# Properties of Frank-Wolfe's Method and Linear Minimization Oracle

**Tao Hu**

October 1, 2025

**Abstract Keywords** Frank-Wolfe · Robust Algorithm · Projection · Linear Minimization Oracle

My final project mainly discusses the performance of Frank-Wolfe's method when a disturbed gradient is used. Also, I proved that an accurate linear minimization oracle is no slower than an inexact projection.

## 1 Introduction

We consider the maximization problems:

$$\begin{aligned} \max_{\lambda} \ & h(\lambda) \\ \text{s.t. } & \lambda \in Q \ , \end{aligned} \tag{1}$$

where $Q \subset E$ is convex and compact, and $h(\cdot) : Q \to \mathbb{R}$ is differentiable on $Q$.

In order to prove the convergence of the Frank-Wolfe algorithm, the following two auxiliary sequences are frequently used, where $\alpha_k$ and $\beta_k$ are determined by the first $k$ step-sizes, $\bar{\alpha}_1, \ldots, \bar{\alpha}_k$ in Algorithm 1:

$$\beta_k = \frac{1}{\prod_{j=1}^{k-1}(1-\bar{\alpha}_j)} \ , \qquad \alpha_k = \frac{\beta_k \bar{\alpha}_k}{1-\bar{\alpha}_k} \ , \qquad k \geq 1 \ . \tag{2}$$

(We follow the conventions: $\prod_{j=1}^{0} \cdot = 1$ and $\sum_{i=1}^{0} \cdot = 0$ .)

Tao Hu
Xi'an Jiaotong University
Tel.: +1 (510)-384-9593
E-mail: tao_hu@berkeley.edu

**Algorithm 1** Frank-Wolfe Method for maximizing $h(\lambda)$

Pick initial point $\lambda_0 \in Q$, initial upper bound $U_{init}$ (usually determined by prior knowledge, if no prior knowledge is assumed, then $U_{init} = +\infty$), $k \leftarrow 0$ .

At iteration $k$:

1. Evaluate $\nabla h(\lambda_k)$ .
2. Evaluate $\tilde{\lambda}_k \leftarrow \arg\max_{\lambda \in Q}\{h(\lambda_k) + \nabla h(\lambda_k)^T(\lambda - \lambda_k)\}$ .

   $U_k^w \leftarrow h(\lambda_k) + \nabla h(\lambda_k)^T(\tilde{\lambda}_k - \lambda_k)$ .
   $G_k \leftarrow \nabla h(\lambda_k)^T(\tilde{\lambda}_k - \lambda_k)$ .
   $U_k \leftarrow \min\{U_{k-1}, U_k^w\}$, if $k \geq 1$.
   $U_0 \leftarrow \min\{U_{init}, U_0^w\}$, if $k = 0$.
3. Set $\lambda_{k+1} \leftarrow \lambda_k + \bar{\alpha}_k(\tilde{\lambda}_k - \lambda_k)$, where $\bar{\alpha}_0 = 1$ and $\bar{\alpha}_k \in [0, 1), \forall\, k \geq 1$ .

## 2 Frank-Wolfe Method for smooth convex functions

Assume that $h$ is concave and its gradient is $L_{h,Q}-$ Lipschitz, that is,

$$\|\nabla h(\lambda) - \nabla h(\bar{\lambda})\|_* \leq L_{h,Q}\|\lambda - \bar{\lambda}\|,\ \forall\ \lambda, \bar{\lambda} \in Q\ .$$

We denote $C_{h,Q} = L_{h,Q}(Diam_Q)^2$.

In this scenario, several results have been discussed by Freund and Grigas [1].

**Theorem 1** *If the step-size sequence $\{\bar{\alpha}_k\}$ is used in the iterate sequences of the Frank-Wolfe method (Method 1), then the following inequality holds for all $k \geq 0$,*

$$U_k - h(\lambda_{k+1}) \leq \frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \frac{\frac{1}{2}C_{h,Q}\sum_{i=1}^{k}\frac{\alpha_i^2}{\beta_{i+1}}}{\beta_{k+1}}\ . \tag{3}$$

