AIMS Mathematics, 11(5): 12580–12605.
DOI: 10.3934/math.2026517
Received: 20 January 2026
Revised: 31 March 2026
Accepted: 01 April 2026
Published: 07 May 2026

http://www.aimspress.com/journal/Math

Research article

A Hager–Zhang Riemannian conjugate gradient method for matrix approximation

Nasiru Salihu[1], **Seyed Yaser Mousavi Siamakani**[2,*], **Auwal Bala Abubakar**[3,4,5,6,*] **and Also Mohammed Saleh**[7]

[1] Department of Mathematics, Faculty of Sciences, Modibbo Adama University, Yola, 652105, Nigeria

[2] College of Engineering, Department of Civil Engineering, Rangsit University, Mueang, Pathum Thani, Thailand, 12000, Thailand

[3] Department of Art and Science, George Mason University, Songdomunhwa-ro 119-4, Yeonsu-gu, Incheon 21985, Republic of Korea

[4] Numerical Optimization Research Group, Department of Mathematical Sciences, Faculty of Physical Sciences, Bayero University, Kano 700241, Nigeria

[5] Department of Mathematics and Applied Mathematics, Sefako Makgatho Health Sciences University, Ga-Rankuwa, Pretoria 0204, Medunsa, South Africa

[6] Faculty of Education and Arts, Sohar University, Sohar 311, Oman

[7] College of Agriculture, Science and Technology, Jalingo, Nigeria

* **Correspondence:** Email: yaser.m@rsu.ac.th, ababubakar.mth@buk.edu.ng.

Abstract: We introduce a class of Hager–Zhang-type Riemannian conjugate gradient (CG) methods that generalize the framework of Sakai et al. (*Applied Mathematics and Computation*, 441 (2023) 127685) to arbitrary retractions while significantly extending its theoretical and practical scope. These methods ensure global convergence for non-convex problems without relying on strong convexity and inherently satisfy the sufficient descent property, independent of Riemannian line search conditions. A key algorithmic innovation is the introduction of an adaptive min(max) strategy to adjust the CG parameter β_{k+1} using bounded parameters to ensure numerical stability. Furthermore, we extended classical Euclidean CG portfolio optimization to the sphere manifold, naturally enforcing budget constraints and improving robustness. Numerical experiments, available at GitHub repo, showed that our methods outperform classical Riemannian CG in iterations and computational time for large-scale problems.

Keywords: manifold optimization; Hager–Zhang-type CG methods; sufficient descent property; scaled vector transport; convergence analysis

Mathematics Subject Classification: 65K05, 90C30

1. Introduction

Optimization on smooth manifolds has become a powerful tool for addressing complex problems in science and engineering. In this paper, we consider an unconstrained optimization problem on a Riemannian manifold:

$$\min_{x\in\mathcal{M}} f(x), \tag{1.1}$$

where $f : \mathcal{M} \to \mathbb{R}$ is a smooth function. The manifold $\mathcal{M}$ is equipped with a Riemannian metric, which defines a smoothly varying inner product on the tangent spaces $T_x\mathcal{M}$ at each point. The collection of all tangent spaces forms the tangent bundle:

$$T\mathcal{M} = \bigcup_{x\in\mathcal{M}} T_x\mathcal{M}.$$

For an in-depth understanding of Riemannian manifolds and optimization within this framework, we refer the reader to the works of Absil et al. [1] and Boumal [2]. Adapting classical optimization methods to Riemannian manifolds requires modifying key components, such as gradients, line searches, and search directions, to respect the geometry of the manifold. This is achieved via retractions, Riemannian gradients, and vector transports [2]. Among these, nonlinear conjugate gradient (CG) methods are particularly attractive for large-scale problems due to their low memory requirements and robust convergence [3].

These tools enable defining the iterative scheme

$$x^{k+1} = x^k + \alpha_k\delta^k,$$

where α_k is the step size and δ^k is the search direction, updated as

$$\delta^k = -g^k + \beta_k\delta^{k-1}, \tag{1.2}$$

with $g^k = \nabla f(x^k)$ the Euclidean gradient and β_k the CG parameter controlling the search structure.

In this work, we denote the Riemannian gradient by $\operatorname{grad} f(x) = g^x$, distinguishing it from the Euclidean gradient by context. Riemannian CG methods were introduced by Ring and Wirth [4], who analyzed the Riemannian Fletcher–Reeves (RFR) method under generalized Wolfe conditions, paving the way for subsequent developments and broader applications.

A major advancement came with the introduction of scaled vector transports by Sato [5], which facilitated the transfer of Euclidean convergence results to Riemannian settings, including methods such as the Dai–Yuan (DY) CG variant [6]. Thereafter, researchers have explored hybrid Riemannian CG schemes that combine update formulas [7, 8], denoted by HHSDY (Hybrid Hestenes–Stiefel–Dai–Yuan) and HFRPR (Hybrid Fletcher–Reeves–Polak–Ribière), respectively, as well as spectral Riemannian CG methods [9, 10] and three-term Riemannian CG structures [11].

Moreover, researchers have focused on modern CG parameters. Retraction-based convergence analyses have been developed for the Hager–Zhang (HZ) method [12, 13] and the RMIL

(Rivaie–Mustafa–Ismail–Leong) method [14], as discussed in [15–17]. These methods mainly differ in their choice of retraction: HZ-type methods typically use the exponential map, while RMIL-type methods employ more computationally efficient retractions, such as QR-based projections on Stiefel and oblique manifolds, or normalization on the sphere.

Consider the formula for the HZ CG parameter β_k [12, 13], which has been extended to the Riemannian setting in [8] as follows:

$$\beta_{k+1}^{\mathrm{RHZ}} = \gamma_{k+1}^{1} - \mu\gamma_{k+1}^{2}, \tag{1.3}$$

where

$$\Gamma_k = \langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}} - \langle g^k, \delta^k\rangle_{x^k},$$

and

$$\gamma_{k+1}^{1} = \frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}} - \langle g^k, \delta^k\rangle_{x^k}}, \quad \gamma_{k+1}^{2} := \frac{\|y^k\|_{x^{k+1}}^2 \langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\left(\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}} - \langle g^k, \delta^k\rangle_{x^k}\right)^2},$$

with $y^k = g^{k+1} - \mathcal{T}_{\alpha_k\delta^k}(g^k)$ and $\|\cdot\|_x$ represents the norm associated with the Riemannian inner product $\langle\cdot,\cdot\rangle_x$ on the tangent space $T_x\mathcal{M}$. This formulation utilizes the concept of vector transport $\mathcal{T}$, specifically the transport $\mathcal{T}_{\alpha_k\delta^k}$ along the step $\alpha_k\delta^k$, and $\mu = 2$ is a scalar that controls the influence of the correction term, based on Riemannian geometry (see Section 2).

Parameter μ plays a critical role in controlling the descent direction and ensuring the convergence behavior of the HZ method [18]. However, the classical versions [12, 13] and their Riemannian extensions [8, 15] do not adapt μ during the iterative process. Notably, μ can be made adaptive, and several dynamic tuning strategies have been proposed to adjust CG parameters automatically, as demonstrated in various studies [11, 19, 20]. Thus, the HZ framework ensures the steep-descent condition under $c = 1 - \frac{1}{4\mu}$, which is a fundamental property for ensuring convergence:

$$\langle g^k, \delta^k\rangle_x \leq -c\|g^k\|_x^2, \quad \text{for some constant } c > 0. \tag{1.4}$$

The convergence properties of the HZ-type Riemannian CG method are shown by replacing $\beta_{k+1}^{\mathrm{RHZ}}$ in (1.3) by

$$\hat{\beta}_{k+1}^{\mathrm{RHZ}} := \max\left\{\beta_{k+1}^{\mathrm{RHZ}}, \zeta_{k+1}\right\}, \quad \zeta_{k+1} := -\frac{1}{\|\delta^{k+1}\|_{x^{k+1}} \min\left\{\zeta, \|g^{k+1}\|_{x^{k+1}}\right\}},$$

where $\zeta > 0$ is a constant [15].

