A More on Frank-Wolfe-type Algorithms

A.1 Standard Frank-Wolfe Algorithm

Algorithm 2 Frank-Wolfe algorithm (Frank et al., 1956)

Input: obj. $f : \mathcal{Y} \mapsto \mathbb{R}$, oracle $O(\cdot)$, init. $\boldsymbol{w}_0 \in \mathcal{Y}$

for t=1, 2, 3 $\ldots$, T **do**
  $\boldsymbol{v}_t \leftarrow \text{Oracle}(\nabla f(\boldsymbol{w}_{t-1})) = \arg\min_{\boldsymbol{v}\in\mathcal{Y}} \boldsymbol{v}^T \nabla f(\boldsymbol{w}_{t-1})$
  $\boldsymbol{w}_t \leftarrow (1 - \eta_t \boldsymbol{w}_{t-1}) + \eta_t \boldsymbol{v}_t$, for $\eta := \frac{2}{t+2}$
end for
return $\boldsymbol{w}_T$

For a convex function $f : \mathcal{X} \mapsto \mathbb{R}$ the Frank-Wolfe algorithm (FW) solves the constrained optimization problem over a compact and convex set $\mathcal{X}$. The standard FW is known to has sublinear convergence rate, and various methods are proposed to improve the convergence rate. For example, when the underlying feasible set is a polytope, and the objective function is strongly convex, multiple variants, such as away-step FW (Wolfe, 1970; Jaggi, 2013), pairwise FW (Mitchell et al., 1974), and Wolfe's method (Wolfe, 1976) are shown to enjoy linear convergence rate.

A.2 Wolfe's Method for Minimum Norm Point

Algorithm 3 Wolfe's Method for Minimum Norm Point

Initialize $\boldsymbol{x} \in \mathcal{P}$, active set $\mathcal{S} = [\boldsymbol{x}]$ and weight $\boldsymbol{\lambda} = [1]$.
Output: $\boldsymbol{x} \in \mathcal{P}$ that has the minimum Euclidean norm.

while true **do** // Major cycle
  $\boldsymbol{s} \leftarrow \texttt{Oracle}(\boldsymbol{x})$ // Potential improving point
  if $||\boldsymbol{x}||^2 \leq \boldsymbol{x}^T\boldsymbol{s} + \epsilon$ **then** break
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathrm{s}\}$
  while true **do** // Minor cycle
    $\boldsymbol{y}, \boldsymbol{\alpha} \leftarrow \texttt{AffineMinimizer}(\mathcal{S})$ // $\boldsymbol{y} = \arg\min_{\boldsymbol{s}\in\texttt{aff}(\mathcal{S})} ||\boldsymbol{s}||_2$
    if $\alpha_{\boldsymbol{s}} > 0$ for all $\boldsymbol{s}$ **then** break // $\boldsymbol{y} \in \texttt{conv}(\mathcal{S})$
    // If $\boldsymbol{y} \notin \texttt{conv}(\mathcal{S})$, then update $\boldsymbol{y}$ to the intersection of $\texttt{conv}(\mathcal{S})$ and segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$. Then remove points in $\mathcal{S}$ unnecessary for describing $\boldsymbol{y}$.
    $\theta \leftarrow \min_{i:\alpha_i\leq 0} \frac{\lambda_i}{\lambda_i - \alpha_i}$ // Recall $\boldsymbol{\lambda}$ satisfies $\boldsymbol{x} = \sum_{\boldsymbol{s}\in\mathcal{S}} \lambda_{\boldsymbol{s}}\boldsymbol{s}$
    $\boldsymbol{y} \leftarrow \theta\boldsymbol{y} + (1-\theta)\boldsymbol{x}, \lambda_i = \theta\alpha_i + (1-\theta)\lambda_i$
    $\mathcal{S} \leftarrow \{\boldsymbol{s}_i | \boldsymbol{s}_i \in \mathcal{S}$ and $\lambda_i > 0\}$
  end while
  Update $\boldsymbol{x} = \boldsymbol{y}$ and $\boldsymbol{\lambda} = \boldsymbol{\alpha}$.
end while
return $\boldsymbol{x}$

Wolfe's method is an iterative algorithm for finding the point with minimum Euclidean norm in a polytope, which is defined as the convex hull of a set of finitely many points.

