PRINCETON UNIVERSITY

Massachusetts Institute of Technology

Riemannian Frank-Wolfe Methods

Melanie Weber
Princeton University

Suvrit Sra
MIT (LIDS)

Contact

mw25@math.princeton.edu
https://web.math.princeton.edu/~mw25/

Abstract

We study projection-free methods for constrained geodesically convex optimization on Riemannian manifolds:

$$\min_{z \in \mathcal{X} \subseteq \mathcal{M}} \phi(z)$$

In particular, we propose a Riemannian version of the Frank-Wolfe (RFW) method.

Contributions:

(1) In the geodesically constraint setting, we show global, non-asymptotic sublinear convergence. We also present a setting under which RFW can attain a linear rate.

(2) We further introduce a stochastic RFW for nonconvex optimization. In addition we consider two variance-reduced approaches for finite-sum settings.

(3) We then specialize RFW to the PSD manifold and show, that in this case, the log-linear oracle can be solved in closed form. In particular, we apply RFW to the computation of the Kacher mean and Wasserstein barycenters.

Riemannian Frank-Wolfe

Algorithm

Alg. 1 defines a Riemannian Frank-Wolfe method, where the update direction is computed by a log-linear oracle

$$\min_{z \in \mathcal{X}} \langle \text{grad } (x_k), \ \text{Exp}_{x_k}^{-1}(z) \rangle \ ;$$

a geodesic map implements the update. We show the following sublinear convergence guarantee:

Theorem (Convergence RFW)
Let $s_k = 2/(2+k)$. Then $\phi(x_k) - \phi(x^*) = O(1/k)$.

Algorithm 2 Riemannian Frank-Wolfe (RFW) for g-convex optimization

1: Initialize $x_0 \in \mathcal{X} \subseteq \mathcal{M}$; assume access to the geodesic map $\gamma : [0,1] \to \mathcal{M}$
2: **for** $k = 0, 1, \ldots$ **do**
3: $z_k \leftarrow \text{argmin}_{z \in \mathcal{X}} \langle \text{grad}\, \phi(x_k), \text{Exp}_{x_k}^{-1}(z) \rangle$
4: Let $s_k \leftarrow \frac{2}{k+2}$
5: $x_{k+1} \leftarrow \gamma(s_k)$, where $\gamma(0) = x_k$ and $\gamma(1) = z_k$
6: **end for**

For μ-strong g-convex objectives, we can establish linear convergence rates:

Theorem (Linear convergence RFW)
Let $s_k = r\sqrt{\mu \Delta_k}/(\sqrt{2} M_\phi)$ (M_ϕ: curvature constant and $\Delta_k = \phi(x_k) - \phi(x^*)$), then:

$$\Delta_{k+1} \leq \left(1 - \frac{r^2 \mu}{4 M_\phi}\right) \Delta_k \quad .$$

Application to the PSD manifold

On the manifold of positive definite matrices with interval constraints, i.e.,

$$\min_{X \subseteq \mathcal{X} \subseteq \mathbb{P}_d} \phi(X) \ , \quad \mathcal{X} = \{X \in \mathbb{P}_d \mid L \preceq X \preceq U\}$$

the log-linear oracle is solvable in closed form.

Theorem (Linear convergence RFW)
Let $L, U \in \mathbb{P}_d, \ L \preceq U$. Then for arbitrary $S \in \mathbb{H}_d, \ X \in \mathbb{P}_d$ there exists a closed form solution to the oracle

$$\max_{L \preceq Z \preceq U} \text{tr}\left(S \log(XZX)\right) \ .$$

Applications

(1) ***Karcher mean:*** Center of mass w.r.t. the Riemannian distance

$$\text{argmin}_{X \in \mathbb{P}_d} \sum_{i=1}^{n} \| \log \left(X^{-1/2} A_i X^{-1/2} \right) \|_F^2 \ .$$

(2) ***Wasserstein barycenters:*** Solution to multimarginal transport problem

$$\text{argmin}_{X \in \mathbb{P}_d} \sum_{i=1}^{n} \left[\text{tr}(X + A_i) - 2\text{tr}(X^{1/2} A_i X^{1/2})^{1/2} \right]^{1/2} .$$

Extension to stochastic and non-convex settings

Stochastic Optimization on Manifolds

Consider the following constraint stochastic and finite-sum problems:

$$(1) \quad \Phi(X) = \mathbb{E}_\xi \left[\phi(X, \xi)\right] = \int \phi(X, \xi) dP(\xi)$$

$$(2) \quad \Phi(X) = \frac{1}{n} \sum_{i=1}^{n} \phi_i(X)$$

Here, $\Phi, \ \{\phi_i\}$ are smooth, but may be ***non-convex***.

