Generalized Conditional Gradient Method with Three Step Size Strategies on Riemannian Manifolds

Member Kyoto University *CHEN Kangming
Member Kyoto University FUKUDA Ellen Hidemi

1. Introduction

The Conditional Gradient (CG) Method, also known as the Frank-Wolfe algorithm, is a fundamental technique in constrained optimization, developed by Marguerite Frank and Philip Wolfe in 1956. It is particularly useful for minimizing convex functions over complex convex constraint sets where projection is computationally challenging. Recent explorations of CG method on Riemannian manifolds [1] and the Generalized Conditional Gradient (GCG) method have been detailed in various studies [2]. Notably, the GCG method's application to composite multi-objective optimization on Riemannian manifolds has also been discussed, but the critical aspect of solving subproblems remains underexplored. This work investigates the GCG method on Riemannian manifolds, focusing on three types of step size and the implementation of subproblem solutions.

2. The generalized conditional gradient method on Riemannian manifolds

We consider the following unconstrained vector optimization problem:

$$\begin{array}{ll} \min & F(x) := f(x) + g(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{1}$$

where $\mathcal{X} \subseteq \mathcal{M}$ is a compact and geodesically convex set and $F\colon \mathcal{M} \to \mathbb{R}$ is a composite function where $f\colon \mathcal{M} \to \mathbb{R}$ is continuously differentiable and $g\colon \mathcal{M} \to \mathbb{R}$ is a proper, closed, geodesically convex and lower semicontinuous (possibly nonsmooth) function with compact domain.

A Riemannian manifold $\mathcal{M}$ is a manifold endowed with a Riemannian metric $(\eta_x, \sigma_x) \mapsto \langle \eta_x, \sigma_x \rangle_x \in \mathbb{R}$, where η_x and σ_x are tangent vectors in the tangent space of $\mathcal{M}$ at x. The tangent space of a manifold $\mathcal{M}$ at $x \in \mathcal{M}$ is denoted as $T_x\mathcal{M}$, and the tangent bundle of $\mathcal{M}$ is denoted as $T\mathcal{M} := \{(x, d) \mid d \in T_x\mathcal{M}, x \in \mathcal{M}\}$. The norm of $\eta \in T_x\mathcal{M}$ is defined as $\|\eta\|_x := \sqrt{\langle \eta, \eta \rangle_x}$. For a map $F : \mathcal{M} \to \mathcal{N}$ between two manifolds $\mathcal{M}$ and $\mathcal{N}$, $\mathrm{D}F(x) : T_x\mathcal{M} \to T_{F(x)}\mathcal{N}$ denotes the derivative of F at $x \in \mathcal{M}$. The Riemannian gradient $\mathrm{grad} f(x)$ of a function $f : \mathcal{M} \to \mathbb{R}$ at $x \in \mathcal{M}$ is defined as a unique tangent vector at x satisfying $\langle \mathrm{grad} f(x), \eta \rangle_x = \mathrm{D}f(x)[\eta]$ for any $\eta \in T_x\mathcal{M}$.

In Euclidean space, the CG method method aims to minimize a convex function $f(x)$ over a convex set $\mathcal{X}$. At each iteration k, the method solves a linear approximation of the original problem by finding a direction s^k that minimizes the linearized objective function over the constraint set $\mathcal{X}$:

$$s^k = \arg\min_{s \in \mathcal{X}} \langle \nabla f(x^k), s \rangle.$$

Then, it updates the current point x^k using a step size λ_k along the direction $s^k - x^k$:

$$x^{k+1} = x^k + \lambda_k (s^k - x^k).$$

For composite functions of the form $f(x) = h(x) + g(x)$, where h is smooth and g is convex, the CG method can be adapted to handle the non-smooth term g. The update step becomes:

$$x^{k+1} = \arg\min_{s \in \mathcal{X}} \langle \nabla h(x^k), s \rangle + g(s).$$

The step size λ_k can be chosen using various strategies. The CG method is particularly useful when the constraint set $\mathcal{X}$ is complex, making projection-based methods computationally expensive.

Compared to the update iterations in Euclidean space, the process in the Riemannian setting is generalized as

$$x^{k+1} = R_{x^k}(\lambda_k d^k), \quad \text{for } k = 0, 1, 2, \ldots,$$

where d^k is a descent direction, λ_k is a step size, and R is a retraction that projects points from the tangent space of the manifold onto the manifold itself.

Then we can give our algorithm.