This modification ensures that the search direction δ^k remains bounded by the exponential retraction and Gauss's lemma. However, it relies on a strong convexity assumption, limiting its applicability to problems with a unique global minimizer and excluding many real-world non-convex problems, such as those in portfolio selection, machine learning, and deep learning. Its numerical implementation is further restricted to certain Riemannian optimization problems, e.g., on the unit sphere.

Moreover, in [15], the researchers proposed a Riemannian HZ method on the sphere using the exponential retraction and noted that "in a future paper, we will present the HZ method using a general retraction and its convergence analyses". In this work, we fulfill and extend this promise by developing two adaptive HZ-type Riemannian CG frameworks (MRHZ1 and MRHZ2) in Section 2 that: (i) Operate on general Riemannian manifolds beyond the sphere, (ii) handle non-convex problems, and (iii) include a robust parameter adaptation strategy. These enhancements enable

application to a broader class of real-world problems, including portfolio optimization. To validate the proposed methods, we apply them to certain Riemannian optimization problems and introduce a novel Riemannian optimization framework for portfolio selection by reformulating the problem on the sphere manifold, detailed in Section 6.

Summary of contributions: This work makes the following key contributions that collectively realize and substantially advance the research program outlined in [15]:

- **Generalized framework with non-convex convergence:** We introduce a new family of modified HZ-type CG methods (MRZH1 and MRZH2) within a general retraction-based Riemannian framework, thereby fulfilling the core technical promise of [15]. Crucially, our design avoids reliance on the Gauss lemma, relaxing the requirement for strong convexity. Theorems 4.4 and 4.5 ensure global convergence for these methods in non-convex settings via a norm-decreasing property using a scaled vector transport, making them robust for problems like portfolio selection on the sphere (Sections 3 and 4).
- **Robust parameter regulation:** To ensure practical stability in this generalized setting, we propose a novel regulation strategy based on a *min(max)* formulation (Eqs (2.9) and (2.12)). This strategy guarantees that the search directions satisfy the sufficient descent property without requiring the strong Wolfe conditions traditionally needed in Riemannian optimization [4, 7, 8], thereby simplifying implementation and enhancing robustness (Section 3).
- **Practical generalization to real-world problems:** We demonstrate the power of our generalized framework through a portfolio optimization application (Section 6). By reformulating the problem on the sphere manifold, we extend prior Euclidean CG methods [21–23] into a novel geometric framework. This approach naturally enforces the necessary simplex constraints, ensuring valid and independent portfolio weights. Our key findings are:
  - All Riemannian CG variants converge reliably, producing a consistent, efficient, and moderately diversified optimal portfolio.
  - The MRHZ1, MRHZ2, and RHZ methods emerge as the most efficient, offering the fastest convergence and most stable performance.
  - The robustness of the optimal solution is evidenced by the identical Sharpe ratio achieved by all methods.

Organization: In Section 2, we present the geometrical structures required in constructing the Riemannian CG methods by systematically adapting their Euclidean components to the manifold context. In Sections 3 and 4, we provide a rigorous convergence analysis, while in Section 5, we report extensive numerical experiments benchmarking our algorithms against state-of-the-art methods. In Section 6, we introduce the Riemannian optimization framework for portfolio selection by reformulating the problem on the sphere manifold. In Section 7, we conclude the paper.

2. Geometric ingredients and algorithm

In this section, we outline the core theoretical tools employed in our study of Riemannian CG methods. We begin by introducing some key geometric ingredients that are essential for the design and

analysis of Riemannian CG algorithms.

2.1. Retraction and vector transport

Standard gradient-based methods from Euclidean spaces cannot be directly applied in the Riemannian setting due to the manifold's nonlinear geometry. To address this, retraction mappings are used to project updated points onto the manifold, while vector transports facilitate the transfer of search directions across tangent spaces [4]. The Riemannian gradient, grad $f(x)$, is the unique vector in the tangent space $T_x\mathcal{M}$ that satisfies the equation:

$$\langle \operatorname{grad} f(x), \delta \rangle_x = \mathrm{D}f(x)[\delta], \quad \forall \delta \in T_x\mathcal{M}, \tag{2.1}$$

where $\mathrm{D}f(x)[\delta]$ represents the directional derivative of $f(c(t))$ at $t = 0$, with $c(t)$ being any smooth curve on the manifold such that $c(0) = x$ and $\dot{c}(0) = \delta$. This Riemannian gradient is critical for both the update rules and the theoretical analysis of the CG methods [2].

Definition 2.1 (Retraction [1])**.** *A retraction is a smooth mapping $R : T\mathcal{M} \to \mathcal{M}$ such that, for each $x \in \mathcal{M}$, its restriction $R_x : T_x\mathcal{M} \to \mathcal{M}$ satisfies the following properties:*

- *$R_x(0_x) = x$, where 0_x denotes the origin element in the tangent space $T_x\mathcal{M}$;*
- *$\mathrm{D}R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$, meaning that the derivative at 0_x acts as the identity mapping on $T_x\mathcal{M}$.*

To properly combine vectors from different tangent spaces in CG updates, we require a mechanism to "transport" vectors between them. This is achieved via:

Definition 2.2 (Vector transport [1])**.** *Let R be a retraction on $\mathcal{M}$. A smooth map $\mathcal{T} : T\mathcal{M} \oplus T\mathcal{M} \to T\mathcal{M}$ is called a vector transport if, for all $x \in \mathcal{M}$ and $\delta, \eta, \xi \in T_x\mathcal{M}$, it holds that:*

- *$\mathcal{T}_\delta(\eta) \in T_{R_x(\delta)}\mathcal{M}$;*
- *$\mathcal{T}_{0_x}(\eta) = \eta$;*
- *$\mathcal{T}_\delta(a\eta + b\xi) = a\mathcal{T}_\delta(\eta) + b\mathcal{T}_\delta(\xi)$, for all scalars $a, b \in \mathbb{R}$.*

A common example of vector transport is based on the differentiated retraction defined in [1] as:

$$\mathcal{T}_{\delta_x}(\xi_x) := \mathrm{D}R_x(\delta_x)[\xi_x], \quad \delta_x, \xi_x \in T_x\mathcal{M},\ x \in \mathcal{M}. \tag{2.2}$$

In practical implementations of Riemannian CG methods, it is often beneficial to adopt scaled vector transports that preserve or reduce the norm of transported vectors. A notable approach introduced by Sato and Iwai [5] defines transport as:

$$\mathcal{T}_{\delta_x}(\xi_x) := \begin{cases} \mathcal{T}_{\delta_x}(\xi_x), & \text{if } \|\mathcal{T}_{\delta_x}(\xi_x)\|_{R_x(\delta_x)} \le \|\xi_x\|_x, \\ \mathcal{T}^0_{\delta_x}(\xi_x), & \text{otherwise,} \end{cases} \tag{2.3}$$

where $\mathcal{T}^0_{\delta_x}(\xi_x) = \frac{\|\xi_x\|_x}{\|\mathcal{T}^R_{\delta_x}(\xi_x)\|_{R_x(\delta_x)}} \mathcal{T}_{\delta_x}(\xi_x)$, which guarantees norm-decreasing property:

$$\|\mathcal{T}_{\delta_x}(\xi_x)\|_{R_x(\delta_x)} \le \|\xi_x\|_x. \tag{2.4}$$

This scaling is crucial to maintaining sufficient descent, a property central to the convergence of CG methods on manifolds.

In [15], the exponential map is considered as a retraction $R := \exp$, and the vector transport is defined by the differential of the exponential retraction:

$$\mathcal{T} : T\mathcal{M} \oplus T\mathcal{M} \to T\mathcal{M} : (\delta, \xi) \mapsto \mathcal{T}_\delta(\xi) := (d \exp_x)_\delta(\xi), \tag{2.5}$$

for $\delta, \xi \in T_x\mathcal{M}$, and it is shown from the Gauss lemma that:

$$\langle g^k, \delta^k \rangle_{x^k} = \langle \mathcal{T}_{\alpha_k \delta^k}(g^k), \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle_{x^{k+1}}. \tag{2.6}$$

Although the exponential map $R_x := \exp_x$ and its differential (2.5) provide satisfactory theoretical results, particularly in preserving inner products via the Gauss Lemma, this approach has significant practical drawbacks for iterative optimization.

