Nonlinear conjugate gradient method for vector optimization on Riemannian manifolds

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

1. Introduction

Recently, many Riemannian conjugate gradient methods have been analyzed in [1]. Moreover, the nonlinear conjugate gradient method for vector optimization was first proposed in [2]. Here, we propose a nonlinear conjugate gradient method for vector optimization on Riemannian manifolds. We notice that a similar work, that uses general retractions and vector transports but for multiobjective problems, was also done in [4]. They establish the convergence of the algorithm for the Fletcher–Reeves (FR) and Dai–Yuan (DY) parameters with the corresponding sufficient descent condition and a desired inequality. We do not rely on satisfying these conditions. Instead, we demonstrate that our algorithm can obtain descent directions and establish its convergence not only under FR and DY but also conjugate descent (CD) parameters.

2. Nonlinear vector Riemannian conjugate gradient method

We consider the following unconstrained vector optimization problem:

$$\begin{array}{ll} \min_K & F(x) \\ \text{s.t.} & x \in \mathcal{M}, \end{array} \tag{1}$$

where $F : \mathcal{M} \to \mathbb{R}^m$ is continuously differentiable, $\mathcal{M}$ is an n-dimensional smooth Riemannian manifold and $K \subset \mathbb{R}^m$ is a closed, convex, and pointed (i.e., $K \cap (-K) = \{0\}$) cone with a nonempty interior. 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 $x \in \mathcal{M}$.

Now, consider the extension of the notion of the steepest descent direction to the vector-valued case. We denote it as $v \colon \mathbb{R}^n \to \mathbb{R}^n$ and define it as

$$v(x) := \arg\min_d \left\{ \varphi(\mathrm{D}F(x)d) + \frac{\|d\|^2}{2} \,\middle|\, d \in \mathbb{R}^n \right\}. \tag{2}$$

Here $\varphi(\mathrm{D}F(x)d) = \sup\{[\mathrm{D}F(x)d]^T w \mid w \in C\}$, where $C = \{\omega \in K^* \mid \|\omega\| = 1\}$, $\cdot^T$ denotes transpose, and K^* is the dual cone of K.

Assume that we have an iterative method generating iterates $\{x^k\}$. In Euclidean spaces, the update takes the form $x_{k+1} = x_k + t_k d_k$, while in the Riemannian case it is generalized as $x_{k+1} = R_{x_k}(t_k d_k)$, for $k = 0, 1, 2, \ldots$, where d_k is a descent direction, t_k is a step size, and R is a retraction that project points from the tangent space of the manifold onto the manifold itself.

In the Euclidean case, the search direction of the nonlinear vector conjugate gradient method is given by $d_{k+1} = v(x_{k+1}) + \beta_{k+1} d_k$, for $k \geq 0$, where $\beta_{k+1} \in \mathbb{R}$ is a parameter. To extend it to the Riemannian case, we use a vector transport called the differentiated retraction [1]. Using $\mathcal{T}^k : T_{x_k}\mathcal{M} \to T_{x_{k+1}}\mathcal{M}$ with $\mathcal{T}^k(d_k) := \mathcal{T}^R_{t_k d_k}(d_k) = \mathrm{D}R_{x_k}(t_k d_k)[d_k]$, we have

$$d_{k+1} = v(x_{k+1}) + \beta_{k+1} \mathcal{T}^k(d_k). \tag{3}$$

In order to get a proper decrease, we extend Wolfe conditions to vector optimization on Riemannian manifolds. Let $e \in K$ be given such that $0 < \langle \omega, e \rangle \leq 1$ for all $\omega \in C$. Letting $0 < c_1 < c_2 < 1$ and $e = [1, \ldots, 1]^T \in \mathbb{R}^m$, we propose the (weak) Wolfe conditions as follows:

$$F(R_{x_k}(t_k d_k)) \preceq_K F(x_k) + c_1 t_k \varphi(\mathrm{D}F(x_k)[d_k])\, e, \tag{4}$$

$$\varphi(\mathrm{D}F(R_{x_k}(t_k d_k))[\mathrm{D}R_{x_k}(t_k d_k)[d_k]]) \geq c_2 \varphi(\mathrm{D}F(x_k)[d_k]), \tag{5}$$

and the strong Wolfe conditions are given by (4), together with

$$|\varphi\left(\mathrm{D}F(R_{x_k}\left(t_k d_k\right))[\mathrm{D}R_{x_k}\left(t_k d_k\right)\left[d_k\right]]\right)| \\ \leq c_2|\varphi\left(\mathrm{D}F(x_k)[d_k]\right)|. \quad (6)$$

Then we can give our algorithm.

