Frank-Wolfe Beyond $1/t$ Convergence

Sebastian Pokutta

Technische Universität Berlin & Zuse Institute Berlin pokutta@zib.de

Take-home. If the feasible region is curved around the minimizers, vanilla Frank-Wolfe beats $1/t$: $f(x_t) - f^\star = o(1/t)$ for every smooth convex f.

The Frank-Wolfe Algorithm

Problem. $\min_{x\in\mathcal{X}} f(x)$, f L-smooth convex, $\mathcal{X}$ compact convex, accessed only via a *linear minimization oracle* (LMO).

Algorithm. For $t = 0, 1, 2, \ldots$

$$s_t \leftarrow \operatorname*{argmin}_{v\in\mathcal{X}} \langle \nabla f(x_t), v\rangle \qquad \text{(LMO)}$$
$$x_{t+1} \leftarrow x_t + \gamma_t (s_t - x_t)$$

Step sizes. Short step, exact line search, or open loop $\gamma_t = 2/(t+2)$.

Rate. All three give, with $D = \operatorname{diam}\mathcal{X}$,

$$f(x_t) - f^\star \le \frac{2LD^2}{t+2} = O(1/t) \quad (t \ge 1).$$

[Frank & Wolfe 1956; Levitin & Polyak 1966; Jaggi 2013]

Simple, projection-free, but $O(1/t)$.

Where Does $1/t$ Come From?

Classical instance. $\mathcal{X} = \operatorname{conv}\{e_1, \ldots, e_n\}$, $f(x) = \frac{1}{2}\|x\|^2$.

Sparsity vs. error. After t steps from a vertex, x_t uses at most $t+1$ vertices, so

$$f(x_t) - f^\star = \Omega(1/t) \quad (n \text{ large}).$$

[Jaggi 2013; Lan 2013]

What this exploits. Flat faces: the LMO keeps returning distant vertices while progress stays small.

Known $\Omega(1/t)$ bounds rely on flat faces!

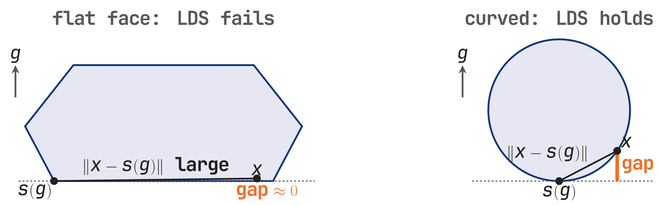
Question. With f only smooth and convex: can the *curvature of $\mathcal{X}$ alone* beat $1/t$?

Local Dual Sharpness

LMO atom. $s(g) \in \operatorname{argmin}_{y\in\mathcal{X}} \langle g, y\rangle$, for a fixed tie-breaking rule.

Definition. There are $A > 0$, $q \ge 2$ such that for all g and all x near $\mathcal{M} = \operatorname{argmin}_{\mathcal{X}} f$:

$$A\,\|g\|\,\|x - s(g)\|^q \le \langle g, x - s(g)\rangle$$



Interpretation. A small dual gap forces the LMO atom close to x.

Note. (α, q)-uniformly convex sets (e.g. ℓ_p-balls) satisfy LDS with $A = \alpha/2$; LDS is strictly more general.

[Def. 3.1, Prop. 3.4]

Main Result: Curvature Alone Beats $1/t$

Theorem. Lean-verified Let f be smooth and convex and let $\mathcal{X}$, with its fixed LMO rule, satisfy LDS around $\mathcal{M} = \operatorname{argmin}_{\mathcal{X}} f$. Then

$$f(x_t) - f^\star = o(1/t)$$

for short steps, exact line search, and open loop $2/(t+2)$.

[Cor. 3.8, Thm. 3.10]

No strong convexity, error bound, or gradient bound needed.

In particular. On uniformly convex sets: $o(1/t)$ for all smooth convex f.

Scope. Per instance, $t\,(f(x_t) - f^\star) \to 0$; not a uniform (minimax) rate.

Rates. Add a local Hölder error bound of order r: $f(x_t) - f^\star = O\big(t^{-rq/(rq-2)}\big)$, e.g. $O(1/t^2)$ for $q = r = 2$.

[Thm. 4.2]

Proof Idea: Reciprocal Progress

Short steps; the hard zero-gradient case.

$F_t = f(x_t) - f^\star$, $g_t = \langle \nabla f(x_t), x_t - s_t\rangle$, $d_t = \|x_t - s_t\|$, and $\delta_t = \|x_t - p_t\| \to 0$ for nearest minimizers p_t (with $\nabla f(p_t) = 0$).

1. Key estimate. Convexity at p_t plus LDS:

$$g_t \ge A\,\|\nabla f(x_t)\|\, d_t^q \ge A\,\frac{F_t}{\delta_t}\, d_t^q.$$

2. Reciprocal progress. Short-step descent and step 1 give, for $H_t = 1/F_t$ and some $c > 0$,

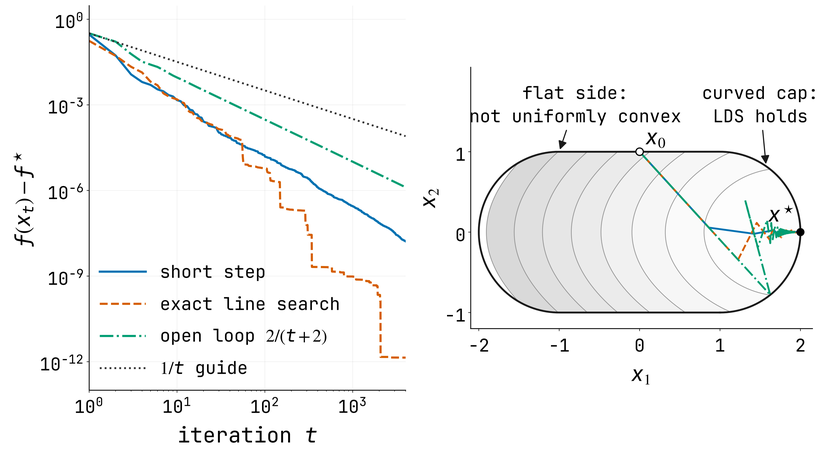
$$H_{t+1} - H_t \ge c\,\delta_t^{-2/q} \longrightarrow \infty.$$

3. Cesàro. $H_t/t \to \infty$, i.e. $t\,F_t \to 0$. □

[Thm. 3.5]

Example: The Stadium

$\mathcal{X} = [-1, 1] \times \{0\} + B_2(0, 1)$, $f(x) = \frac{1}{2}x_2^2 + \psi(2 - x_1)$, $\psi(u) = u - \arctan u$; minimizer $x^\star = (2, 0)$ on the cap, $\nabla f(x^\star) = 0$.



All rules fall below the $1/t$ guide (finite horizon; illustrative).

[Sec. 5, Fig. 1]

Lean-verified. Every theorem from the paper shown here is kernel-checked in Lean 4, in real Hilbert spaces and for every LMO tie-breaking rule (formalization: S. Pokutta & C. Spiegel).

Supported by MATH$^+$ (EXC-2046/2, project ID 390685689), project EF-LiOpt-3 *Agentic AI for Mathematics*.

Lean proofs
lemmata.iol.berlin
html/2611.00012

Paper
arXiv:2604.28006