Acceleration of Frank-Wolfe Algorithms with Open-Loop Step-Sizes

Elias Wirth [1] Thomas Kerdreux [2] Sebastian Pokutta [1,3]

[1]Berlin Institute of Technology, [2]Geolabe LLC, [3]Zuse Institute Berlin

Abstract

Frank-Wolfe algorithms (FW) are popular first-order methods for solving constrained convex optimization problems that rely on a linear minimization oracle instead of potentially expensive projection-like oracles. Many works have identified accelerated convergence rates under various structural assumptions on the optimization problem and for specific FW variants when using line-search or short-step, requiring feedback from the objective function. Little is known about accelerated convergence regimes when utilizing open-loop step-size rules, a.k.a. FW with pre-determined step-sizes, which are algorithmically extremely simple and stable. We derive several accelerated convergence results for FW with open-loop step-size rules and characterize a general setting for which FW with open-loop step-size rules converges non-asymptotically faster than FW with line-search or short-step. Numerical experiments show that vanilla FW with open-loop step-sizes can compete with momentum-based FW variants.

The Frank-Wolfe algorithm

We study the constrained convex optimization problem

$$\min_{x\in C} f(x), \qquad \text{(OPT)}$$

where $C \subseteq \mathbb{R}^d$ is a compact convex set and $f\colon C \to \mathbb{R}$ is a convex and L-smooth function. Let $x^* \in \operatorname{argmin}_{x\in C} f(x)$ be the constrained optimal solution. We address (OPT) with the *Frank-Wolfe algorithm* (FW) [4], which enjoys several attractive properties for practitioners who work at scale:

1. First-order.
2. Projection-free.
3. Affine-invariant.
4. Easy to implement.

Algorithm 1 Frank-Wolfe algorithm (FW) [4]

Input: $x_0 \in C$, step-size rule $\eta_t \in [0,1]$ for $t \in \{0,\ldots,T-1\}$.

```
for t = 0, ..., T−1 do
    p_t ∈ argmin_{p∈C} ⟨∇f(x_t), p − x_t⟩
    x_{t+1} ← (1 − η_t) x_t + η_t p_t
```

Why open-loop step-sizes $\eta_t = \frac{\ell}{t+\ell}$, where $\ell \in \mathbb{N}_{\geq 1}$?

1. Not governed by Wolfe's lower bound [12].
2. Problem-agnostic.
3. Easy to compute since no knowledge of the smoothness constant is required.

Numerical experiments: logistic regression

We consider the problem of logistic regression, which for feature vectors $a_1,\ldots,a_m \in \mathbb{R}^d$, label vector $b \in \{-1,+1\}^m$, $p \in \mathbb{R}_{\geq 1}$, and radius $r > 0$, leads to the problem formulation

$$\min_{x\in\mathbb{R}^d} \frac{1}{m}\sum_{i=1}^m \log(1+\exp(-b_i a_i^\intercal x))$$
$$\text{subject to } \|x\|_p \leq r.$$

For $p \in \{1,2,5\}$ and $r = 1$, we compare FW, the *primal-averaging Frank-Wolfe algorithm* (PAFW) [9], and the *momentum-guided Frank-Wolfe algorithm* (MFW) [11], with open-loop step-sizes $\eta_t = \frac{\ell}{t+\ell}$, where $\ell \in \{2,6\}$, on the Gisette dataset[a] [7]. Results are presented in Figure 1:

1. On uniformly convex sets, all algorithms converge at rates of order $\mathcal{O}(1/t^\ell)$.
2. Acceleration of momentum-based variants relies on choice of ℓ.
3. Vanilla FW with open-loop step-sizes competes with momentum-based variants.

[a]Available online at https://archive.ics.uci.edu/ml/datasets/Gisette.

Results

References	Region C	Objective f	Location of x^*	Rate	Step-size rule
[8]	-	-	unrestricted	$\mathcal{O}(1/t)$	any
[6]	-	strongly convex	interior	$\mathcal{O}(e^{-t})$	line-search, short-step
This paper	-	strongly convex	interior	$\mathcal{O}(1/t^2)$	open-loop $\eta_t = \frac{4}{t+4}$
[10, 2, 3]	strongly convex	$\|\nabla f(x)\|_2 \geq \lambda > 0$ for all $x \in C$	unrestricted	$\mathcal{O}(e^{-t})$	line-search, short-step
This paper	strongly convex	$\|\nabla f(x)\|_2 \geq \lambda > 0$ for all $x \in C$	unrestricted	$\mathcal{O}(1/t^2)$	open-loop $\eta_t = \frac{4}{t+4}$
This paper	strongly convex	$\|\nabla f(x)\|_2 \geq \lambda > 0$ for all $x \in C$	unrestricted	$\mathcal{O}(1/t^{\ell/2})$	open loop $\eta_t = \frac{\ell}{t+\ell}$ for $\ell \in \mathbb{N}_{\geq 4}$
[5]	strongly convex	strongly convex	unrestricted	$\mathcal{O}(1/t^2)$	line-search, short-step
This paper	strongly convex	strongly convex	unrestricted	$\mathcal{O}(1/t^2)$	open-loop $\eta_t = \frac{4}{t+4}$
[12]	polytope	strongly convex	interior of face	$\Omega(1/t^{1+\epsilon})^*$	line-search, short-step
[1]	polytope	strongly convex	interior of face	$\mathcal{O}(1/t^2)^*$	open-loop $\eta_t = \frac{2}{t+2}$
This paper	polytope	strongly convex	interior of face	$\mathcal{O}(1/t^2)$	open-loop $\eta_t = \frac{4}{t+4}$