*Proof*

$$\begin{aligned} h(\lambda_{k+1}) =& h(\lambda_k) + \nabla h(\lambda_k)^T(\lambda_{k+1} - \lambda_k) - \frac{1}{2}L_{h,Q}\|\lambda_{k+1} - \lambda_k\|^2 \\ =& h(\lambda_k) + \bar{\alpha}_k\nabla h(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k) - \frac{1}{2}\bar{\alpha}_k^2L_{h,Q}\|\widetilde{\lambda}_k - \lambda_k\|^2 \\ \geq& h(\lambda_k) + \bar{\alpha}_k\nabla h(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k) - \frac{1}{2}\bar{\alpha}_k^2C_{h,Q}. \end{aligned}$$

We have that

$$U_k^w = h(\lambda_k) + \nabla h(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k).$$

Therefore,

$$\begin{aligned} h(\lambda_{k+1}) \geq& (1 - \bar{\alpha}_k)h(\lambda_k) + \bar{\alpha}_k(h(\lambda_k) + \nabla h(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k)) - \frac{1}{2}\bar{\alpha}_k^2C_{h,Q} \\ =& (1 - \bar{\alpha}_k)h(\lambda_k) + \bar{\alpha}_kU_k^w - \frac{1}{2}\bar{\alpha}_k^2C_{h,Q}. \end{aligned}$$

Hence,

$$\begin{aligned}\beta_{k+1}h(\lambda_{k+1}) \geq&\beta_{k+1}(1-\bar{\alpha}_k)h(\lambda_k)+\bar{\alpha}_k\beta_{k+1}U_k^w-\frac{1}{2}\bar{\alpha}_k^2\beta_{k+1}C_{h,Q}\\ =&\beta_k h(\lambda_k)+(\beta_{k+1}-\beta_k)U_k^w-\frac{1}{2}\frac{\alpha_k^2}{\beta_{k+1}}C_{h,Q}\\ \geq&\beta_k h(\lambda_k)+(\beta_{k+1}-\beta_k)U_k-\frac{1}{2}\frac{\alpha_k^2}{\beta_{k+1}}C_{h,Q}.\end{aligned}$$

Therefore,

$$\begin{aligned}\beta_{k+1}(U_k-h(\lambda_{k+1})) \leq&\beta_k(U_k-h(\lambda_k))+\frac{1}{2}\frac{\alpha_k^2}{\beta_{k+1}}C_{h,Q}\\ \leq&\beta_k(U_{k-1}-h(\lambda_k))+\frac{1}{2}\frac{\alpha_k^2}{\beta_{k+1}}C_{h,Q}.\end{aligned}$$

Taking summation, we get that

$$\beta_{k+1}(U_k-h(\lambda_{k+1}))\leq\beta_1(U_0-h(\lambda_1))+\frac{1}{2}\sum_{j=1}^{k}\frac{\alpha_j^2}{\beta_{j+1}}C_{h,Q}.$$

We conclude that, since $\beta_{k+1}>0$,

$$U_k-h(\lambda_{k+1})\leq\frac{U_0-h(\lambda_1)}{\beta_{k+1}}+\frac{C_{h,Q}\sum_{j=1}^{k}\frac{\alpha_j^2}{\beta_{j+1}}}{2\beta_{k+1}}.$$

□

## 2.1 Inexact Gradient

Assuming that we are using a $\delta$-oracle, that is, every time we are doing a gradient evaluation, we don't get an exact gradient; instead, we get a "gradient" $g_\delta(\lambda)$ with error $\delta/(DiamQ)$, that is,

$$\|g_\delta(\lambda)-\nabla h(\lambda)\|\leq\frac{\delta}{DiamQ},\forall\,\lambda\in Q.$$

We conclude that, $\forall\,\lambda_1,\lambda_2\in Q$,

$$\begin{aligned}h(\lambda_2)\leq&h(\lambda_1)+\nabla h(\lambda_1)^T(\lambda_2-\lambda_1)\\ \leq&h(\lambda_1)+g_\delta(\lambda_1)^T(\lambda_2-\lambda_1)+\|g_\delta(\lambda_1)-\nabla h(\lambda_1)\|\|\lambda_2-\lambda_1\|\\ \leq&h(\lambda_1)+g_\delta(\lambda_1)^T(\lambda_2-\lambda_1)+\delta;\end{aligned}$$