Advantages

(1) Compute stochastic gradients instead of full gradients: Improved oracle complexity.

(2) Not restricted to g-convex objectives.

Algorithm	RFW	SRFW	SVR-RFW	SPIDER-RFW
SFO/ IFO	$O(n/\epsilon^2)$	$O\left(\frac{1}{\epsilon^4}\right)$	$O\left(n + \frac{n^{2/3}}{\epsilon^2}\right)$	$O\left(\frac{1}{\epsilon^3}\right)$
LO	$O(1/\epsilon^2)$	$O\left(\frac{1}{\epsilon^2}\right)$	$O\left(\frac{1}{\epsilon^2}\right)$	$O\left(\frac{1}{\epsilon^2}\right)$

Algorithms

We introduce three stochastic RFW methods:

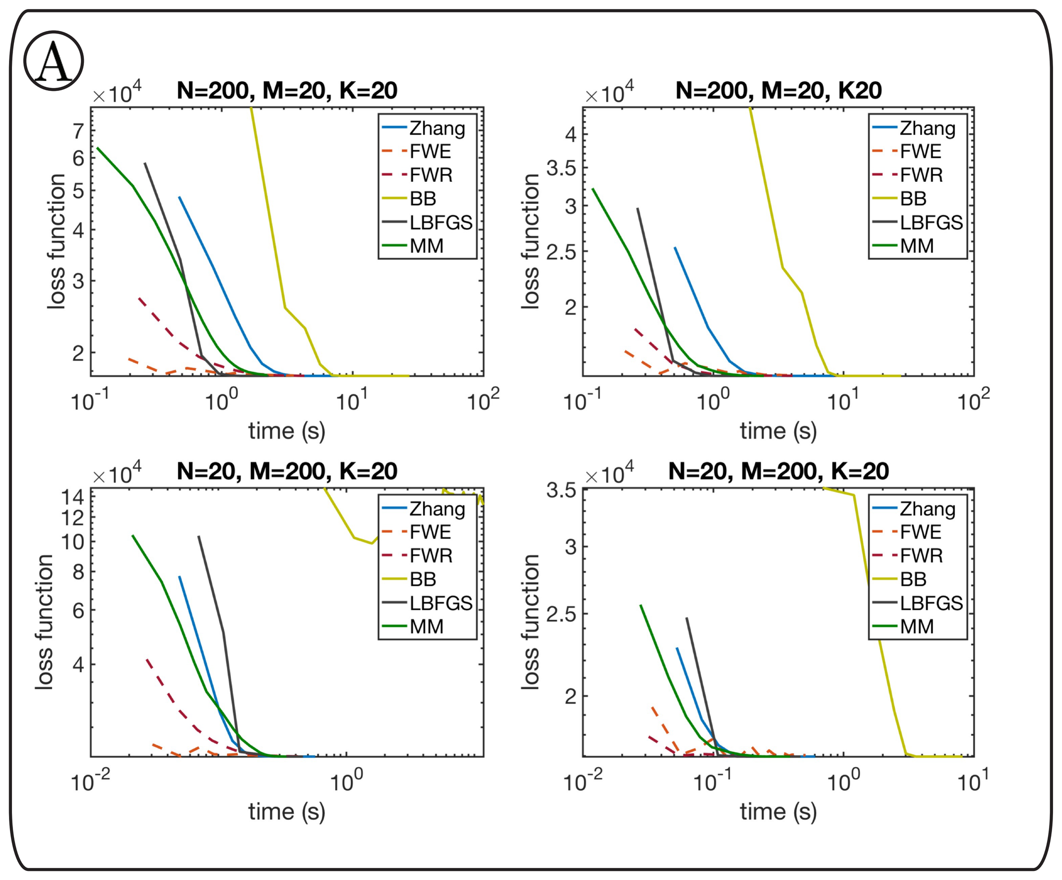
(a) a purely stochastic RFW (***SRFW***) for solving (1);