Table 1: Comparison of convergence rates for the Frank-Wolfe algorithm under different assumptions, where $x^* \in \operatorname{argmin}_{x\in C} f(x)$.

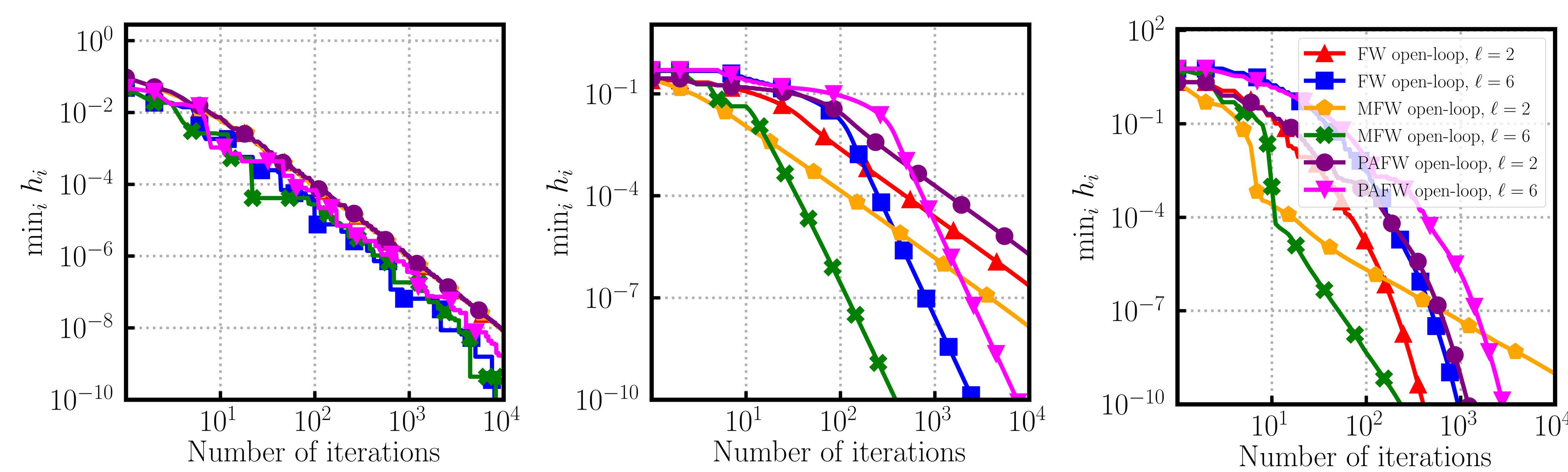

Figure 1: Logistic regression for different ℓ_p-balls, $p \in \{1,2,5\}$. The y-axis represents the minimum primal gap.

References

[1] Francis Bach. On the effectiveness of richardson extrapolation in data science. *SIAM Journal on Mathematics of Data Science*, 3(4):1251–1277, 2021.

[2] Vladimir Fedorovich Demianov and Aleksandr Moiseevich Rubinov. *Approximate methods in optimization problems*. Number 32. Elsevier Publishing Company, 1970.

[3] Joseph C Dunn. Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals. *SIAM Journal on Control and Optimization*, 17(2):187–211, 1979.

[4] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[5] Dan Garber and Elad Hazan. Faster rates for the Frank-Wolfe method over strongly-convex sets. In *Proceedings of the International Conference on Machine Learning*. PMLR, 2015.

[6] Jacques Guélat and Patrice Marcotte. Some comments on Wolfe's 'away step'. *Mathematical Programming*, 35(1):110–119, 1986.

[7] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182, 2003.

[8] Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the International Conference on Machine Learning*, pages 427–435. PMLR, 2013.

[9] Guanghui Lan. The complexity of large-scale convex programming under a linear optimization oracle. *arXiv preprint arXiv:1309.5550*, 2013.

[10] Evgeny S Levitin and Boris T Polyak. Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):1–50, 1966.

[11] Bingcong Li, Mario Coutino, Georgios B Giannakis, and Geert Leus. A momentum-guided frank-wolfe algorithm. *IEEE Transactions on Signal Processing*, 69:3597–3611, 2021.

[12] Philip Wolfe. Convergence theory in nonlinear programming. *Integer and Nonlinear Programming*, pages 1–36, 1970.