Riemannian Frank-Wolfe with Application to the Geometric Matrix Mean

PRINCETON UNIVERSITY
Massachusetts Institute of Technology

Melanie Weber, Princeton University
Suvrit Sra, Laboratory for Information and Decision Systems, MIT

arXiv: 1710.10770
mw25@math.princeton.edu
https://web.math.princeton.edu/~mw25

Abstract

We consider constrained optimization of geodesically convex objectives over geodesically convex subsets of Riemannian manifolds. We address these constraints via a Riemannian Frank-Wolfe (RFW) approach, that offers to be promising due to its "projection free" nature.

First, we prove an abstract convergence result for RFW; then we specialize it to the manifold of positive definite matrices, for the specific task of computing the Riemannian centroid (also known as Karcher mean). This specialization relies crucially on a "log-linear oracle," a subroutine key to implementing RFW – remarkably, this oracle is seen to admit a closed form solution, which may be of independent interest.

We discuss two other variations, including a non-convex Euclidean Frank-Wolfe (EFW) method, as well a setting under which RFW attains a linear rate of convergence. Experiments against recently published methods for the Riemannian centroid highlight the competitiveness of RFW.

Background

In this work, we consider the optimization of a geodesically (g-)convex function on a Riemannian manifold $\mathcal{M}$ over a compact, g-convex set $\mathcal{X}$.

If $\mathcal{X}$ is simple, Riemannian projected-gradient methods offer a practical solution. However, in several settings the computation of these (metric) projections onto the constraint set is computationally expensive - driving the search for projection-free methods.

In the Euclidean case ($\mathcal{M} = \mathbb{R}^n$), Frank-Wolfe (FW) schemes provide a projection-free approach: Instead of projection, FW relies on a "linear" oracle that maximizes a conditional gradient - which can often be much simpler.

FW methods have been applied to a variety of (Euclidean) optimization problems, including convex, non-convex, submodular and stochastic settings. However, they have not been studied in the manifold case - a gap in the literature, that we attempt to fill.

Algorithm 1 Euclidean Frank-Wolfe without line-search

Initialize with a feasible point $x_0 \in \mathcal{X} \subset \mathbb{R}^n$
for $k = 0, 1, \dots$ **do**
  Compute $z_k \leftarrow \text{argmin}_{z \in \mathcal{X}} \langle \nabla\phi(x_k),\, z - x_k \rangle$
  Let $s_k \leftarrow \frac{2}{k+2}$
  Update $x_{k+1} \leftarrow (1 - s_k)x_k + s_k z_k$
end for

M. Frank & P. Wolfe (1956)

Geometric Optimization

Consider
$$\min_{x \in \chi \subseteq \mathcal{M}} \phi(x)$$
where $\phi : \mathcal{M} \to \mathbb{R}$ is differentiable and g-convex. The g-convexity guarantees the following optimality result:

Proposition (Optimality)
Let $x^* \in \chi \subset \mathcal{M}$ be a local optimum. Then, x^* is globally optimal, and for all $y \in \chi$
$$\langle \text{grad}\, \phi(x^*), \text{Exp}_y^{-1}(x^*) \rangle \geq 0 \,.$$

Using this argument, we can replace the linear oracle in the Euclidean FW by the following log-linear oracle:
$$\min_{z \in \chi} \langle \text{grad}\, \phi(x_k), \text{Exp}_{x_k}^{-1}(z) \rangle \,,$$
where χ is compact and g-convex. This allows for a Riemannian FW scheme. One can show the following global convergence result for RFW:

Theorem (Convergence of RFW)
Let $s_k = \frac{2}{k+2}$ be the stepsize of RFW. Then,
$$\phi(x_k) - \phi(x^*) = O(1/k) \,.$$

By introducing a notion of optimal transport to the RFW scheme and requiring μ-strong g-convexity, we can establish linear convergence rates by applying the PL inequality. In particular, we can prove the following linear result:

Theorem (Linear Convergence of RFW)
Let $s_k = \frac{r\sqrt{\mu \Delta_k}}{\sqrt{2M_\phi}}$ (M_ϕ: curvature constant). Then, RFW converges linearly at
$$\Delta_{k+1} \leq \left(1 - \frac{r^2\mu}{4M_\phi}\right)\Delta_k \,.$$

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

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

Geometric Matrix Mean

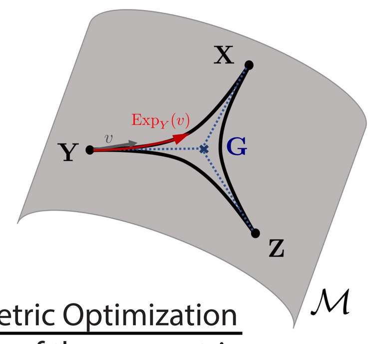


Fig. 1: Geometric Optimization
Computation of the geometric matrix mean as the minimization problem of finding the center of mass on a Riemannian manifold.

We demonstrate RFW on a simple, but important class of constraint optimization problems on the Riemannian manifold of Hermitian positive definite (HPD) matrices , i. e. $\mathcal{M} = \mathbb{P}_d$. Consider
$$\min_{X \in \chi \subseteq \mathbb{P}_d} \phi(X), \quad \text{where } \chi = \{X \in \mathbb{P}_d | L \preceq X \preceq U\}$$
where ϕ and the "HPD-interval" χ are g-convex.

