Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

THE DERIVATIVE-FREE FULLY-CORRECTIVE FRANK-WOLFE ALGORITHM FOR OPTIMIZING FUNCTIONALS OVER PROBABILITY SPACES

Di Yu[1]

[1]Dept. of Statistics, Purdue University, West Lafayette, IN, USA

ABSTRACT

The challenge of optimizing a smooth convex functional over probability spaces is highly relevant in experimental design, emergency response, variations of the problem of moments, etc. A viable and provably efficient solver is the fully-corrective Frank-Wolfe (FCFW) algorithm. We propose an FCFW recursion that rigorously handles the zero-order setting, where the derivative of the objective is known to exist, but only the objective is observable. Central to our proposal is an estimator for the objective's *influence function*, which gives, roughly speaking, the directional derivative of the objective function in the direction of point mass probability distributions, constructed via a combination of Monte Carlo, and a projection onto the orthonormal expansion of an L_2 function on a compact set. A bias-variance analysis of the influence function estimator guides step size and Monte Carlo sample size choice, and helps characterize the recursive rate behavior on smooth non-convex problems.

1 INTRODUCTION

The challenge of optimizing a smooth functional on the space of compactly supported probability measures is stated as follows:

$$\begin{aligned} \min. \quad & J(\mu) \\ \text{s.t.} \quad & \mu \in \mathscr{P}(\mathscr{X}), \end{aligned} \tag{P}$$

where $\mathscr{X} \subset \mathbb{R}^d$ is compact, $\mathscr{P}(\mathscr{X})$ is the space of probability measures supported on $\mathscr{X}$, and $J : \mathscr{P}(\mathscr{X}) \to \mathbb{R}$ is a smooth functional. The problem (P) has received considerable attention due to its applicability in many contexts. Of particular interest in this paper is the frequently encountered *zero-order* (derivative-free) setting. While $J(\mu)$ is observable at any μ, its derivative analogue—understood here as the von Mises derivative J'_μ with influence function h_μ—is typically unavailable.

Problem (P) arises in many contexts. For instance, the P-means problem (Molchanov and Zuyev 2002), often described as a randomized version of the k-means clustering problem, can be formulated in this framework. In the field of experimental design, regression models provide another example: one seeks a randomized design (sampling distribution) μ that minimizes criteria such as the trace (A-optimality) or determinant (D-optimality) of the covariance of least-squares estimators. When derivatives J'_μ are accessible, first-order methods such as the fully corrective Frank–Wolfe (FCFW) algorithm (Yu et al. 2024) can be applied. In contrast, when derivatives are unavailable or expensive to approximate, a zeroth-order analogue is required. This paper develops such a method based on estimating the influence function h_μ.

Our main contribution is a derivative-free variant of FCFW for optimization over probability measures. The approach replaces the unavailable von Mises derivative with an estimated influence function $\hat{h}_\mu$, obtained via a truncated L_2 expansion with Monte Carlo sampling and finite differencing. We provide a bias–variance decomposition (truncation bias, finite-difference bias, sampling variance) and establish an almost-sufficient decrease inequality together with a consistency result. This framework enables practical derivative-free optimization over measures without requiring prior discretization of the domain.

2 METHOD AND RESULTS

Frank-Wolfe in measure space. In Euclidean space $\mathbb{R}^d$, the Frank–Wolfe (FW) method (Bubeck 2015) minimizes a smooth convex function f over a compact convex set Z via updates

$$y_{k+1} = (1-\eta_k)y_k + \eta_k s_k, \quad s_k := \arg\min_{s\in Z} \nabla f(y_k)^\top s.$$

To extend FW to probability measures, we use the *influence function* of J at $\mu \in \mathscr{P}(\mathscr{X})$,

$$h_\mu(x) := \lim_{t\to 0^+} \frac{1}{t}\Big(J((1-t)\mu + t\delta_x) - J(\mu)\Big), \quad x \in \mathscr{X}. \tag{1}$$

and the linear functional *von Mises derivative* can be written as $J'_\mu(\nu-\mu) = \int h_\mu(x)\, d(\nu-\mu)(x)$. The FW update in measure space (Yu et al. 2024) then becomes

$$\mu_{k+1} = (1-\eta_k)\mu_k + \eta_k \delta_{x^*(\mu_k)}, \quad x^*(\mu_k) \in \arg\min_{x\in\mathscr{X}} h_{\mu_k}(x), \tag{2}$$

which iteratively adds atoms to form a sparse discrete measure. The FCFW variant then re-optimizes the weights over all previously selected atoms, yielding improved practical performance.