$$\begin{aligned}h(\lambda_2)\geq&h(\lambda_1)+\nabla h(\lambda_1)^T(\lambda_2-\lambda_1)-\frac{1}{2}L_{h,Q}\|\lambda_2-\lambda_1\|^2\\ \geq&h(\lambda_1)+g_\delta(\lambda_1)^T(\lambda_2-\lambda_1)-\|g_\delta(\lambda_1)-\nabla h(\lambda_1)\|\|\lambda_2-\lambda_1\|-\frac{1}{2}L_{h,Q}\|\lambda_2-\lambda_1\|^2\\ \geq&h(\lambda_1)+g_\delta(\lambda_1)^T(\lambda_2-\lambda_1)-\delta-\frac{1}{2}L_{h,Q}\|\lambda_2-\lambda_1\|^2;\end{aligned}$$

**Algorithm 2** Frank-Wolfe Method Using $\delta$-Oracle

Pick initial point $\lambda_0 \in Q$, initial upper bound $U_{init}$ (usually determined by prior knowledge, if know prior knowledge is assumed, then $U_{init} = +\infty$), $k \leftarrow 0$ .

At iteration $k$:

1. Evaluate $g_\delta(\lambda_k)$ .
2. Evaluate $\tilde{\lambda}_k \leftarrow \arg\max_{\lambda \in Q}\{h(\lambda_k) + g_\delta(\lambda_k)^T(\lambda - \lambda_k)\}$ .

   $U_k^w \leftarrow h(\lambda_k) + \delta + g_\delta(\lambda_k)^T(\tilde{\lambda}_k - \lambda_k)$ .

   $U_k \leftarrow \min\{U_{k-1}, U_k^w\}$, if $k \geq 1$.

   $U_0 \leftarrow \min\{U_{init}, U_0^w\}$, if $k = 0$.
3. Set $\lambda_{k+1} \leftarrow \lambda_k + \bar{\alpha}_k(\tilde{\lambda}_k - \lambda_k)$, where $\bar{\alpha}_0 = 1$ and $\bar{\alpha}_k \in [0, 1), \forall\, k \geq 1$ .

Under inexact gradient evaluation, the bound is given as follows [1],

**Theorem 2** *If the step-size sequence $\{\bar{\alpha}_k\}$ is used in the iterate sequences of the Frank-Wolfe method with the $\delta$-oracle (Algorithm 2),*

$$U_k - h(\lambda_{k+1}) \leq \frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \frac{\frac{1}{2}C_{h,Q}\sum_{i=1}^{k}\frac{\alpha_i^2}{\beta_{i+1}}}{\beta_{k+1}} + \frac{2\delta\sum_{i=1}^{k}\beta_{i+1}}{\beta_{k+1}}, \forall\, k \geq 0. \quad (4)$$

*Proof*

$$\begin{aligned}
h(\lambda_{k+1}) \geq & h(\lambda_k) + g_\delta(\lambda_k)^T(\lambda_{k+1} - \lambda_k) - \delta - \frac{1}{2}L_{h,Q}\|\lambda_{k+1} - \lambda_k\|^2 \\
= & h(\lambda_k) + \bar{\alpha}_k g_\delta(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k) - \delta - \frac{1}{2}\bar{\alpha}_k^2 L_{h,Q}\|\widetilde{\lambda}_k - \lambda_k\|^2 \\
\geq & (1 - \bar{\alpha}_k)h(\lambda_k) + \bar{\alpha}_k(h(\lambda_k) + g_\delta(\lambda_k)^T(\widetilde{\lambda}_k - \lambda_k) + \delta) - (1 + \bar{\alpha}_k)\delta - \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2 \\
\geq & (1 - \bar{\alpha}_k)h(\lambda_k) + \bar{\alpha}_k U_k^w - 2\delta - \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2 \\
\geq & (1 - \bar{\alpha}_k)h(\lambda_k) + \bar{\alpha}_k U_k - 2\delta - \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2
\end{aligned}$$

Hence,

$$\begin{aligned}
U_k - h(\lambda_{k+1}) \leq & U_k - (1 - \bar{\alpha}_k)h(\lambda_k) - \bar{\alpha}_k U_k + 2\delta + \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2 \\
= & (1 - \bar{\alpha}_k)(U_{k-1} - h(\lambda_k)) + 2\delta + \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2.
\end{aligned}$$