Practical limitations: Although the exponential map and its differential preserve inner products via the Gauss Lemma, they present two major practical challenges:

- **High computational cost:** Evaluating $\exp_x$ and $(d \exp_x)_\delta$ requires integrating geodesics or computing matrix exponentials, making each iteration expensive for large-scale problems. Therefore, cheaper retractions and their differentiated transports, such as in (2.2), are preferred.
- **Numerical instability:** Differential $(d \exp_x)_\delta$ can amplify vector norms in curved regions, destabilizing search directions and iterative convergence [5]. Scaled vector transports, as in (2.3), enforce a norm-decreasing property (2.4), providing a robust and practical alternative.

Consequently, the update scheme of a Riemannian CG method, adapted from the classical case, is as follows:

$$\begin{aligned} x^{k+1} &= R_{x^k}(\alpha_k \delta^k), \\ \delta^{k+1} &= -\operatorname{grad} f(x^{k+1}) + \beta_{k+1} \mathcal{T}_{\alpha_k \delta^k}(\delta^k), \end{aligned} \tag{2.7}$$

where $\alpha_k > 0$ is the step size, δ^k is the search direction, and β_{k+1} is a Riemannian CG parameter.

The search direction in (2.7) is motivated by the classical CG update (1.2), adapted to the Riemannian setting. Since the gradient $\operatorname{grad} f(x^{k+1})$ and the previous search direction δ^k lie in different tangent spaces, the vector transport $\mathcal{T}_{\alpha_k \delta^k}$ is used to map δ^k from $T_{x^k}\mathcal{M}$ to $T_{x^{k+1}}\mathcal{M}$. This ensures that the combination of $-\operatorname{grad} f(x^{k+1})$ and the transported direction $\mathcal{T}_{\alpha_k \delta^k}(\delta^k)$ is well-defined in the tangent space at x^{k+1}, preserving the conjugacy properties and enabling the Riemannian CG method to progress properly on the manifold.

Using the above tools, we provide HZ-type Riemannian CG methods next.

2.2. Adaptive HZ-type Riemannian CG frameworks

To relax the requirement of strong convexity in formulation (1.3), we propose two adaptive Riemannian CG HZ-type frameworks, denoted as modified HZ methods (MRHZ1 and MRHZ2), defined as follows:

$$\beta_{k+1}^{\mathrm{MRHZ1}} = \beta_{k+1}^{\mathrm{RHZ}} - \min\{\beta_{k+1}^{\mathrm{RHZ}}, \ -\mu_{k+1}^{\mathrm{RHZ1}} \gamma_{k+1}^2\}, \tag{2.8}$$

where

$$\mu_{k+1}^{\mathrm{RHZ1}} = \min\left(\mu_{\max}, \max\left(\mu_{\min}, \rho_{k+1}^{RHZ1}\right)\right), \tag{2.9}$$

and ρ_{k+1}^{RHZ1} is further defined by

$$\rho_{k+1}^{RHZ1} = \begin{cases} \rho_{k+1}, & \text{if } \Gamma_k \geq \|y^k\|_{x^{k+1}} \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}, \\ \mu_{\min}, & \text{otherwise,} \end{cases} \tag{2.10}$$

with $\rho_{k+1} := \|g^{k+1}\|_{x^{k+1}} - 2$.

The second parameter, denoted $\beta_{k+1}^{\mathrm{MRHZ2}}$, is defined as

$$\beta_{k+1}^{\mathrm{MRHZ2}} = \gamma_{k+1}^3 - \min\{\gamma_{k+1}^3,\ \mu_{k+1}^{\mathrm{RHZ2}}\, \gamma_{k+1}^4\}, \tag{2.11}$$

where

$$\gamma_{k+1}^3 = \frac{\langle g^{k+1}, y^k \rangle_{x^{k+1}}}{\|g^k\|_{x^k}^2}, \quad \gamma_{k+1}^4 = \frac{\|y^k\|_{x^{k+1}}^2 \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle_{x^{k+1}}}{\|g^k\|_{x^k}^4},$$

and the adaptive parameter is updated via:

$$\mu_{k+1}^{\mathrm{RHZ2}} = \min\left(\mu_{\max}, \max\left(\mu_{\min}, \rho_{k+1}^{RHZ2}\right)\right), \tag{2.12}$$

a bounded safeguard parameter controlling the correction term, where ρ_{k+1}^{RHZ2} is defined by

$$\rho_{k+1}^{RHZ2} := \frac{\|g^k\|_{x^k}^4 - \|g^{k+1}\|_{x^{k+1}} \|y^k\|_{x^{k+1}} \|g^k\|_{x^k}^2 \|\delta^k\|_{x^k}}{\|y^k\|_{x^{k+1}}^2 \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}^2}. \tag{2.13}$$

Parameter $\rho_{k+1}^{\mathrm{RHZ2}}$ in (2.13) is motivated by the Riemannian HZ-type CG parameter ξ_{k+1} and is designed to ensure the global convergence of MRHZ2; without it, the search direction may become unbounded, violating the sufficient descent condition. We therefore set $\mu_{k+1} = \rho_{k+1}$ and enforce the bounds

$$\mu_{k+1} = \min\left(\mu_{\max}, \max(\mu_{\min}, \rho_{k+1})\right),$$

typically with $\mu_{\min} = 1$ and $\mu_{\max} = 100$, to ensure stable and controlled updates. The lower bound prevents the descent factor from becoming too small, avoiding underflow or stalled convergence, while the upper bound limits excessively large updates that could destabilize the iteration. Conceptually, this mechanism is analogous to a trust-region strategy: It dynamically adapts the step size according to the local geometry of the manifold and the conditioning of the Riemannian Hessian, ensuring both stability and efficiency in the optimization process.

2.3. Riemannian optimization problems

The following examples summarize the Riemannian optimization problems (ROPs) employed to assess the proposed HZ-type CG methods:

Example 2.1. *Orthogonal procrustes problem [24]*

$$\begin{aligned} \min_{X \in \mathrm{St(p,n)}} \quad & f(X) = \|AX - B\|_F^2, \\ \textit{where} \quad & \mathrm{St(p,n)} := \{\mathrm{X} \in \mathbb{R}^{\mathrm{n}\times\mathrm{p}} \mid \mathrm{X}^{\mathrm{T}}\mathrm{X} = \mathrm{I_p}\}, \end{aligned}$$

with $\|\cdot\|_F$ *denoting the Frobenius norm,* $A \in \mathbb{R}^{l\times n}$, $B \in \mathbb{R}^{l\times p}$, *and* $l = n$.

Example 2.2. *Rayleigh-quotient minimization on the unit sphere [1]*

$$\begin{aligned} \min_{x\in\mathbb{S}^{n-1}} \quad & f(x) = x^T A x, \\ \textit{where} \quad & \mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \mid \|x\| = 1\}, \end{aligned}$$

with $A \in S_{++}^{n\times n}$, a symmetric positive-definite matrix, and $\|\cdot\|$ denoting the Euclidean norm.

Remark 2.3. *Although we focus on the sphere and Stiefel manifolds, the proposed Riemannian CG framework applies to any Riemannian manifold satisfying standard smoothness and retraction assumptions [1]. Examples include fixed-rank matrices [25], hyperbolic spaces [26], and more general Lie groups [27], provided a smooth retraction and suitable vector transport are available. The adaptive parameter update depends on local gradient information and therefore responds to the local curvature of the manifold. For manifolds with high curvature variations, step sizes and search directions may require careful tuning, but the global convergence framework remains applicable, highlighting the method's generality beyond the examples considered.*

The implementation of the HZ-type methods is described in the following.

Algorithm 1: Riemannian HZ-type conjugate gradient method

Input:

- Riemannian manifold $\mathcal{M}$;
- Retraction R;
- Objective function f on $\mathcal{M}$;
- Initial point $x^1 \in \mathcal{M}$;
- Tolerance $\epsilon > 0$.