Derivative-free variant. Since h_μ is unobservable, we approximate it using a truncated orthonormal expansion. Assuming $h_\mu \in L_2(\mathscr{X})$, write $h_\mu(x) \approx \sum_{j=1}^d a_j u_j(x)$, where $\{u_j\}$ is an orthonormal basis and the coefficients $a_j = \langle h_\mu, u_j\rangle$ are estimated by Monte Carlo. Because $h_\mu(X)$ cannot be directly observed, we use a finite-difference approximation $FD_{s,\mu}(X) = \frac{1}{s}\{J((1-s)\mu + s\delta_X) - J(\mu)\}$ for $X \sim \mathrm{Unif}(\mathscr{X})$. Combining these yields the practical estimator

$$\hat{h}_\mu(x) = \sum_{j=1}^d \hat{a}_j(m,s)\, u_j(x), \quad \hat{a}_j(m,s) = \tfrac{v}{m}\sum_{t=1}^m FD_{s,\mu}(X_t)\, u_j(X_t), \quad X_t \sim \mathrm{Unif}(\mathscr{X}), \tag{3}$$

with parameters $p = (m,s,d)$ controlling sample size, step size, and truncation dimension. Substituting $\hat{h}_\mu$ for h_μ in the FCFW update gives a derivative-free variant (DF-FCFW).

The proposed estimator $\hat{h}_\mu$ admits an explicit bias–variance decomposition: truncation bias from d, finite-difference bias from s, and sampling variance from m. Under mild smoothness and decay assumptions, we obtain finite-sample bounds on $\mathbb{E}[\|\hat{h}_\mu - h_\mu\|_\infty]$ and $\mathbb{E}[\|\hat{h}_\mu - h_\mu\|_\infty^2]$, which guarantee that the estimator converges to h_μ in expectation as $m\to\infty$, $s\to 0$, and $d\to\infty$.

Substituting $\hat{h}_\mu$ into FCFW yields a derivative-free algorithm. We establish an "almost sufficient decrease" inequality that links progress in objective value to the quality of the estimator, and prove a consistency theorem: as long as parameters (m_k, s_k, d_k) are chosen appropriately across iterations, the influence values at selected atoms converge to zero almost surely, aligning with the optimality condition.

Numerical Validation. We validated the DF-FCFW on the Gaussian deconvolution problem $Y_i = W_i + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0,\sigma^2 I)$ and $W_i \sim \mu$. For the discrete case $\mu_a = \frac{1}{3}\delta_{-1} + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_{10}$, where the theoretical optimum and optimality conditions are known, DF-FCFW recovered the distribution and satisfied the optimality test via the influence function ($\min_x h_{\hat{\mu}}(x) \geq 0$). For the continuous case $\mu_b = \mathcal{N}(0, I_d)$ with $d = 10$, DF-FCFW showed steady decrease in objective value, moderate atom growth, and convergence of influence values toward zero, demonstrating both performance and scalability. These results confirm that the proposed estimator enables practical optimization over measures without requiring explicit derivatives.

REFERENCES

Bubeck, S. 2015. "Convex Optimization: Algorithms and Complexity". *Foundations and Trends in Machine Learning* 8(3–4):231–358 https://doi.org/10.1561/2200000050.

Molchanov, I., and S. Zuyev. 2002. "Steepest Descent Algorithms in a Space of Measures". *Statistics and Computing* 12(2):115–123 https://doi.org/10.1023/A:1014878317736.

Yu, D., S. G. Henderson, and R. Pasupathy. 2024. "Deterministic and Stochastic Frank-Wolfe Recursion on Probability Spaces". *arXiv preprint arXiv:2407.00307*.