Algorithm 1 Riemannian GCG method

Step 0. Initialization:
Choose $x^0 \in \mathcal{X}$ and initialize $k=0$.
Step 1. Compute the search direction:
Compute an optimal solution $p(x^k)$ and the optimal value $\theta(x^k)$ as

$$p(x^k) = \arg\min_{u\in\mathcal{X}} \langle \mathrm{grad} f(x^k), R_{x^k}^{-1}(u)\rangle_{x^k} + g(u) - g(x^k),$$

$$\theta(x^k) = \langle \mathrm{grad} f(x^k), R_{x^k}^{-1}(u)\rangle_{x^k} + g(u) - g(x^k).$$

Define the search direction by $d(x^k) = R_{x^k}^{-1}(p(x^k))$.
Step 2. Compute the Step size:
Compute the step size λ_k.
Step 3. Update:
Update the current iterate

$$x^{k+1} = R_{x^k}(\lambda_k d(x^k)).$$

Step 4. Convergence check:
If a convergence criteria is met, stop; otherwise, set $k=k+1$ and return to Step 1.

In this work, we consider three types of step sizes.

Armijo step size:

Let $\zeta \in (0,1)$ and $0 < \omega_1 < \omega_2 < 1$. The step size λ_k is chosen according to the following line search algorithm:

Step 0 Set $\lambda_{k_0} = 1$ and initialize $\ell \leftarrow 0$.

Step 1 If

$$F(R_{x^k}(\lambda_{k_\ell} d(x^k))) \le F(x^k) + \zeta \lambda_{k_\ell} \theta(x^k),$$

then set $\lambda_k := \lambda_{k_\ell}$ and return to the main algorithm.

Step 2 Find $\lambda_{k_{\ell+1}} \in [\omega_1 \lambda_{k_\ell}, \omega_2 \lambda_{k_\ell}]$, set $\ell \leftarrow \ell+1$, and go to Step 1.

Adaptive step size:

$\lambda_k := \min\left\{1, -\theta(x^k)/(LD^2(p\left(x^k\right), x^k))\right\} = \mathrm{argmin}_{\lambda\in(0,1]} \left\{\lambda\theta(x^k) + \frac{L}{2}\lambda^2 D^2(p(x^k), x^k)\right\}$, where $D(x,y)$ is the geodesic distance between x and y and L is the smoothness constant of f.

Diminishing step size:

$$\lambda_k := 2/(k+2).$$

3. Convergence results

The convergence results are established as follows:

Theorem 1. *Let x^* be an optimal point of the problem and $\{x^k\}$ be generated by Algorithm 1 with adaptive or diminishing step size. Assume that f is L-smooth and g is geodesically convex. Then $\{x^k\}$ satisfies $F(x^k) - F\left(x^*\right) = O(1/k)$.*

Theorem 2. *Let $\{x^k\}$ be generated by Algorithm 1 with the Armijo step size. Assume that f is L-smooth and g is geodesically convex and lower semicontinuous. Then $\lim_{k\to\infty} \theta\left(x^k\right) = 0$ and every limit point $x^* \in \mathcal{X}$ of the sequence $\{x^k\}$ is a stationary point.*

4. Solving subproblem

The subproblem involves minimizing the function $\min_{u\in\mathcal{X}} \langle \mathrm{grad} f(x^k), R_{x^k}^{-1}(u)\rangle_{x^k} + g(u)$ or equivalently $\min_{\eta\in T_{x^k}\mathcal{X}} \langle \mathrm{grad} f(x^k), \eta\rangle_{x^k} + g(R_{x^k}(\eta))$. Non-convexity may arise from the inverse retraction $R_{x^k}^{-1}(u)$ and the composition with g. To address this, define the function

$$\ell_{x^k}(\eta) = \langle \mathrm{grad} f(x^k), \eta\rangle_{x^k} + g(R_{x^k}(\eta)).$$

Follow the idea in [3], a local model $\tilde{\ell}_{y_k}(\xi_k)$ is defined as:

$$\left\langle \mathcal{T}_{R_{d_k}}^{-\sharp}\left(\mathrm{grad}\, f(x)\right), \xi \right\rangle_{y_k} + g\left(y_k + \xi\right),$$

where d_k is the current estimate η_k. The optimization strategy involves finding a new estimate by minimizing this local model in the tangent space $\mathrm{T}_{y_k}\mathcal{X}$ to compute a search direction ξ_k^* and then updating d_k along this direction. This iterative approach refines the descent direction and moves toward the optimal solution.

5. Future works

Our future work will focus on developing and analyzing an accelerated version of the Riemannian GCG method. In addition, we plan to conduct numerical experiments to validate the effectiveness of our proposed methods.

Reference

[1] M. Weber, S. Sra. Riemannian optimization via Frank-Wolfe methods. *Mathematical Programming*, 2023, 199(1): 525-556.

[2] P. B. Assunção, O. P. Ferreira, L. F. Prudente. A generalized conditional gradient method for multiobjective composite optimization problems. *Optimization*, 2023: 1-31.

[3] W. Huang, K. Wei. Riemannian proximal gradient methods. *Mathematical Programming*, 2022, 194(1): 371-413.