Steps:

1. Initialize the search direction: $\delta^1 = -g^1$.
2. **For** $k = 1, 2, \ldots$, perform the following:
   - If $\|g^k\|_{x^k} \le \epsilon$, terminate the algorithm.
3. Compute a suitable step size $\alpha_k > 0$ such that:

$$f(R_{x^k}(\alpha_k\delta^k)) \le f(x^k) + \sigma_1\,\alpha_k\,\langle g^k, \delta^k\rangle_{x^k}, \tag{2.14}$$

$$\left|\langle \operatorname{grad} f(R_{x^k}(\alpha_k\delta^k)), \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{R_{x^k}(\alpha_k\delta^k)}\right| \le \sigma_2\left|\langle g^k, \delta^k\rangle_{x^k}\right|, \tag{2.15}$$

   where $0 < \sigma_1 < \sigma_2 < 1$.
4. Generate the updated point:

$$x^{k+1} = R_{x^k}(\alpha_k\delta^k).$$

5. Update the search direction δ^{k+1} by the formula (2.7).
6. Compute the CG coefficient β_{k+1} using (2.8) or (2.11).
7. Increment k and return to Step 2.

Output: The sequence of iterates $\{x^k\}$ on the manifold $\mathcal{M}$, representing the solution after convergence.

This algorithm utilizes two HZ-inspired Riemannian CG parameters, denoted by $\beta_{k+1}^{\mathrm{MRHZ1}}$ and $\beta_{k+1}^{\mathrm{MRHZ2}}$, as specified in Eqs (2.8) and (2.11), respectively.

3. Riemannian Hager–Zhang-type conjugate gradient descent properties

In the subsequent subsections, we show that δ^{k+1} in Algorithm 1, as determined by the HZ-type Riemannian methods, fulfills the sufficient descent condition (1.4) independently of the selected step size α_k.

3.1. Descent property of the MRHZ1 CG method

Theorem 3.1. *Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth objective function. Consider the sequence of iterates $\{x^k\}$ generated by Algorithm 1 using the search direction update* (2.7) *and the modified CG parameter $\beta_{k+1}^{\mathrm{MRHZ1}}$ defined in* (2.8)*. Furthermore, let the scalar sequence $\{\mu_{k+1}\}$ be confined to the interval $[\mu_{\min}, \mu_{\max}]$ with its lower bound satisfying $\mu_{\min} > 1/4$. Then, the generated search direction δ^{k+1} satisfies the sufficient descent condition: For every iteration $k \geq 0$,*

$$\langle g^{k+1}, \delta^{k+1} \rangle_{x^{k+1}} \leq -\Big(1 - \frac{1}{4\mu_{\min}}\Big) \|g^{k+1}\|^2_{x^{k+1}}. \tag{3.1}$$

Proof. If the safeguard term dominates, i.e.,

$$\mu_{k+1} \frac{\langle g^{k+1}, y^k \rangle_{x^{k+1}} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle_{x^{k+1}}}{\Gamma_k^2} > \beta_{k+1}^{\mathrm{RHZ}},$$

then $\beta_{k+1}^{\mathrm{MRHZ1}} = 0$, which implies $\delta^{k+1} = -g^{k+1}$. In this case, the search direction automatically satisfies condition (3.1), as it results in a strictly negative inner product with the gradient.

Otherwise, since $\mu_{\min} = 1$ and using the expression for $\beta_{k+1}^{\mathrm{MRHZ1}}$ with δ^{k+1}, we get:

$$\begin{aligned} \langle g^{k+1}, \delta^{k+1} \rangle_{x^{k+1}} \leq & - \|g^{k+1}\|^2_{x^{k+1}} + \frac{\langle g^{k+1}, y^k \rangle_{x^{k+1}} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle_{x^{k+1}}}{\Gamma_k} \\ & - 2 \frac{\|y^k\|^2_{x^{k+1}} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle^2_{x^{k+1}}}{\Gamma_k^2} + \frac{\|y^k\|^2_{x^{k+1}} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle^2_{x^{k+1}}}{\Gamma_k^2}. \end{aligned}$$

Let us define:

$$a := \frac{\sqrt{\Gamma_k}}{\sqrt{2}} g^{k+1}, \quad b := \frac{\sqrt{2} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle_{x^{k+1}}}{\sqrt{\Gamma_k}} y^k.$$

By Cauchy–Schwarz and arithmetic–geometric mean inequalities:

$$\langle a, b \rangle_{x^{k+1}} \leq \frac{1}{2} (\|a\|^2_{x^{k+1}} + \|b\|^2_{x^{k+1}}).$$

So that

$$\begin{aligned} \|a\|^2_{x^{k+1}} &= \tfrac{\Gamma_k}{2} \|g^{k+1}\|^2_{x^{k+1}}, \\ \|b\|^2_{x^{k+1}} &= \tfrac{2}{\Gamma_k} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle^2_{x^{k+1}} \|y^k\|^2_{x^{k+1}}. \end{aligned}$$

It follows that

$$\langle a, b \rangle_{x^{k+1}} \leq \tfrac{1}{2} \Big(\tfrac{\Gamma_k}{2} \|g^{k+1}\|^2_{x^{k+1}} + \tfrac{2}{\Gamma_k} \langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \rangle^2_{x^{k+1}} \|y^k\|^2_{x^{k+1}} \Big).$$

Substituting into the expression in $\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}}$ for the second term, we obtain:

$$\begin{aligned}\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} &\leq -\|g^{k+1}\|^2_{x^{k+1}} + \frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\Gamma_k} - \frac{\|y^k\|^2_{x^{k+1}}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle^2_{x^{k+1}}}{\Gamma_k^2} \\ &= \frac{-\|g^{k+1}\|^2_{x^{k+1}}\Gamma_k + \langle a, b\rangle_{x^{k+1}} - \frac{\|y^k\|^2_{x^{k+1}}}{\Gamma_k}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle^2_{x^{k+1}}}{\Gamma_k} \\ &= \frac{-\|g^{k+1}\|^2_{x^{k+1}}\Gamma_k + \frac{\Gamma_k}{4}\|g^{k+1}\|^2_{x^{k+1}}}{\Gamma_k}.\end{aligned}$$

Applying the above inequality to $\langle a, b\rangle_{x^{k+1}}$, we get:

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq \frac{1}{\Gamma_k}\left(-\|g^{k+1}\|^2_{x^{k+1}}\Gamma_k + \frac{\Gamma_k}{4}\|g^{k+1}\|^2_{x^{k+1}}\right),$$

yields the sufficient descent condition

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq -\left(1 - \frac{1}{4}\right)\|g^{k+1}\|^2_{x^{k+1}}.$$

Hence, the proof is concluded. □

3.2. Descent property of the MRHZ2 CG method

Theorem 3.2. *Consider a smooth function $f : \mathcal{M} \to \mathbb{R}$ and a sequence of iterates $\{x^k\}$ generated by Algorithm 1. Suppose that the search direction δ^{k+1} is computed according to* (2.7)*, and CG parameter $\beta^{\text{MRHZ2}}_{k+1}$ is determined by* (2.11)*. If the controlling sequence $\{\mu_{k+1}\}$ is chosen such that $\mu_{k+1} \in [\mu_{\min}, \mu_{\max}]$ with $\mu_{\min} > \frac{1}{4}$, then the generated search directions satisfy the sufficient descent condition*

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq -\left(1 - \frac{1}{4\mu_{\min}}\right)\|g^{k+1}\|^2_{x^{k+1}}, \quad \forall k \geq 0. \tag{3.2}$$

Proof. If the safeguard term is dominant, meaning condition $\mu_{k+1}\gamma^4_{k+1} > \gamma^3_{k+1}$ holds, then the CG parameter simplifies to $\beta^{\text{MRHZ2}}_{k+1} = 0$. Consequently, the search direction reduces to the steepest descent direction, $\delta^{k+1} = -g^{k+1}$. In this scenario, the descent condition (3.2) is trivially satisfied, as the inner product evaluates to $\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} = -\|g^{k+1}\|^2_{x^{k+1}}$, which fulfills the inequality.