The Wolfe's method consists of a finite number of major cycles, each of which consists of a finite number of minor cycles. At the start of each major cycle, let $H(\boldsymbol{x}) := \{\boldsymbol{y}^T\boldsymbol{x} = \boldsymbol{x}^{\boldsymbol{x}}\}$ be the hyperplane defined by $\boldsymbol{x}$. If $H(\boldsymbol{x})$ separates the polytope from the origin, then the major cycle is terminated. Otherwise, we invoke an oracle to find any point that on the near side of the hyperplane. The point is then added into the active set $\mathcal{S}$, and minor cycle starts.

In a minor cycle, let $\boldsymbol{y}$ be the point of smallest norm in of the affine hull $\texttt{aff}(\mathcal{S})$. If $\boldsymbol{y}$ is in the relative interior of the convex hull $\texttt{conv}(\mathcal{S})$, the $\boldsymbol{x}$ is updated to $\boldsymbol{y}$ and the minor cycle is terminated. Otherwise, $\boldsymbol{y}$ is updated to the nearest point to $\boldsymbol{y}$ on the line segment $\texttt{conv}(\mathcal{S}) \cap [\boldsymbol{x}, \boldsymbol{y}]$. Thus $\boldsymbol{y}$ is updated to a boundary point of $\texttt{conv}(\mathcal{S})$, and any point that is not on the face of $\texttt{conv}(\mathcal{S})$ in which $\boldsymbol{y}$ lies are deleted. The minor cycles executed repeatedly until the $\mathcal{S}$ becomes a *corral*, that is, a set whose affine minimizer lies inside its convex hull. Since a set of one point is always a corral, the minor cycles is terminated after a finite number of runs.

B PROOF OF THEOREM 4.1

Theorem 4.1 (Approximation Error Strictly Decreases)**.** *For any non-terminal step t, we have $\mathtt{err}(\mu^{t+1}) < \mathtt{err}(\mu^t)$. That is, the measurement vector of μ^t found by Algorithm 1 gets strictly closer to the convex set Ω after major cycle step.*

Proof. If the current step is a major cycle with no minor cycle, then $\boldsymbol{x}^{t+1}$ is the affine minimizer of $\mathtt{aff}(\mathcal{S} \cup \{\boldsymbol{s}^t\})$ with respect to $\boldsymbol{\omega}^t$. Then the affine minimizer property implies $(\boldsymbol{s}^t - \boldsymbol{x}^{t+1})(\boldsymbol{x}^{t+1} - \omega^t) = 0$. Since iteration does not terminate at step t, we have $(\boldsymbol{x}^t - \boldsymbol{\omega}^t)^T(\boldsymbol{x}^t - \boldsymbol{s}^t) > 0$, and therefore $\boldsymbol{x}^{t+1}$ not equal to $\boldsymbol{x}^t$. Then $\boldsymbol{x}^{t+1}$ is the unique affine minimizer implies $f_\Omega(\boldsymbol{x}^{t+1}) = \min_{\omega\in\Omega} ||\boldsymbol{x}^{t+1} - \omega||^2 \leq ||\boldsymbol{x}^{t+1} - \omega^t||^2 < ||\boldsymbol{x}^t - \omega^t||^2 = f_\Omega(\boldsymbol{x}^t)$.

Otherwise the current step contains one or more minor cycles. In this case, we show that the first minor cycle strictly reduces the approximation error, and the (possibly) following minor cycles cannot increase it. For the first minor cycle, the affine minimizer $\boldsymbol{y}^0$ of $\mathtt{aff}(\mathcal{S} \cup \{\boldsymbol{s}^t\})$ with respect to $\boldsymbol{\omega}^t$ is outside $\mathtt{conv}(\mathcal{S} \cup \{\boldsymbol{s}^t\})$. Let $\boldsymbol{z} = \theta\boldsymbol{y}^0 + (1-\theta)\boldsymbol{x}^t$ be the intersection of $\mathtt{conv}(\mathcal{S} \cup \{\boldsymbol{s}_t\})$ and segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$. Let $\mathcal{V}^0 := \mathcal{S}^t$ and $\mathcal{V}^i$ denotes the active set after the i-th minor cycle. Then since $\boldsymbol{y}^1$ is the affine minimizer of $\mathcal{V}^1$ with respect to ω^t, we have

