal{H}\\ Az\preceq b}}{\mathrm{Argmin}}\,\frac{1}{2}\|z\|^2+\lambda^*(\langle x\mid z\rangle-\nu) \tag{17}$$

$$=\underset{\substack{z\in\mathcal{H}\\ Az\preceq b}}{\mathrm{Argmin}}\,\frac{1}{2}\|-\lambda^*x-z\|^2=\mathrm{Proj}_C(-\lambda^*x), \tag{18}$$

where (18) makes use of the fact that the minimization is not changed by addition of the constant $\lambda^*\nu+\|\lambda^*x\|^2/2$. This establishes that, for finite $\lambda^*\geq 0$, $\mathrm{Proj}_C(-\lambda^*x)\in\mathrm{LMO}_C(x)$. In consequence, under Assumption 1, the computation requied to perform $\mathrm{Proj}_C(-\lambda^*x)$ (namely $P+1$) is an upper bound on $L(0)$, which completes the proof. □

3 Conclusion

While this note demonstrates two connections between the complexities of projection and linear minimization, the general relationship between P and $L(0)$ warrants further investigation.

For some sets (e.g., singletons), $P=L(0)$. Hence, even though current evidence suggests that $P>L(0)$ on a myriad of important sets [3], strict inequality cannot hold in general. Nonetheless, it remains an open question as to whether or not there exists *any* compact convex set such that its projection operator has a faster runtime complexity than exact linear minimization.

Question 1. *Does there exist a nonempty compact convex set such that $P<L(0)$?*

References

[1] 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.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, New York, 2017.

[3] C. W. Combettes and S. Pokutta, Complexity of linear minimization and projection on some sets, *Oper. Res. Lett.*, vol. 49 (4), pp. 565–571, 2021.

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

[5] 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.

[6] 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.

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

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

[9] 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.

[10] 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.

[11] 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.

Acknowledgements

Many thanks to Mathieu Besançon, who first suggested trying a duality approach for the proof of Proposition 3. Thanks to Jannis Halbey, Sebastian Pokutta, Elias Wirth, and the anonymous user 'gerw' (on Math Stack Exchange) for early conversations on this topic.

Data availability statement

No data was used in the creation of this article. Citations are available above.

Conflict of interest statement

The author declares no conflict of interest.