When the safeguard term is not dominant, the expression for $\beta^{\text{MRHZ2}}_{k+1}$ must be analyzed directly. Substituting its full definition yields:

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} = -\|g^{k+1}\|^2_{x^{k+1}} + \beta^{\text{MRHZ2}}_{k+1}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}. \tag{3.3}$$

Given that $\mu_{\min} = 1$, the expression for $\beta^{\text{MRHZ2}}_{k+1}$ yields:

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq -\|g^{k+1}\|^2_{x^{k+1}} + \left(\frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\|g^k\|^2_{x^k}} - \gamma^4_{k+1}\right)\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}.$$

Now define the vectors:

$$a := \frac{\|g^k\|_{x^k}}{\sqrt{2}}g^{k+1}, \quad b := \frac{\sqrt{2}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\|g^k\|_{x^k}}y^k.$$

Using inequality $\langle a, b\rangle_{x^{k+1}} \leq \frac{1}{2}(\|a\|^2_{x^{k+1}} + \|b\|^2_{x^{k+1}})$, we obtain:

$$\langle a, b\rangle_{x^{k+1}} \leq \frac{1}{2}\left(\frac{\|g^{k+1}\|^2_{x^{k+1}}\,\|g^k\|^2_{x^k}}{2} + \frac{2\,\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle^2_{x^{k+1}}\|y^k\|^2_{x^{k+1}}}{\|g^k\|^2_{x^k}}\right).$$

Substituting into the expression in $\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}}$ for the second term, we obtain:

$$\begin{aligned}\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} &\leq -\|g^{k+1}\|^2_{x^{k+1}} + \frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\|g^k\|^2_{x^k}} - \frac{\|y^k\|^2_{x^{k+1}}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle^2_{x^{k+1}}}{\|g^k\|^4_{x^k}} \\ &= \frac{-\|g^{k+1}\|^2_{x^{k+1}}\|g^k\|^2_{x^k} + \langle a, b\rangle_{x^{k+1}} - \frac{\|y^k\|^2_{x^{k+1}}}{\|g^k\|^2_{x^k}}\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle^2_{x^{k+1}}}{\|g^k\|^2_{x^k}} \\ &= \frac{-\|g^{k+1}\|^2_{x^{k+1}}\|g^k\|^2_{x^k} + \frac{\|g^k\|^2_{x^k}}{4}\|g^{k+1}\|^2_{x^{k+1}}}{\|g^k\|^2_{x^k}}.\end{aligned}$$

Applying the earlier inequality for $\langle a, b\rangle_{x^{k+1}}$ and simplifying, we get:

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq \frac{1}{\|g^k\|^2_{x^k}}\left(-\|g^{k+1}\|^2_{x^{k+1}}\,\|g^k\|^2_{x^k} + \frac{\|g^{k+1}\|^2_{x^{k+1}}\,\|g^k\|^2_{x^k}}{4}\right).$$

Thus, the inequality

$$\langle g^{k+1}, \delta^{k+1}\rangle_{x^{k+1}} \leq -\left(1 - \frac{1}{4}\right)\|g^{k+1}\|^2_{x^{k+1}}$$

holds, which completes the proof. □

4. Global convergence of Riemannian HZ-type methods

To establish the global convergence of the proposed Riemannian HZ-type CG methods, we require the following set of assumptions on the objective function f.

4.1. Basic assumptions on the problem

Assumption 4.1 (Smoothness and Boundedness)**.** *Let $f : \mathcal{M} \to \mathbb{R}$ be continuously differentiable.*

(1) ***Boundedness from below and level set compactness****: The objective function is bounded below on the level set ϱ defined with respect to the initial iterate x^1:*

$$\varrho := \{x \in \mathcal{M} : f(x) \leq f(x^1)\}. \tag{4.1}$$

Furthermore, we assume level set ϱ is compact. This condition prevents the objective value from diverging to $-\infty$ and guarantees that all iterates remain within a bounded region of the manifold.

(2) ***Lipschitz-type condition along retractions****: There exists a constant $L > 0$ such that, for every point x in the level set ϱ and every tangent direction $\delta \in T_x\mathcal{M}$ with unit norm ($\|\delta\|_x = 1$), the following inequality holds:*

$$\left|\mathrm{D}(f \circ R_x)(t\delta)[\delta] - \mathrm{D}(f \circ R_x)(0)[\delta]\right| \leq L\,t, \tag{4.2}$$

for all $t \geq 0$. Here, R_x denotes the retraction map. This condition controls the variation of the directional derivative of the objective along retraction paths and is a standard requirement to ensure that the step size can be selected effectively via a Wolfe-type line search.

Assumption 4.1 guarantees the well-posedness of the optimization problem. The compactness of the level set and the boundedness of f ensure that the sequence of iterates remains in a controlled domain. The Lipschitz-type condition in (4.2) is a mild smoothness requirement on the pullback $f \circ R_x$; it is strictly weaker than assuming Lipschitz continuity of the gradient of the pullback and is satisfied for a wide range of practical problems on matrix manifolds (e.g., the Brockett cost function and the Rayleigh quotient) [5].

4.2. Boundedness of the gradient sequence

Building on the compactness of the level set ϱ and the smoothness of f, we note an immediate and crucial consequence, which we adopt as an additional condition for our analysis. It follows from standard results in Riemannian optimization [28] that the Riemannian gradient remains uniformly bounded on ϱ. Formally, there exists a constant $\psi > 0$, such that

$$\| \operatorname{grad} f(x^k) \|_{x^k} \leq \psi \quad \text{for all } k. \tag{4.3}$$

For notational brevity, as mentioned in the analysis that follows, we will denote $\operatorname{grad} f(x^k)$ as g^k.

Under Assumption 4.1 and the gradient boundedness condition (4.3), if the step sizes α_k are chosen to satisfy the standard Wolfe conditions (see conditions (2.14)), it can be shown that the generated search directions are descent directions and that the corresponding sequence of objective values $\{f(x^k)\}$ is monotonically non-increasing.

The global convergence analysis for Riemannian HZ-type CG methods will be established using a Riemannian adaptation of the classical Zoutendijk theorem [4, 5]. The boundedness of the gradients (4.3) and the descent property of the search directions are key ingredients in this argument.

Lemma 4.2. *Let $(\mathcal{M}, g)$ be a Riemannian manifold with retraction R, and let $f : \mathcal{M} \to \mathbb{R}$ satisfy Assumption 4.1. If the step sizes $\alpha_k > 0$ in Algorithm 1 satisfy* (2.14) *and* (2.15)*, then the following series converges:*

$$\sum_{k=0}^{\infty} \frac{\langle g^k, \delta^k \rangle_{x^k}^2}{\left\| \delta^k \right\|_{x^k}^2} < \infty. \tag{4.4}$$

Proof. Following the approach outlined in [9, 11], we combine the Wolfe condition (2.15) with (2.1) and (2.2), along with the Lipschitz continuity assumption in Assumption 4.1, to derive the lower bound for α_k:

$$\alpha_k \geq \frac{(\sigma_2 - 1)\langle g^k, \delta^k \rangle_{x^k}}{L \|\delta^k\|_{x^k}^2}.$$

By substituting this bound into the sufficient decrease condition, and summing over all iterations, we obtain the desired result, utilizing the boundedness of f.

