Projection-free first-order methods for nonsmooth optimization

Antonio SILVETI-FALLS antonio.silveti-falls@centralesupelec.fr
Fédération de Mathématiques de CentraleSupélec

Keywords

Constrained optimization, nonsmooth optimization, stochastic optimization, Frank-Wolfe, Conditional Gradient, machine learning, data science.

Context

Projection-free first-order optimization methods, such as the Frank-Wolfe algorithm [4] and conditional gradient methods [7], have proven to be useful for many machine learning and data science problems [1] due to their ability to handle complex constraint sets without requiring possibly expensive projection operations. These methods rely on solving a linear minimization subproblem over the feasible domain at each iteration, making them attractive for large-scale optimization problems [3] when this operation is cheaper than projections. However, their analysis was traditionally relegated to smooth objective functions.

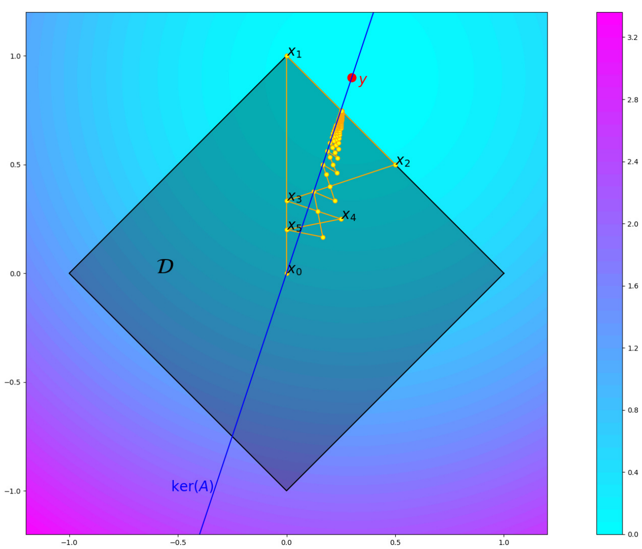

Figure 1: A nonsmooth variant of the Frank-Wolfe algorithm for solving a toy problem: project onto the intersection of an ℓ^1 ball and a linear subspace.

$$\min_{x\in\mathbb{R}^2:\ Ax=0,\|x\|_1\leq 1} \frac{1}{2}\|x-y\|_2^2$$

State of the Art

Recent works in this area have focused on minimizing objectives of the form $f+g$ over a convex, compact constraint set $\mathcal{C}$, where f is $C^{1,1}$ smooth (continuously differentiable with Lipschitz-continuous gradient) and convex, and g is nonsmooth but convex, proper, and lower semicontinuous [5, 10, 11, 12, 13]. This problem structure arises in various applications, including sparse recovery, matrix completion, and more. We have also come up with some preliminary results indicating that a Frank-Wolfe approach is also capable of tackling $f+g$ when f is nonconvex or g is nonsmooth and weakly-convex (such as the SCAD regularizer, the MCP regularizer, and so on), which we plan to expand on in this project.

Novelty

Building upon our preliminary results, this post-doctoral research project aims to push the boundaries of nonsmooth projection-free optimization by exploring several innovative directions:

1. **Adaptive step sizes and smoothing**: We will investigate strategies to adapt the step size schedule based on the local geometry of the problem, utilizing a local curvature estimate. The goal in this direction is to accelerate convergence in practice compared to the short-step or open-loop step sizes that are used currently.

We plan to investigate the smoothing schedule for the nonsmooth function g and its effect on convergence as well, since it is known from our previous work [10] that an augmented Lagrangian approach is more stable than simple smoothing and can be interpreted as adaptive smoothing. Although convergence is guaranteed under abstract summability assumptions on the parameters [10], the effects of different choices of initial values or sequences is not yet understood and no explorations of adaptivity in this context exist yet.

2. **Accelerated algorithms**: We will explore accelerated variants related to the Conditional Gradient Sliding [6] and Boosted Frank-Wolfe [2] approaches, which utilize more than one call to the linear minimization oracle per iteration. The aim will be to prove convergence rates and investigate their performance numerically. For sets with a cheap linear minimization oracle, advances in this direction will yield algorithms that are faster than vanilla Frank-Wolfe.

3. **Extensions to stochastic problems**: Although we don't expect the adaptive step size studies to extend to the stochastic setting, we do expect the other developments to lead to novel algorithms in stochastic contexts as well, as has been done for instance in [8, 9, 11].