Algorithm 1 Nonlinear vector Riemannian conjugate gradient method (NVRCG)

Step 0. Let $x_0 \in \mathcal{M}$ and initialize $k \leftarrow 0$.
Step 1. Compute $v\left(x_k\right)$ as in (2). If $v\left(x_k\right) = 0$, then stop.
Step 2. Compute

$$d_k = \begin{cases} v\left(x_k\right), & \text{if } k = 0, \\ v\left(x_k\right) + \beta_k \mathcal{T}^{k-1}(d_{k-1}), & \text{if } k \geq 1, \end{cases}$$

where β_k is an algorithmic parameter.
Step 3. Compute a step size $t_k > 0$ by a line search procedure and set $x_{k+1} = R_{x_k}(t_k d_k)$.
Step 4. Set $k \leftarrow k+1$, and go to Step 1.

3. Convergence analysis

We extend Zoutendjik's type condition to Riemannian manifolds and analyze the convergence of the NVRCG method with the Riemannian extensions of FR, CD, and DY parameters.

First we give the Lipschitz-like continuity for vector optimization on Riemannian manifolds: For all $x \in \mathcal{M}, d \in T_x\mathcal{M}$ with $\|d\| = 1, t \geq 0$, there exist constant $L > 0$ such that $\|\mathrm{D}\left(F \circ R_x\right)(td)[d] - \mathrm{D}\left(F \circ R_x\right)(0)[d]\| \leq Lt$.

Assume this condition holds and consider the iteration, where d_k is a K-descent direction for F at x_k and t_k satisfies the Wolfe conditions. Then, we have

$$\sum_{k \geq 0} \frac{\varphi^2\left(\mathrm{D}F(x_k)[d_k]\right)}{\|d_k\|^2} < \infty,$$

which is an extension of standard Zoutendijk's condition.

Here, we extend FR, CD and DY parameters to vector optimization on Riemannian manifolds as follows:

$$\beta_k^{\mathrm{FR}} = \frac{\psi_{x_k,v_k}(0)}{\psi_{x_{k-1},v_{k-1}}(0)}, \quad (7)$$

$$\beta_k^{\mathrm{CD}} = \frac{\psi_{x_k,v_k}(0)}{\psi_{x_{k-1},d_{k-1}}(0)}, \quad (8)$$

$$\beta_k^{\mathrm{DY}} = \frac{-\psi_{x_k,v_k}(0)}{\psi_{x_{k-1},d_{k-1}}(t_{k-1}d_{k-1}) - \psi_{x_{k-1},d_{k-1}}(0)}. \quad (9)$$

Under strong Wolfe conditions, we can prove that d_k generated by (3) with CD or DY parameter is a descent direction satisfying some inequality. For FR parameter, we get the same result by adding an additional restriction on c_2 of strong Wolfe conditions. According to the above results, we have proved that if $0 \leq \beta_k \leq \beta_k^{\mathrm{FR}}$ and t_k satisfies the strong Wolfe conditions, then $\liminf_{k\to\infty} \|v\left(x_k\right)\| = 0$. We can get the same result for β^{CD} and β^{DY}. Moreover, the PRP-, HS-, and LS-type hybrid scheme of conjugate gradient methods to generate descent directions are also discussed in our work.

Theorem 1. *Each accumulation point of the sequence $\{x_k\}$ generated by Algorithm 1, if exists, is a Pareto stationary point.*

4. Numerical experiments

The results of the numerical experiments will be presented on the presentation day of the conference.

5. Future works

For future work, the convergence with the exact PRP, HS, and LS parameters will be discussed. Additionally, an analysis of the convergence rate will be conducted.

Reference

[1] Sato H. Riemannian *Optimization and Its Applications*. Cham: Springer, 2021.

[2] Lucambio Pérez L R, Prudente L F. Nonlinear conjugate gradient methods for vector optimization. *SIAM Journal on Optimization*, 2018, 28(3): 2690–2720.

[3] Boumal N. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

[4] Najafi S, Hajarian M. Multiobjective conjugate gradient methods on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 2023, 197(3): 1229–1248.