Next, observe that the step size α_k that satisfies the conditions in (2.14) and (2.15) must also fulfill the Wolfe line search criterion (2.14) and

$$\left\langle \operatorname{grad} f(R_{x_k}(\alpha_k \delta^k)), \mathcal{T}_{\alpha_k \delta^k}(\delta^k) \right\rangle_{R_{x^k}(\alpha_k \delta^k)} \geq \sigma_2 \langle g^k f(x_k), \delta^k \rangle_{x^k}.$$

Therefore, under the strong Wolfe line search conditions as noted in [29], Lemma 4.2 is satisfied. □

Remark 4.3. *The global convergence of HZ-type Riemannian CG methods is typically established under convexity assumptions. The proof of Theorem 4.2 follows a convergence analysis analogous to the classical Zoutendijk theorem in Euclidean optimization [30]. In contrast, Euclidean HZ-type CG methods [13] rely on strong convexity and require step lengths satisfying the Wolfe conditions, specifically,* $(\nabla f(x^{k+1}) - \nabla f(x^k))^\top (x^{k+1} - x^k) \geq \gamma\alpha_k \|\delta^k\|^2$. *The analogous condition in the Riemannian setting is* $\langle y^k, \mathcal{T}_{\alpha_k\delta^k}\delta^k\rangle_{x^{k+1}} \geq \gamma\alpha_k \|\delta^k\|^2_{x^k}$, *as shown in the convergence analysis of the RHZ method [15]. In Riemannian terms, this condition ensures that the eigenvalues of the Riemannian Hessian are bounded below by a positive constant. However, Theorems 4.4 and 4.5 guarantee global convergence of the MRHZ1 and MRHZ2 methods (Eqs* (2.8) *and* (2.11)*) without requiring convexity. Instead, convergence is ensured through a norm-decreasing property, and there is no need for the eigenvalues of the Riemannian Hessian to be uniformly bounded below, as discussed in [15]. Regarding local convergence rates or worst-case iteration complexity, this framework does not assume convexity or Hessian bounds, so establishing such results would require additional assumptions (e.g., a Łojasiewicz inequality) and is beyond the scope of this work. Our focus here is on global convergence, which is the standard and practically relevant criterion for nonlinear Riemannian CG methods.*

4.3. Convergence of the MRHZ1 CG method

Theorem 4.4. *Let* $\{x^k\}$ *be the sequence generated by Algorithm 1 with* β_{k+1} *defined by* (2.8) *under Assumption 4.1 and with step sizes* α_k *chosen by the Riemannian strong Wolfe conditions. Then*

$$\liminf_{k\to\infty} \|g^k\|_{x^k} = 0. \tag{4.5}$$

Proof. Assume that the condition in (4.5) does not hold. Since the sequence $\{g^k\}$ generated by Algorithm 1 satisfies $\|g^k\|_{x^k} > \varrho$ for some constant $\varrho > 0$, there exists an integer k_0 such that for all $k \geq k_0$, the following condition holds:

$$\|g^k\|_{x^k} \geq \varrho. \tag{4.6}$$

Now, referring to the expression for $\beta^{\mathrm{MRHZ1}}_{k+1}$ in (2.8), we proceed to analyze this further:

$$\beta^{\mathrm{MRHZ1}}_{k+1} = \beta^{\mathrm{RHZ}}_{k+1} - \min\{\beta^{\mathrm{RHZ}}_{k+1},\ -\mu^{\mathrm{RHZ1}}_{k+1}\,\gamma^2_{k+1}\}.$$

Using this, we estimate the norm of $\beta^{\mathrm{MRHZ1}}_{k+1}$ as follows:

$$\begin{aligned}
\left\|\beta^{\mathrm{MRHZ1}}_{k+1}\right\| &= \left\|\beta^{\mathrm{RHZ}}_{k+1} + \mu^{\mathrm{RHZ1}}_{k+1}\gamma^2_{k+1}\right\| \\
&\leq \left|\frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\Gamma_k}\right| + 2\|y^k\|^2_{x^{k+1}} \left|\frac{\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\Gamma^2_k}\right| \\
&\quad + \mu^{\mathrm{RHZ1}}_{k+1}\|y^k\|^2_{x^{k+1}} \left|\frac{\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\Gamma^2_k}\right|.
\end{aligned}$$

Applying the Cauchy–Schwarz inequality and the inequality (2.4), we obtain:

$$|\langle g^{k+1}, y^k\rangle_{x^{k+1}}| \leq \|g^{k+1}\|_{x^{k+1}}\|y^k\|_{x^{k+1}},$$

$$|\langle g^{k+1}, \mathcal{T}_{\alpha_k \delta^k}(\delta^k)\rangle_{x^{k+1}}| \leq \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}.$$

Substituting these together with $\Gamma_k \geq \|g^{k+1}\|_{x^{k+1}} \|y^k\|_{x^{k+1}} \|\delta^k\|_{x^k}$ into the inequality for $\left\|\beta_{k+1}^{\mathrm{MRHZ1}}\right\|$, we obtain:

$$\begin{aligned}\left\|\beta_{k+1}^{\mathrm{MRHZ1}}\right\| =& \frac{\|g^{k+1}\|_{x^{k+1}} \|y^k\|_{x^{k+1}}}{\Gamma_k} + 2\frac{\|y^k\|_{x^{k+1}}^2 \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}}{\Gamma_k^2} \\ &+ \mu_{k+1}^{\mathrm{RHZ1}} \frac{\|y^k\|_{x^{k+1}}^2 \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}}{\Gamma_k^2} \\ \leq& \frac{1}{\|\delta^k\|_{x^k}} + \frac{2}{\|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}} + \mu_{k+1}^{\mathrm{RHZ1}} \frac{1}{\|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}}.\end{aligned}$$

From the update rule (2.9) and (2.10), we set

$$\mu_{k+1}^{\mathrm{RMHZ1}} = \|g^{k+1}\|_{x^{k+1}} - 2.$$

Substituting this into the previous inequality and simplifying leads to

$$\left\|\beta_{k+1}^{\mathrm{MRHZ1}}\right\| \leq \frac{2}{\|\delta^k\|_{x^k}}.$$

Let us define:

$$\vartheta_0 = \max\left\{\|\delta^k\|_{x^k} : 0 \leq k \leq k_0\right\},$$

and let

$$\vartheta = \vartheta_0 + \vartheta_1.$$

Applying (2.4) together with the update rule (2.7), we obtain:

$$\begin{aligned}\|\delta^{k+1}\|_{x^{k+1}} &\leq \|g^{k+1}\|_{x^{k+1}} + |\beta_{k+1}^{\mathrm{MRHZ1}}| \|\mathcal{T}_{\alpha_k \delta^k}(\delta^k)\|_{x^{k+1}} \\ &\leq \|g^{k+1}\|_{x^{k+1}} + 2 \leq \psi + 2 \triangleq \vartheta_1.\end{aligned}$$

Thus, we obtain the following bound for the norm of the search direction:

$$\|\delta^k\|_{x^k} \leq \vartheta, \quad \forall k. \tag{4.7}$$

From (4.6) and (4.7), we conclude:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \geq \sum_{k=k_0}^{\infty} \frac{\varrho^4}{\vartheta^2} = +\infty.$$

Next, using the inequality (3.1), we know:

$$\langle g^k, \delta^k\rangle_{x^k}^2 \geq \left(1 - \frac{1}{4}\right)^2 \|g^k\|_{x^k}^4.$$

This leads to:

$$\frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \leq \frac{1}{\left(1 - \frac{1}{4}\right)^2} \frac{\langle g^k, \delta^k\rangle_{x^k}^2}{\|\delta^k\|_{x^k}^2}.$$

Now, summing both sides over k, we get:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \leq \frac{1}{\left(1-\frac{1}{4}\right)^2} \sum_{k=0}^{\infty} \frac{\langle g^k, \delta^k\rangle_{x^k}^2}{\|\delta^k\|_{x^k}^2}.$$

Since the series on the right-hand side is finite by (4.4), it follows that the entire sum must be finite:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} < +\infty,$$

which leads to a contradiction. Therefore, the assumption that (4.5) does not hold is false. □

4.4. Convergence of the MRHZ2 CG method

Theorem 4.5. *Let $\{x^k\}$ be generated by Algorithm 1 with β_{k+1} given by* (2.11)*, under Assumption 4.1. If the step sizes α_k satisfy the Riemannian strong Wolfe conditions, then*

$$\liminf_{k\to\infty} \|g^k\|_{x^k} = 0. \tag{4.8}$$

Proof. We proceed by assuming, for the sake of contradiction, that (4.8) does not hold. By the properties of the sequence $\{g^k\}$ generated by Algorithm 1, we know that $\|g^k\|_{x^k} > \varrho$ for some constant $\varrho > 0$. Hence, there exists an integer k_0 such that for all $k \geq k_0$, the following condition is satisfied

$$\|g^k\|_{x^k} \geq \varrho. \tag{4.9}$$

Next, we consider the expression for $\beta_{k+1}^{\mathrm{MRHZ2}}$ from (2.11). This can be written as:

$$\beta_{k+1}^{\mathrm{MRHZ2}} = \frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\|g^k\|_{x^k}^2} - \min\left\{\frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\|g^k\|_{x^k}^2}, \mu_{k+1}^{\mathrm{MRHZ2}} \frac{\|y^k\|_{x^{k+1}}^2 \langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}}{\|g^k\|_{x^k}^4}\right\}.$$