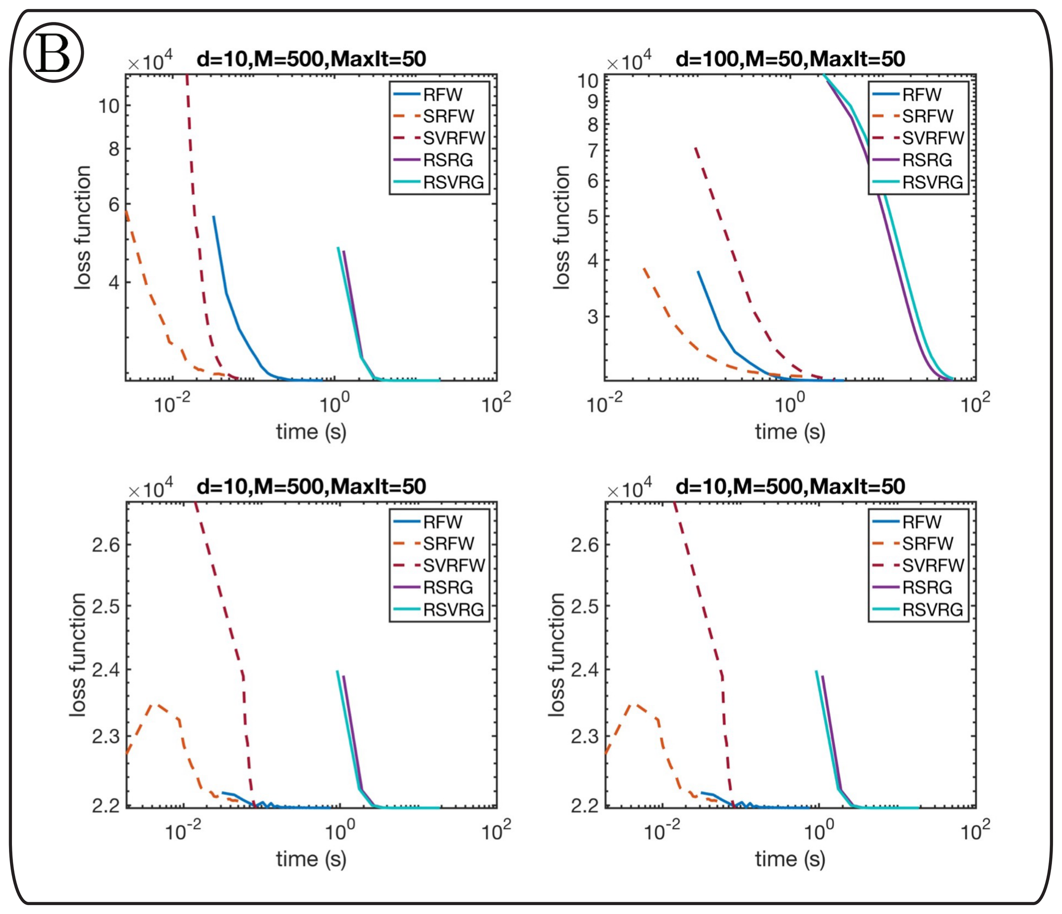
(b) a variance-reduced approach (***SVRFW***) for solving (2);

(c) an improved variance-reduced approach using SPIDER (***SPIDER-RFW***).

The performance of all four RFW methods is comparable to their Euclidean counterparts.

Experiments





Experiments (Karcher mean)

A: Performance of RFW in comparison to state-of-the-art methods for well-conditioned (left) and ill-conditioned (right) inputs.

B: Performance of stochastic RFW (SRFW and SVRFW) against RFW and state-of-the-art stochastic methods for well-conditioned (left) and ill-conditioned (right) inputs.