**Problem setup**

Consider optimization problem with the decision vector $x$:

\[
\begin{align*}
\text{minimize} & \quad f(x) := g(Ax) + \langle c, x \rangle \\
\text{subject to} & \quad x \in \Omega.
\end{align*}
\]

- $\Omega$ convex and compact with diameter $D$
- $g$ smooth
- $A$ a linear map, and $c$ a vector

Applications: LASSO, SVM, matrix completion, phase retrieval, and one-bit matrix completion, etc.

**Frank-Wolfe**

FW: choose $x_0 \in \Omega$, iterate

1. Linear Optimization Oracle (LOO): Find a direction $v_1$ that solves $\min_{v \in \Omega} \langle \nabla f(x), v \rangle$.
2. Line Search: Find $x_{t+1}$ that solves $\min_{x \in \Omega \cap \{1 - \theta(x, x_t)\}} f(x)$.

**Slow convergence of FW and Zigzag**

- FW: slow in both theory and practice, $O(\frac{1}{t})$ convergence rate.
- Zigzag: cause of slow convergence when when the optimal solution $x^*$ is not a convex combination of extreme points.

See Figure 1 for $r = 2$. The grey arrows are the negative gradients $-\nabla f$.

**Our key insight**

The sparsity $r_*$ is small in many applications and $\nabla f(x_* \Omega)$ has the smallest inner product with $v_1^*, \ldots, v_{r_*}^*$, among all $v \in \Omega$.

Our key insight:

- Compute all extreme points $v_1^*$ that minimize $\langle \nabla f(x), v \rangle$;
- Solve the smaller problem $\min_{v \in \Omega} \langle \nabla f(x_* \Omega), v \rangle$.

See Figure 2 for an illustration.

**Our key insight**

The sparsity $r_*$ is small in many applications and $\nabla f(x_* \Omega)$ has the smallest inner product with $v_1^*, \ldots, v_{r_*}^*$, among all $v \in \Omega$.

Our key insight:

- Compute all extreme points $v_1^*$ that minimize $\langle \nabla f(x), v \rangle$;
- Solve the smaller problem $\min_{v \in \Omega} \langle \nabla f(x_* \Omega), v \rangle$.

See Figure 2 for an illustration.

**Theoretical Result**

**Analytical Conditions**

- Sparsity measure $r_*$: number of extreme points of the smallest face $F(x_* \Omega)$ containing $x_* \Omega$.
- Strict complementarity (SC) and its measure $\delta$: a unique solution $x_* \in \partial \Omega$ and $\nabla f(x) \in \text{relint}(V_0, V_1)$ normal cone). The SC measure is $\delta = \min_{v \in \Omega} \langle \nabla f(x), v \rangle$ extreme point).
- Quadratic growth (QG): for all $x \in \Omega, f(x) - f(x_*) \geq \gamma \|x - x_*\|^2$.

**Theorem Statement**

Suppose $f$ is $L_f$-smooth and convex and $\Omega$ is convex compact with diameter $D$.

If then for any $k \geq 1$ and for all $t \geq 1$, the iterate $x_t \in k$FW satisfies $f(x_t) - f(x_*) \leq \frac{\delta \sqrt{L_f}}{kD}$.

Moreover, suppose Problem (1) satisfies strict complementarity and quadratic growth, and $k \geq r_*$. If the constraint set $\Omega$ is a polytope or a unit norm ball, then the gap $\delta > 0$ if $k$FW finds $x_*$ in at most $T + 1$ iterations, where $T$ is

\[ T = \frac{4L_f \delta}{\gamma \|x_*\|^2}. \]

If the constraint set is the spechedron or the unit nuclear norm ball, the gap $\delta > 0$ and $k$FW satisfies that for any $t \geq T_1 := \frac{\gamma \|x_*\|^2}{\delta}$, $f(x_{T_1}) - f(x_*) \leq \left(1 - \frac{1}{T_1 + 1}\right)(f(X_I) - f(X_I))$.

**Numerics**

We compare our method $k$FW with FW, away-step FW (awayFW), pairwise FW (pairFW), DICG [Garber and Meshi 2016], and blockFW [Allen-Zhu et al. 2017] for the Lasso, support vector machine (SVM), group Lasso, and matrix completion problems on synthetic data. All algorithms terminate when the relative change of the objective is less than $10^{-6}$ or after 1000 iterations. As shown in Figure 1, $k$FW converges in many fewer iterations than other methods. Table 1 shows that $k$FW also converges faster in wall-clock time, with one exception (blockFW in matrix completion).