$$||\boldsymbol{z} - \omega^t|| = ||\theta\boldsymbol{y}^0 + (1-\theta)\boldsymbol{x}^t - \omega^t|| \leq \theta||\boldsymbol{y}^0 - \omega^t|| + (1-\theta)||\boldsymbol{x}^t - \omega^t|| < ||\boldsymbol{x}^t - \omega^t||, \quad (14)$$

where the second step uses the triangle inequality and the last step follows since the segment $\boldsymbol{x}^t\boldsymbol{y}^0$ intersects the interior of $\mathtt{conv}(\mathcal{S} \cup \{\boldsymbol{s}^t\})$, and the distance to ω^t strictly decreases along this segment. Therefore the point $\boldsymbol{z}$ found by first minor cycle satisfies

$$f_\Omega(\boldsymbol{z}) = \min_{\omega\in\Omega} ||\boldsymbol{z} - \omega||^2 \leq ||\boldsymbol{z} - \omega^t||^2 < ||\boldsymbol{x}^t - \omega^t|| = f_\Omega(\boldsymbol{x}^t). \quad (15)$$

Hence $h(\boldsymbol{y}^1) < h(\boldsymbol{x}^t)$, and the first minor cycle strictly decreases the approximation error. By a similar argument, in subsequent minor cycles the approximation error cannot be increased. However, after the first minor cycle, the iterating point may already at the intersection point and the strict inequality in last step of Eq. 14 need to be replaced by non-strict inequality.

Therefore any major cycle either finds an improving point and continue, or enter minor cycles where the first minor cycle finds an improving point, and the subsequent minor cycles does not increase the distance. Adding both side of $f_\Omega(\boldsymbol{x}^{t+1}) < f_\Omega(\boldsymbol{x}^t)$ by $f_\Omega(\boldsymbol{x}^*)$ and we have the approximation error $h(\boldsymbol{x}^{t+1}) < h(\boldsymbol{x}^t)$ strictly decreases. □

C PROOF OF THEOREM 4.2

We first prove the Theorem 4.2, using Lemma 4.3 and Lemma 4.4. Then present the proof of the lemmas.

Theorem 4.2 (Convergence in Approximation Error)**.** *For $t \geq 1$, the mixed policy μ^t found by Algorithm 1 satisfies*

$$\mathtt{err}(\mu^t) \leq 16Q^2/(t+2). \quad (16)$$

where $Q := \max_{\mu\in\Delta(\mathcal{U})} ||\boldsymbol{c}(\mu)||$ is the maximum norm of a measurement vector.

Proof. Since Lemma 4.4 shows that drop steps are no more than half of total major cycle steps, and Theorem 4.1 guarantees these drop steps reducing the approximation error, we can safely skip these step, and re-index the step numbers to include non-drop steps only using k.

For these non-drop steps, we claim that $\mathtt{err}(\mu^k) \leq 8Q^2/(k+1)$. Using Lemma 4.3, we prove the convergence rate using induction. We first bound the error of any $\mathtt{err}(\mu^k)$. For any $k \geq 1$

$$\begin{aligned}
\mathtt{err}(\mu^k) &= \mathtt{dist}(\boldsymbol{c}(\mu^k), \Omega) - \mathtt{dist}(\boldsymbol{c}(\mu^*), \Omega) && (17)\\
&= 1/2||\boldsymbol{c}(\mu^k) - \mathtt{Proj}_\Omega(\boldsymbol{c}(\mu^k))||^2 - 1/2||\boldsymbol{c}(\mu^*) - \mathtt{Proj}_\Omega(\boldsymbol{c}(\mu^*))||^2 && (18)\\
&\leq 1/2(||\boldsymbol{c}(\mu^k)||^2 + ||\mathtt{Proj}_\Omega(\boldsymbol{c}(\mu^k))||^2 - ||\boldsymbol{c}(\mu^*)||^2 - ||\mathtt{Proj}_\Omega(\boldsymbol{c}(\mu^*))||^2) && (19)\\
&\leq ||\boldsymbol{c}(\mu^k)||^2 - ||\boldsymbol{c}(\mu^*)||^2 && (20)\\
&\leq ||\boldsymbol{c}(\mu^k)||^2 && (21)\\
&\leq Q^2, && (22)
\end{aligned}$$

where Eq. 18 uses the definition of our squared Euclidean distance function. Eq. 19 follows from triangle inequality, and Eq. 20 is by the contractive property of the Euclidean distance.