Therefore,

$$\beta_k(U_k - h(\lambda_{k+1})) \leq \beta_{k-1}(U_{k-1} - h(\lambda_k)) + 2\delta\beta_k + \frac{1}{2}C_{h,Q}\bar{\alpha}_k^2\beta_k.$$

By taking summation, we get that

$$\beta_k(U_k - h(\lambda_{k+1})) \leq (U_0 - h(\lambda_1)) + 2\delta\sum_{j=1}^{k}\beta_j + \frac{1}{2}C_{h,Q}\sum_{j=1}^{k}\bar{\alpha}_j^2\beta_j.$$

To conclude,

$$U_k - h(\lambda_{k+1}) \leq \frac{U_0 - h(\lambda_1)}{\beta_k} + 2\delta \frac{\sum_{j=1}^{k} \beta_j}{\beta_k} + \frac{1}{2} C_{h,Q} \sum_{j=1}^{k} \bar{\alpha}_j^2 \beta_j.$$

Dividing both side by $\beta_k$ yields the result. □

Consider the right hand side of Equation (4), we have that

$$\begin{aligned}
&\frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \frac{\frac{1}{2} C_{h,Q} \sum_{i=1}^{k} \frac{\alpha_i^2}{\beta_{i+1}}}{\beta_{k+1}} + \frac{\sum_{i=1}^{k} \beta_{i+1}\delta}{\beta_{k+1}} \\
=&\frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \frac{\sum_{i=1}^{k} (\frac{1}{2} C_{h,Q} \frac{\alpha_i^2}{\beta_{i+1}} + \beta_{i+1}\delta)}{\beta_{k+1}} \\
\geq&\frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \frac{\sum_{i=1}^{k} \alpha_i \sqrt{2C_{h,Q}\delta}}{\beta_{k+1}} \\
=&\frac{U_0 - h(\lambda_1)}{\beta_{k+1}} + \sqrt{2C_{h,Q}\delta} \frac{\beta_{k+1} - 1}{\beta_{k+1}}.
\end{aligned}$$

For $k$ sufficiently large, we have that

$$0 < \frac{\beta_2 - 1}{\beta_2} \leq \frac{\beta_{k+1} - 1}{\beta_{k+1}} \leq 1.$$

Therefore, in order to achieve a first-order accurate solution, we need a second-order accurate gradient.

## 3 Frank-Wolfe Method with constant step sizes

We now take a special example of the results above.

### 3.1 Concave Functions with Lipschitz Gradient

Given $\bar{\alpha} \in (0, 1)$, consider using the following constant step-size rule:

$$\bar{\alpha}_i = \bar{\alpha} \qquad \text{for } i \geq 1 \ . \tag{5}$$

.

**Proposition 1** *Under the step-size sequence (5), the following inequality holds for all $k \geq 1$:*

$$U_k - h(\lambda_{k+1}) \leq (U_k - h(\lambda_1))\,(1 - \bar{\alpha})^k + \tfrac{1}{2} C_{h,Q} \left[\bar{\alpha} - \bar{\alpha}(1 - \bar{\alpha})^k\right] \ . \tag{6}$$

*As a consequence:*

$$U_k - h(\lambda_{k+1}) \leq \tfrac{1}{2} C_{h,Q} \left[(1 - \bar{\alpha})^{k+1} + \bar{\alpha}\right] \ . \tag{7}$$

It can be easily observed that the right hand side is a convex function. By taking derivatives and finding the stationary point, we can get that the optimal $\bar{\alpha}$ to choose is

$$\bar{\alpha} = 1 - \frac{1}{\sqrt[k]{k+1}}, \tag{8}$$

if we fix the iteration number before the algorithm starts.

The following proposition and the idea of prestart is provided by Freund and Grigas [1].

**Proposition 2** $U_0 - h(\lambda_1) \le \frac{1}{2}C_{h,Q}$.

*Proof* We have $\lambda_1 = \tilde{\lambda}_0$ and $U_0 \le U_0^w$, whereby from the definition of $C_{h,Q}$ using $\alpha = 1$ we have:

$$h(\lambda_1) = h(\tilde{\lambda}_0) \ge h(\lambda_0) + \nabla h(\lambda_0)^T(\tilde{\lambda}_0 - \lambda_0) - \tfrac{1}{2}C_{h,Q} = U_0^w - \tfrac{1}{2}C_{h,Q} \ge U_0 - \tfrac{1}{2}C_{h,Q}\ ,$$

and the result follows by rearranging terms.