To solve this problem, we implement RFW with the following adapted log-linear oracle:
$$\min_{L \preceq Z \preceq U} \langle X_k^{1/2} \nabla^{\mathbb{H}} \phi(X_k)\, X_k^{1/2}, \log(X_k^{-1/2} Z X_k^{-1/2}) \rangle \,.$$
For this, we can derive a closed-form solution:

Theorem (log-linear oracle)
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
$$\max_{L \preceq Z \preceq U} \text{tr}(S \log(XZX)) \,.$$

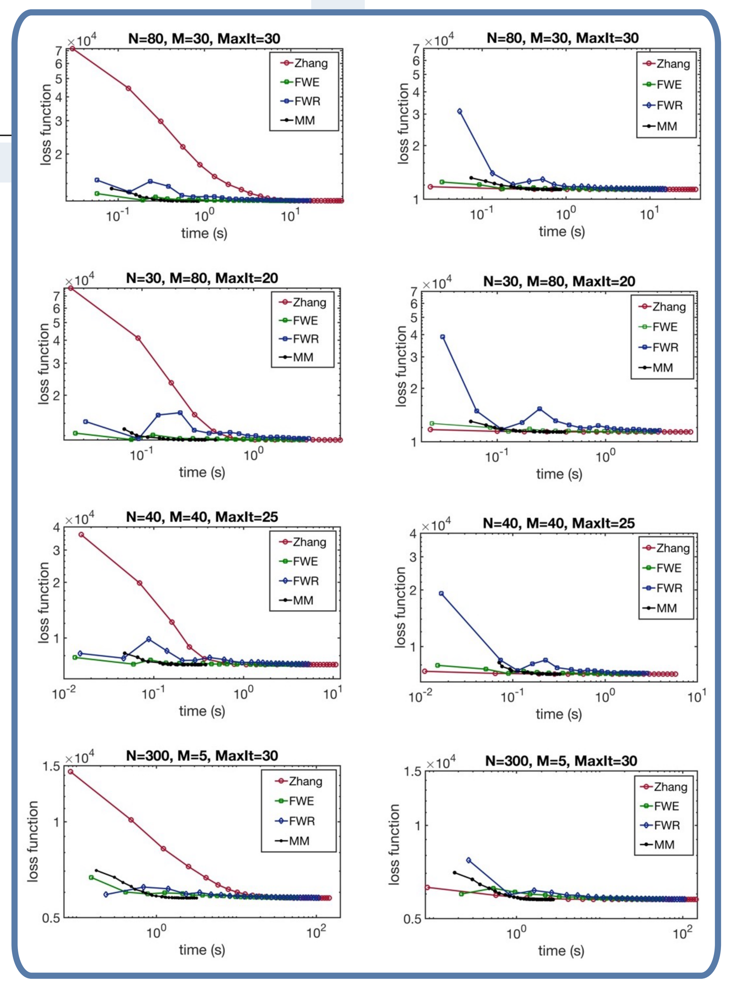
The constraint set χ is not only g-convex, but also convex in the usual sense. Therefore, we can also apply a (non-convex) Euclidean Frank-Wolfe scheme (EFW), albeit with a slower convergence rate. We can derive an analogous closed-form solution for the Euclidean linear oracle.

An implementation of both RFW and EFW is given below.

Algorithm 3 Frank-Wolfe for fast geometric mean

$(A_1, \dots, A_N)$, $\boldsymbol{w} \in \mathbb{R}_+^N$
$\bar{X} \approx \text{argmin}_{X>0} \sum_i w_i \delta_R^2(X, A_i)$
for $k = 0, 1, \dots$ **do**
  $\nabla\phi(X_k) = X_k^{-1}(\sum_i w_i \log(X_k A_i^{-1}))$
  Compute Z_k using (i) for EFW and (ii) for RFW:
  (i) $Z_k \leftarrow \text{argmin}_{H \leq Z \leq A} \langle \nabla\phi(X_k),\, Z - X_k \rangle$
  (ii) $Z_k \leftarrow \text{argmin}_{H \leq Z \leq A} \langle X_k^{1/2} \nabla\phi(X_k) X_k^{1/2},\, \log(X_k^{-1/2} Z X_k^{-1/2}) \rangle$
  Let $\alpha_k \leftarrow \frac{2}{k+2}$
  Update X using (i) for EFW and (ii) for RFW:
  (i) $X_{k+1} \leftarrow X_k + \alpha_k(Z_k - X_k)$
  (ii) $X_{k+1} \leftarrow X_k \#_{\alpha_k} Z_k$
end for
return $\bar{X} = X_k$

Fig. 2: Performance of RFW and EFW
Performance of RFW/ EFW in comparison to state-of-the-art methods by T. Zhang [SIAM, 2017] and the Matrix Mean Toolbox (MM) by D. A. Bini and B. Iannazzo [Lin. Alg. Appl., 2014]. Here, N is the size of the matrix, M the number of matrices and K the number of iterations. The left column shows results for initial guess $X_0 = H$, the right column for $X_0 = \frac{1}{2}(H + A)$, where H is the harmonic mean and A the arithmetic.



Contributions

- We introduce Riemannian Frank-Wolfe (RFW) for constrained g-convex optimization on Riemannian manifolds. Analogous to the Euclidean case, we show that RFW attains a non-asymptotic sublinear rate of convergence. Furthermore, under additional assumptions on the objective function and the constraint set, we show that RFW can attain linear convergence rates.
- We specialize RFW for g-convex problems on the manifold of Hermitian positive definite (HPD) matrices, for which we present a closed-form solution to the required RFW "linear" oracle. In addition, we discuss a direct non-convex Euclidean FW (EFW) method (which converges to the global optimum due to g-convexity); here the linear oracle involves a semi-definite program (SDP), which is shown to have a closed-form solution.
- We apply RFW and EFW to the computation of the geometric matrix mean of HPD matrices. We believe the closed-form solutions for the "log-linear" and linear oracles for RFW and EFW, respectively should be of wider interest too. These oracles lie at the heart of why RFW and EFW result in outperforming state-of-the-art methods for computing the Riemannian mean of HPD matrices.