When $k = 1$, the Eq. 22 established the based case. Now for $k \geq 1$, assume that $\texttt{err}(\mu^k) \leq 8Q^2/(k+1)$ for $k \geq 1$, then Lemma 4.3 gives $\texttt{err}(\mu^{k+1}) \leq \texttt{err}(\mu^k) - \texttt{err}^2(\mu^k)/8Q^2$. Since the quadratic function of the right hand side is monotonically increasing on $(-\infty, 4Q^2]$, using the inductive hypothesis

$$\texttt{err}(\mu^{k+1}) \leq \texttt{err}(\mu^k) - \texttt{err}^2(\mu^k)/8Q^2 \leq 8Q^2/(k+1) - 8Q^2/(k+1)^2 \leq Q^2/(k+2) \quad (23)$$

Then since for t steps of major cycle steps, the number of non-drop steps $k > t/2$, we conclude that $\texttt{err}(\mu^t) \leq 16Q^2/(t+2)$.

□

Then we prove the lemmas.

Lemma 4.3. *For a non-drop step, we have* $\texttt{err}(\mu^t) - \texttt{err}(\mu^{t+1}) \geq \texttt{err}^2(\mu^t)/8Q^2$.

Proof. The non-drop step contains either no minor cycle or one minor cycle. We first consider the no minor cycle case.

If a major cycle contains no minor cycle, then $\boldsymbol{x}^{t+1}$ is the affine minimizer of the $\mathcal{S} \cup \{\boldsymbol{s}^t\}$.

$$\begin{aligned}
\texttt{err}(\mu^t) - \texttt{err}(\mu^{t+1}) &= \texttt{dist}(\boldsymbol{x}^t, \Omega) - \texttt{dist}(\boldsymbol{x}^{t+1}, \Omega) &(24)\\
&= 1/2(||\boldsymbol{x}^t - \boldsymbol{\omega}^t||^2 - \min_{\boldsymbol{\omega}\in\Omega} ||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}||^2) &(25)\\
&\geq 1/2(||\boldsymbol{x}^t - \boldsymbol{\omega}^t||^2 - ||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t||^2) &(26)\\
&= 1/2(||\boldsymbol{x}^t - \boldsymbol{\omega}^t||^2 + ||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t||^2 - 2||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t||^2) &(27)\\
&= 1/2(||\boldsymbol{x}^t - \boldsymbol{\omega}^t||^2 + ||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t||^2 - 2(\boldsymbol{x}^t - \boldsymbol{\omega}^t)^T(\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t)) &(28)\\
&= 1/2(||\boldsymbol{x}^t - \boldsymbol{x}^{t+1}||^2), &(29)
\end{aligned}$$

where the equation (28) follows from the affine minimizer property Eq. (9). For $||\boldsymbol{x}^t - \boldsymbol{x}^{t+1}||$ in the last equation, and $\forall \boldsymbol{q} \in \texttt{aff}(\mathcal{S} \cup \{\boldsymbol{s}^t\})$, we have

$$\begin{aligned}
||\boldsymbol{x}^t - \boldsymbol{x}^{t+1}|| &\geq ||\boldsymbol{x}^t - \boldsymbol{x}^{t+1}|| \frac{||\boldsymbol{x}^t|| + ||\boldsymbol{q}||}{2Q} &&(\text{ Definition of } Q) &(30)\\
&\geq ||\boldsymbol{x}^t - \boldsymbol{x}^{t+1}|| \frac{||\boldsymbol{x}^t - \boldsymbol{q}||}{2Q} &&(\text{ Triangle inequality}) &(31)\\
&\geq \frac{1}{2Q}(\boldsymbol{x}^t - \boldsymbol{x}^{t+1})(\boldsymbol{x}^t - \boldsymbol{q}) &&(\text{ Cauchy-Schwarz inequality}) &(32)\\
&= \frac{1}{2Q}(\boldsymbol{x}^t - \boldsymbol{\omega}^t)(\boldsymbol{x}^t - \boldsymbol{q}) &&(\text{ Affine minimizer property}). &(33)
\end{aligned}$$