When we are using a constant step size $\bar{\alpha}_k = \bar{\alpha}$, and we take $\delta_i = \delta$, then we have that

$$\beta_k = (1-\bar{\alpha})^{-k+1}, \alpha_k = \bar{\alpha}(1-\bar{\alpha})^{-k}.$$

By substitute them in, we get that

$$\begin{aligned}
&U_k - h(\lambda_{k+1}) \\
\le&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \frac{1}{2}\mathrm{Diam}_Q^2(1-\bar{\alpha})^k \sum_{i=1}^k L_i\bar{\alpha}^2(1-\bar{\alpha})^{-i} + (1-\bar{\alpha})^k \sum_{i=1}^k (1-\bar{\alpha})^{-i}\delta \\
\le&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \frac{1}{2}\mathrm{Diam}_Q^2(1-\bar{\alpha})^k \sum_{i=1}^k L\bar{\alpha}^2(1-\bar{\alpha})^{-i} + (1-\bar{\alpha})^k \sum_{i=1}^k (1-\bar{\alpha})^{-i}\delta \\
=&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \frac{1}{2}\mathrm{Diam}_Q^2(1-\bar{\alpha})^k \sum_{i=1}^k L\bar{\alpha}^2(1-\bar{\alpha})^{-i} + (1-\bar{\alpha})^k \sum_{i=1}^k (1-\bar{\alpha})^{-i}\delta \\
=&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \left(\frac{1}{2}\mathrm{Diam}_Q^2 L\bar{\alpha}^2 + \delta\right)(1-\bar{\alpha})^k \sum_{i=1}^k (1-\bar{\alpha})^{-i} \\
=&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \left(\frac{1}{2}\mathrm{Diam}_Q^2 L\bar{\alpha}^2 + \delta\right)\frac{1-(1-\bar{\alpha})^k}{\bar{\alpha}} \\
=&(1-\bar{\alpha})^{k-1}(U_k - h(\lambda_1)) + \left(\frac{1}{2}\mathrm{Diam}_Q^2 L\bar{\alpha} + \frac{\delta}{\bar{\alpha}}\right)\left(1-(1-\bar{\alpha})^k\right).
\end{aligned}$$

By Proposition 2, we can proceed that,

$$
\begin{aligned}
&U_k - h(\lambda_{k+1})\\
\leq&\frac{1}{2}(1-\bar{\alpha})^{k-1}C_{h,Q} + \left(\frac{1}{2}C_{h,Q}\bar{\alpha} + \frac{\delta}{\bar{\alpha}}\right)\left(1-(1-\bar{\alpha})^k\right)\\
\leq&\frac{1}{2}C_{h,Q}(1-\bar{\alpha})^{k-1} + \left(\frac{1}{2}C_{h,Q}\bar{\alpha} + \frac{\delta}{\bar{\alpha}}\right)\\
=&\frac{1}{2}C_{h,Q}\left((1-\bar{\alpha})^{k-1} + \bar{\alpha}\right) + \frac{\delta}{\bar{\alpha}}.
\end{aligned}
$$

If we take $\bar{\alpha} = 1-(k-1)^{-1/(k-2)}$, then

$$
\bar{\alpha} = 1-(k-1)^{-1/(k-2)} = 1-e^{-\frac{\log(k-1)}{k-2}} \in \left[\frac{1}{e}\frac{\log(k-1)}{k-2}, \frac{\log(k-1)}{k-2}\right]
$$

$$
\begin{aligned}
&U_k - h(\lambda_{k+1})\\
\leq&\frac{1}{2}C_{h,Q}\left((1-\bar{\alpha})^{k-1} + \bar{\alpha}\right) + \frac{\delta}{\bar{\alpha}}\\
\leq&\frac{1}{2}C_{h,Q}\left((1-\bar{\alpha})^{k-2} + \bar{\alpha}\right) + \frac{\delta}{\bar{\alpha}}\\
=&\frac{1}{2}C_{h,Q}\left(\frac{1}{k-1} + 1 - (k-1)^{-\frac{1}{k-2}}\right) + \frac{\delta}{\bar{\alpha}}\\
=&\frac{1}{2}C_{h,Q}\left(\frac{1}{k-1} + 1 - e^{-\frac{\log(k-1)}{k-2}}\right) + \frac{\delta}{\bar{\alpha}}\\
\leq&\frac{1}{2}C_{h,Q}\left(\frac{1}{k-1} + \frac{\log(k-1)}{k-2}\right) + \frac{\delta}{\bar{\alpha}}
\end{aligned}
$$