We then estimate the norm of $\beta_{k+1}^{\mathrm{MRHZ2}}$:

$$\left\|\beta_{k+1}^{\mathrm{MRHZ2}}\right\| \leq \left|\frac{\langle g^{k+1}, y^k\rangle_{x^{k+1}}}{\|g^k\|_{x^k}^2}\right| + \left|\mu_{k+1}^{\mathrm{RHZ2}} \gamma_{k+1}^4\right|.$$

Now, applying the Cauchy–Schwarz inequality and the norm-decreasing property (2.4), we derive the following bounds:

$$|\langle g^{k+1}, y^k\rangle_{x^{k+1}}| \leq \|g^{k+1}\|_{x^{k+1}} \|y^k\|_{x^{k+1}},$$

and

$$|\langle g^{k+1}, \mathcal{T}_{\alpha_k\delta^k}(\delta^k)\rangle_{x^{k+1}}| \leq \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}.$$

Substituting these bounds into the inequality for $\left\|\beta_{k+1}^{\mathrm{MRHZ2}}\right\|$, we obtain:

$$\left\|\beta_{k+1}^{\mathrm{MRHZ2}}\right\| \leq \frac{\|g^{k+1}\|_{x^{k+1}} \|y^k\|_{x^{k+1}}}{\|g^k\|_{x^k}^2} + \mu_{k+1}^{\mathrm{RHZ2}} \frac{\|y^k\|_{x^{k+1}}^2 \|g^{k+1}\|_{x^{k+1}} \|\delta^k\|_{x^k}}{\|g^k\|_{x^k}^4}.$$

Now, setting the adaptive update rule (2.12) as (2.13):

$$\mu_{k+1}^{\text{RHZ2}} = \frac{\|g^k\|_{x^k}^4 - \|g^{k+1}\|_{x^{k+1}}\|y^k\|_{x^{k+1}}\|g^k\|_{x^k}^2\|\delta^k\|_{x^k}}{\|y^k\|_{x^{k+1}}^2\|g^{k+1}\|_{x^{k+1}}\|\delta^k\|_{x^k}^2}.$$

Substituting this into the previous inequality yields:

$$\left\|\beta_{k+1}^{\text{MRHZ2}}\right\| \leq \frac{1}{\|\delta^k\|_{x^k}}.$$

Using the norm-decreasing property from Eq (2.4) and the search direction update rule from Eq (2.7), we obtain the following bound for the search direction at iteration $k+1$:

$$\|\delta^{k+1}\|_{x^{k+1}} \leq \|g^{k+1}\|_{x^{k+1}} + |\beta_{k+1}^{\text{MRHZ2}}|\|\mathcal{T}_{\alpha_k\delta^k}(\delta^k)\|_{x^{k+1}}.$$

This can be simplified to:

$$\|\delta^{k+1}\|_{x^{k+1}} \leq \|g^{k+1}\|_{x^{k+1}} + 1 = \psi + 1 \triangleq \vartheta_2.$$

Now, let:

$$\vartheta_0 = \max\left\{\|\delta^k\|_{x^k} : 0 \leq k \leq k_0\right\}, \qquad \vartheta = \vartheta_0 + \vartheta_2.$$

Thus, we have the following bound for the search direction at all iterations:

$$\|\delta^k\|_{x^k} \leq \vartheta, \qquad \forall\ k.$$

From this bound and using Eq (4.9), we get the following inequality:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \geq \sum_{k=k_0}^{\infty} \frac{\varrho^4}{\vartheta^2} = +\infty.$$

Next, from the sufficient descent condition (3.2), we have:

$$\langle g^k, \delta^k\rangle_{x^k}^2 \geq \left(1 - \frac{1}{4}\right)^2 \|g^k\|_{x^k}^4.$$

Therefore, we can write:

$$\frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \leq \frac{1}{\left(1 - \frac{1}{4}\right)^2} \frac{\langle g^k, \delta^k\rangle_{x^k}^2}{\|\delta^k\|_{x^k}^2}.$$

Summing both sides over all iterations k, we obtain:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} \leq \frac{1}{\left(1 - \frac{1}{4}\right)^2} \sum_{k=0}^{\infty} \frac{\langle g^k, \delta^k\rangle_{x^k}^2}{\|\delta^k\|_{x^k}^2}.$$

By the bound on $\|\delta^k\|_{x^k}$ from Eq (4.4), the series on the right-hand side converges, so the sum on the left must also be finite:

$$\sum_{k=0}^{\infty} \frac{\|g^k\|_{x^k}^4}{\|\delta^k\|_{x^k}^2} < +\infty.$$

This leads to a contradiction with the previous divergence, which means our assumption that Eq (4.8) does not hold must be false. □

5. Numerical experiments

In this section, we present numerical experiments to assess the performance of the methods outlined in Algorithm 1. The two HZ-type Riemannian CG parameters, `MRHZ2` and `MRHZ1`, as defined in Eqs (2.8) and (2.11), respectively, are compared with several Riemannian CG algorithms. These include the `RFR` method [4], the `HHSDY` method [7], the `HFRPR` method [8], the `RHZ` method [8, 15], and the modified `RMIL` (Rivaie, Mustafa, Ismail, Leong) method by Salihu [17], referred to as `SMRMIL`. The implementation relies on the `Manopt` toolbox [31] and is executed on a laptop equipped with a 2.90 GHz Intel® Core i7-1195G7 CPU and 16 GB of RAM, with the codes used for the experiments are available at https://github.com/NasiruSalihu/A-Hager-Zhang-Riemannian-Conjugate-Gradient-Method-for-Matrix-Approximation/tree/main . The algorithm stops when the gradient norm reaches a tolerance of 10^{-6} or when the iteration count exceeds 2000. For the Wolfe conditions in (2.14) and (2.15), the parameters are set to $\sigma_1 = 10^{-4}$ and $\sigma_2 = 0.01$ for all methods tested.

The scaling factor for HZ-type methods in `MRHZ2` and `MRHZ1` is restricted to $\mu_{\min} = 1$ and $\mu_{\max} = 100$, using the following formula:

$$\mu_{k+1} = \min\left(\mu_{\max}, \max\left(\mu_{\min}, \rho_{k+1}\right)\right),$$

where ρ_{k+1} is computed using (2.10) and (2.13) for `MRHZ1` and `MRHZ2`, respectively.

5.1. Performance profile analysis

In this section, we utilize performance profiles [32] to objectively assess the efficiency of the algorithm across all test problems. For a given solver set S and a problem set P, the performance ratio $r_{p,s} = \frac{f_{p,s}}{\min_s f_{p,s}}$ quantifies the relative performance, while $\rho_s(\tau)$ indicates the proportion of problems where solver s performs within a factor τ of the best.

The evaluation is based on the following metrics:

- **Robustness**: The fraction of test problems where a CG method outperforms all other methods (i.e., "wins").
- **Success rate**: The percentage of problems where the method meets the stopping criterion within the given number of iterations, regardless of its performance ranking.

Each method is tested in 100 trials in all the Examples considered. Figures 1–4 show profiles for iterations and computational time, with $\rho_s(1)$ indicating efficiency (problems where solver s is best) and asymptotic $\rho_s(\tau)$ reflecting robustness and success rate, and summary results are given in Tables 1 and 2, reporting success and failure based on NI (number of iterations) and ET (elapsed time).

Observing Figures 1 and 2 based on the technique in [32], we summarize $\rho_s(\tau)$ (reflecting robustness and success rate) for both NI and ET in the following table for Example 2.1. This also applies for Figures 3 and 4 in Example 2.2.

In Example 2.1, the parameters are set to $n = 1000$, $p = 5$, $A = I_n \in \mathbb{R}^{n\times n}$ (the identity matrix), and $B = \frac{1}{\sqrt{n}}\mathbf{1}_{n\times p} \in \mathbb{R}^{n\times p}$ (the matrix of all ones normalized by $\frac{1}{\sqrt{n}}$). Figures 1, 2, and Table 1 reveal consistent performance trends in iteration and time metrics for Example 2.1. All methods successfully solve the 100 trials without failure, though with varying levels of robustness. The `RHHSDY` method achieves the highest robustness rate (68–71%), followed by `MRHZ2` with a robustness rate of 22–27%.