Then it suffices to show that $(\boldsymbol{x}^t - \boldsymbol{\omega}^t)(\boldsymbol{x}^t - \boldsymbol{q}) \geq \texttt{err}(\mu^t)$.

Since Ω is a convex set, the squared Euclidean distance function $\texttt{dist}(\boldsymbol{x}, \Omega)$ is convex for $\boldsymbol{x}$, which implies

$$\texttt{dist}(\boldsymbol{x}^t, \Omega) + (\boldsymbol{q} - \boldsymbol{x}^t)\nabla\texttt{dist}(\boldsymbol{x}^t, \Omega) \leq \texttt{dist}(\boldsymbol{q}, \Omega). \quad (34)$$

Putting in $\nabla\texttt{dist}(\boldsymbol{x}^t, \Omega) = (\boldsymbol{x}^t - \texttt{Proj}_\Omega(\boldsymbol{x}^t)) = (\boldsymbol{x}^t - \boldsymbol{\omega}^t)$, we get $(\boldsymbol{x}^t - \boldsymbol{\omega}^t)(\boldsymbol{x}^t - \boldsymbol{q}) \geq \texttt{err}(\mu^t)$, which together with Eq. 29 and Eq. 33 concludes that for non-drop step with no minor cycles, we have $\texttt{err}(\mu^t) - \texttt{err}(\mu^{t+1}) \geq \texttt{err}^2(\mu^t)/8Q^2$.

For non-drop step with one minor cycle, we use the Theorem 6 of (Chakrabarty et al., 2014). By a linear translation of adding all points with $-\omega^t$, it gives

$$||\boldsymbol{x}^t - \boldsymbol{\omega}^t||^2 - ||\boldsymbol{x}^{t+1} - \boldsymbol{\omega}^t||^2 \geq ((\boldsymbol{x}^t - \boldsymbol{\omega}^t)(\boldsymbol{x}^t - \boldsymbol{q}))^2/8Q^2. \quad (35)$$

Then applying the same argument as Eq. 34, and we finished our proof.

□

Lemma 4.4. *After t major cycle steps of Algorithm 1, the number of drop steps is less than $t/2$.*

Proof. Recall that at the termination of a minor cycle, the size of the active set $|\mathcal{S}_c| \in [1, m]$. Since in each major cycle steps, the size of active set $\mathcal{S}_t$ increases by one, and each drop step reduces the size of $\mathcal{S}_t$ by at least one, the number of drop steps is always less than half of total number of the major cycle steps. $\square$

D Proof of Theorem 4.5

Theorem 4.5 (Memory Complexity Bound)**.** *For an constrained RL problem with m-dimensional measurement vector, in the worst case, a mixed policy needs to randomize among $m + 1$ individual policies to ensure convergence of RL oracles that search for deterministic policies.*

Proof. We give a constructive proof. Consider a m-dimensional vector-valued MDP with a single state, $m + 1$ actions, and $\boldsymbol{c}(a_i) := \boldsymbol{e}_i$ is the unit vector of i-th dimension for $i \in [1, m]$, and $\boldsymbol{c}(a_{m+1}) := \mathbf{0}$, and the episode terminates after 1 steps. The constrained RL problem is to find a policy whose measurement vector lies in the convex set of a single point $\{\mathbf{1}/2m\}$. By linear programming, it is clear that the only feasible mixed deterministic policy is to select a_{m+1} with $1/2$ probability, and the rest m actions with $1/2m$ probability; i.e. the unique feasible policy to this problem has an active set containing $m + 1$ deterministic policies. Therefore any method randomize among less than $m + 1$ individual policies does not ensure convergence when used with RL algorithms searching for deterministic policies. $\square$