Therefore, in order to achieve a convergence rate of $\widetilde{O}(\frac{1}{k})$, it suffices to let

$$
\delta \sim O(\frac{\bar{\alpha}}{k}) = O(\frac{\log(k)}{k^2}).
$$

## 4 A Comparison between the Projection Problem and the Linear Minimization Problem

### 4.1 Coarse Projection is More Expensive than High-accuracy Linear Minimization

**Definition 1** Suppose that $\mathcal{H}$ is a real Hilbert space equipped with inner product $\langle \cdot \mid \cdot \rangle$ which induces the norm $\|\cdot\|$. Assume that $C \subset \mathcal{H}$ is a nonempty compact convex set. From the compactness of $C$, we are allowed to define the projection operator and the set of linear minimization oracle points separately as follows:

$$
\mathrm{Proj}_C \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \mathrm{Argmin}_{c\in C}\|c-x\|,
$$

$$\mathrm{LMO}_C\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \mathrm{Argmin}_{c\in C}\langle c \mid x\rangle.$$

In the studies of robust algorithms, the errors in both the projection and the minimization should be considered. For $\varepsilon \geq 0$, we define the set of $\epsilon$-approximate projection oracle points and $\varepsilon$-approximate linear minimization oracle as follows:

$$\varepsilon\text{-}\operatorname{Proj}_C : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \{v \in \mathcal{H} | 0 \leq \frac{1}{2}\|v - x\|^2 - \min_{c\in C} \frac{1}{2}\|c - x\|^2 \leq \varepsilon\},$$
$$\varepsilon\text{-}\operatorname{LMO}_C : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \{v \in \mathcal{H} | 0 \leq \langle v \mid x\rangle - \min_{c\in C} \langle c \mid x\rangle \leq \varepsilon\}.$$

The definition above actually provides four common minimization problems. An interesting question is which one of them is the hardest and which one of them is the easiest. Zev Woodstock [2] proved that an exact projection evaluation would be no easier than finding a point in $\varepsilon\text{-}\operatorname{LMO}_C$, no matter how small $\varepsilon$ is. The main contribution of this chapter is to give a small modification of Woodstock's proof and prove that finding a point in $K\text{-}\operatorname{Proj}_C$ would be no easier than finding a point in $\varepsilon\text{-}\operatorname{LMO}_C$.

**Proposition 3** *Suppose that $C \subset \mathcal{H}$ is a nonempty compact convex set and $x \in \mathcal{H}$, and that $p \in \varepsilon\text{-}\operatorname{Proj}(x)$. Then*

$$\langle c - p \mid x - p\rangle \leq \varepsilon + \frac{1}{2}\|c - p\|^2, \forall\, c \in C.$$

*Proof* Since $p \in \varepsilon\text{-}\operatorname{Proj}(x)$,

$$\frac{1}{2}\|p - x\|^2 \leq \frac{1}{2}\|c - x\|^2 + \varepsilon, \forall\, c \in C.$$

This could be written in the inner product form,

$$\frac{1}{2}\langle c - p \mid 2x - p - c\rangle \leq \varepsilon, \forall\, c \in C.$$

We thus get that

$$\langle c - p \mid x - p\rangle \leq \varepsilon + \frac{1}{2}\|c - p\|^2, \forall\, c \in C.$$