Objectives

The primary objective of this post-doctoral research project is to find and analyze new variants of projection-free optimization methods for nonsmooth problems and then implement them towards some applications in data science or machine learning. In terms of theory, we aim to develop novel algorithms with rigorous convergence guarantees in the form of worst-case convergence rates that match or improve upon the current state-of-the-art theoretical convergence rates. Towards this, we will examine the theoretical relationship between the Frank-Wolfe gap at $x \in \mathcal{C}$,

$$\min_{g' \in \partial g(x)} \max_{u \in \mathcal{C}} \langle g', x - u \rangle$$

a notion of optimality used in projection-free optimization, and the *smoothed* Frank-Wolfe gap at $x \in \mathcal{C}$,

$$\max_{u \in \mathcal{C}} \langle \nabla g^{\beta}(x), x - u \rangle$$

which is Frank-Wolfe gap evaluated with the gradient of a smoothed surrogate objective function $\nabla g^{\beta}(x)$. Some recent works studying Frank-Wolfe on nonsmooth functions take for granted that these gaps are different, and typically focus on decreasing the smoothed gap. We have preliminary evidence showing this reasoning is flawed, and that the gap-between-gaps must also be taken into account to get the best convergence for the original, nonsmooth problem.

On a practical level, we expect to improve on existing techniques by leveraging adaptive step sizes, smoothing schedules, and multiple calls to the linear minimization oracle at each iteration, in combination with our theoretical insights linking the optimality between the nonsmooth and smoothed problems. To demonstrate our theoretical claims in practice we will be implementing the proposed algorithms as efficient, open-source software packages in python or julia. This will facilitate their adoption by the broader research community and enable their application to problems in various domains such as machine learning, imaging, and deep learning[8, 9].

Desired profile

The desired candidate should have experience with theoretical analysis of optimization algorithms, in particular first-order methods and, if possible, projection-free methods. Experience and familiarity with vectorized Python (or possibly julia) programming, in particular with the common deep learning libraries (PyTorch, JAX, or TensorFlow) or at least with NumPy, will be necessary.

References

[1] Gábor Braun, Alejandro Carderera, Cyrille W Combettes, Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Sebastian Pokutta. Conditional gradient methods. *arXiv preprint arXiv:2211.14103*, 2022.

[2] Cyrille Combettes and Sebastian Pokutta. Boosting frank-wolfe by chasing gradients. In *International Conference on Machine Learning*, pages 2111–2121. PMLR, 2020.

[3] Cyrille W Combettes and Sebastian Pokutta. Complexity of linear minimization and projection on some sets. *Operations Research Letters*, 49(4):565–571, 2021.

[4] Marguerite Frank, Philip Wolfe, et al. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[5] Gauthier Gidel, Fabian Pedregosa, and Simon Lacoste-Julien. Frank-wolfe splitting via augmented lagrangian method. In *International Conference on Artificial Intelligence and Statistics*, pages 1456–1465. PMLR, 2018.

[6] Guanghui Lan and Yi Zhou. Conditional gradient sliding for convex optimization. *SIAM Journal on Optimization*, 26(2):1379–1409, 2016.

[7] Evgeny S Levitin and Boris T Polyak. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5):1–50, 1966.

[8] Thomas Pethick, Wanyun Xie, Kimon Antonakopoulos, Zhenyu Zhu, Antonio Silveti-Falls, and Volkan Cevher. Training deep learning models with norm-constrained lmos. *arXiv preprint arXiv:2502.07529*, 2025.

[9] Thomas Pethick, Wanyun Xie, Mete Erdogan, Kimon Antonakopoulos, Tony Silveti-Falls, and Volkan Cevher. Generalized gradient norm clipping & non-euclidean (l_0, l_1)-smoothness. *arXiv preprint arXiv:2506.01913*, 2025.

[10] Antonio Silveti-Falls, Cesare Molinari, and Jalal Fadili. Generalized conditional gradient with augmented lagrangian for composite minimization. *SIAM Journal on Optimization*, 30(4):2687–2725, 2020.

[11] Antonio Silveti-Falls, Cesare Molinari, and Jalal Fadili. Inexact and stochastic generalized conditional gradient with augmented lagrangian and proximal step. *Journal of Nonsmooth Analysis and Optimization*, 2(Original research articles), 2021.

[12] Alp Yurtsever, Olivier Fercoq, Francesco Locatello, and Volkan Cevher. A conditional gradient framework for composite convex minimization with applications to semidefinite programming. In *International Conference on Machine Learning*, pages 5727–5736. PMLR, 2018.

[13] Alp Yurtsever, Suvrit Sra, and Volkan Cevher. Conditional gradient methods via stochastic path-integrated differential estimator. In *International Conference on Machine Learning*, pages 7282–7291. PMLR, 2019.