□

**Theorem 3** *Let $x \in \mathcal{H}$ and $C \subset \mathcal{H}$ be nonempty, convex and compact. We denote $\delta_C := \sup_{(c_1,c_2)\in C^2} \|c_1 - c_2\| \geq 0$ and $\mu_C := \sup_{c\in C} \|c\| \geq 0$. Take any positive real number $K$ and take $v \in \mathrm{LMO}_C(x)$. For any $\lambda > 0$, take*

$$p' \in K\text{-}\operatorname{Proj}_C(-\lambda x).$$

*Then we have that,*

$$0 \leq \langle p' \mid x\rangle - \langle v \mid x\rangle \leq \frac{K + \delta_C^2/2 + \min\{\mu_C\delta_C, \mu_C^2\}}{\lambda}.$$

*In consequence, we have that for any $\varepsilon > 0$, if $\lambda \geq \frac{K+\delta_C^2/2+\min\{\mu_C\delta_C,\mu_C^2\}}{\varepsilon}$, then*

$$0 \leq \langle p' \mid x\rangle - \min_{c\in C} \langle c \mid x\rangle \leq \epsilon.$$

*We conclude that*

$$K\text{-}\mathrm{Proj}(-\lambda x) \subset \varepsilon\text{-}\mathrm{LMO}(x).$$

*Proof* By proposition 3, we have that

$$\langle c - p' \mid -\lambda x - p'\rangle \leq K + \frac{1}{2}\|c - p'\|^2, \forall\, c \in C.$$

Then, and take $c = v$,

$$\lambda\langle p' \mid x\rangle - \lambda\langle v \mid x\rangle \leq K + \frac{1}{2}\|v - p'\|^2 + \langle p' \mid v - p'\rangle.$$

Next,

$$\begin{aligned}\lambda\langle p' - v \mid x\rangle \leq& K + \frac{1}{2}\|v - p'\|^2 + \langle p' \mid v - p'\rangle\\ =& K + \frac{1}{2}\|v - p'\|^2 + (\langle v \mid p'\rangle - \langle p' \mid p'\rangle)\\ \leq& K + \frac{1}{2}\|v - p'\|^2 + \|p'\|(\|v\| - \|p'\|)\\ \leq& K + \frac{1}{2}\|v - p'\|^2 + \|p'\|\|v - p'\|.\end{aligned}$$

Therefore,

$$\lambda\langle p' - v \mid x\rangle \leq K + \frac{1}{2}\delta_C^2 + \min\{\mu_C\delta_C, \mu_C^2\}.$$

Hence,

$$0 \leq \langle p' \mid x\rangle - \langle v \mid x\rangle \leq \frac{K + \frac{1}{2}\delta_C^2 + \min\{\mu_C\delta_C, \mu_C^2\}}{\lambda},$$

where the first inequality is from the fact that $v \in \mathrm{LMO}_C(x)$. □

## 4.2 Comparison Theory for exact problem solvers

**Definition 2** Assume that $C \subset \mathcal{H}$ is a nonempty compact convex set, $0 \in C$, and let

$$f_C : \mathcal{H} \to 2^C\ ,\ g_C : \mathcal{H} \to 2^C.$$

Suppose that $I_C : \mathcal{H} \to \{0, +\infty\}$ is the indicator function of $C$, that is

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C\\ \infty & \text{if } x \notin C\end{cases}$$

We say that evaluating $f_c$ is not significantly harder that evaluating $g_C$, denoted as $f_C \preceq g_C$, if there is a function $F : 2^{\mathcal{H}} \to 2^{\mathcal{H}}$, which does not depend on $C$, and a sequence of points $\{x_n : \mathcal{H} \to \mathcal{H}\}_{n=1}^{+\infty}$, such that $\forall\ x \in \mathcal{H}$, $\exists N(x) \in \mathbb{Z}_{\geq 1}$, such that

$$f_C(x) \supset F(g_C(x_n(x))), \forall\ n \geq N(x).$$

Under this notation we can write proposition 3 of [2] in the following precise way,

**Proposition 4** *Take $\mathcal{H} = \mathbb{R}^n$, then for any $C \in \mathcal{P} := \{\{x \in \mathcal{H} | Ax \leq b\} | A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^m\}$, we have that* $\mathrm{LMO}_C \preceq \mathrm{Proj}_C$.

## References

1. R. M. Freund and P. Grigas, *New analysis and results for the frank–wolfe method*, Mathematical Programming, 155 (2016), p. 199–230.
2. Z. Woodstock, *High-precision linear minimization is no slower than projection*, (2025). arXiv:2501.18454 [math].