The remaining methods (SMRMIL, RHZ, HFRPR, RFR, and MRHZ1) demonstrate lower robustness rates, at 12–13%, 11–13%, 5–9%, 0–2%, and 0–2%, respectively. Comparing MRHZ2, with a robustness rate of 22–27%, to its restricted version RHZ, which has a robustness rate of 11–13%, we can conclude that our method outperforms the restricted version in this problem.

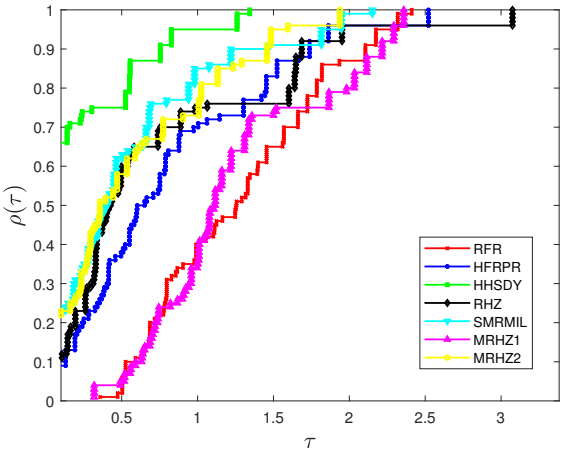

Figure 1. NI performance profile for Example 2.1.

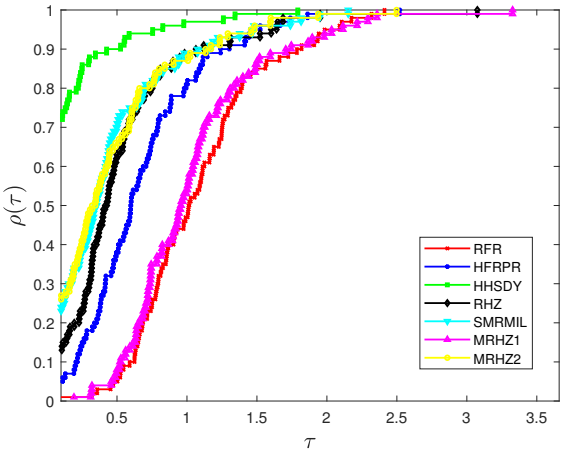

Figure 2. ET performance profile for Example 2.1.

Table 1. Robustness and success rates for Example 2.1.

	RFR	HFRPR	RHHSDY	RHZ	SMRMIL	MRHZ1	MRHZ2
Robustness	0-2%	5-9%	68-71%	11-13%	12-13%	0-2%	22- 27%
Success Rate	100%	100%	100%	100%	100%	100%	100%

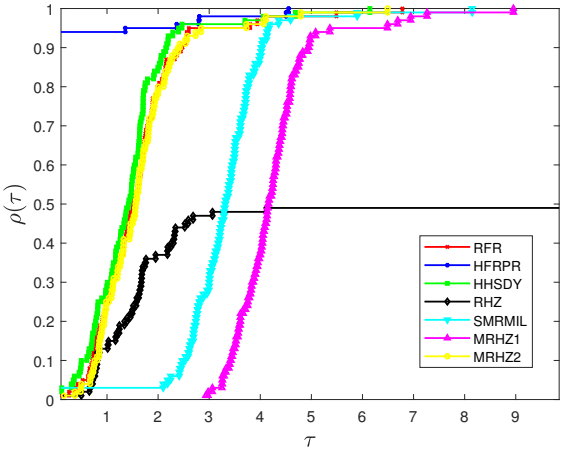

Figure 3. NI performance profile for Example 2.2.

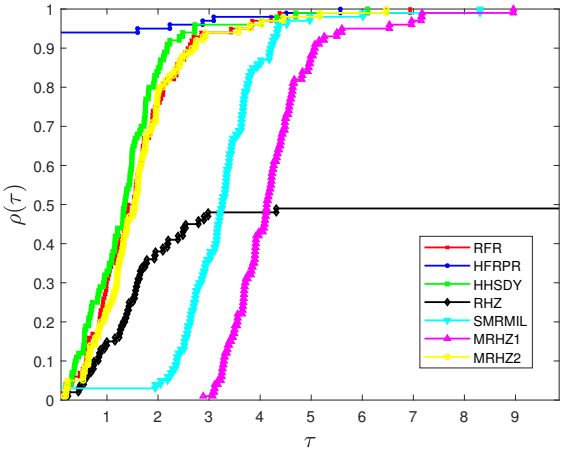

Figure 4. ET performance profile for Example 2.2.

Table 2. Robustness and success rates for Example 2.2.

	RFR	HFRPR	RHHSDY	RHZ	SMRMIL	MRHZ1	MRHZ2
Robustness	0-2%	95%	2-3%	0%	3%	0%	2%
Success Rate	100%	100%	100%	49%	100%	100%	100%

In Example 2.2, the matrix A is a diagonal matrix 1000×1000 with entries $1, 2, \ldots, 1000$, and n set to 1000. The Stiefel manifold is generated using the function stiefelfactory$(n, 1)$ from the `Manopt` toolbox, which creates a 1000-dimensional space of 1-dimensional orthonormal vectors.

Figures 3, 4, and Table 2 demonstrate consistent performance patterns in iteration count and computational time for Example 2.2, with all methods successfully solving all 100 trials except `RHZ`, which achieves a success rate of 49%. `HFRPR` emerges as the superior method, achieving the highest robustness (95%). In contrast, the remaining methods show limited effectiveness with low robustness (0–3%).

Comparing MRHZ1 and MRHZ2, which have a robustness rate of 0-2% and a success rate of 100%, to their restricted version RHZ, with a robustness rate of 0% and a success rate of 49% , we can conclude that our methods also outperform the restricted version in this problem.

6. Riemannian portfolio optimization

CG methods have proven effective in portfolio selection problems by optimizing allocations to minimize risk without explicitly maximizing returns [21–23]. Combining these strategies optimizes returns while controlling risk [33, 34]. Reformulating the problem on the sphere manifold eliminates interdependencies among portfolio weights, ensuring independence and maintaining constraints, extending the CG-based approach to a Riemannian framework.

To define the optimization procedure, let q represent the number of assets in the portfolio, with r_{it} denoting the return of asset i at time t:

$$r_{it} = \frac{P_t - P_{t-1}}{P_{t-1}},$$

where P_t and P_{t-1} are the asset prices at times t and $t-1$, respectively. Let $\mu \in \mathbb{R}^q$ be the vector of expected returns, and $\Sigma \in \mathbb{R}^{q\times q}$ the covariance matrix of asset returns.

The traditional Markowitz mean-variance model aims to minimize risk while maximizing expected return:

$$\min_{w\in\mathbb{R}^q} f(w) = \gamma w^\top \Sigma w - w^\top \mu, \tag{6.1}$$

subject to the full investment constraint $\sum_{i=1}^q w_i = 1$ and the nonnegativity constraints $w_i \geq 0$ for all asset weights $i = 1, \ldots, q$. Here, $\gamma > 0$ is the risk-aversion coefficient; the term $w^\top \Sigma w$ penalizes risk, while $-w^\top \mu$ incentivizes higher expected returns.

To reformulate this in a Riemannian framework, we introduce $x \in \mathbb{R}^q$ as an auxiliary variable and define the portfolio weights as:

$$w_i(x) = \frac{x_i^2}{\|x\|^2}, \quad i = 1, \ldots, q, \tag{6.2}$$

where $\|x\|^2 = x_1^2 + x_2^2 + \cdots + x_q^2$. This ensures the no-short-selling and full-investment constraints, given by:

$$w_i(x) \geq 0 \quad \text{and} \quad \sum_{i=1}^q w_i(x) = 1.$$

Since the portfolio weights depend only on the direction of x, we restrict x to the unit sphere $\|x\| = 1$, yielding:

$$w_i(x) = x_i^2, \quad \text{and} \quad \mathcal{M} = \mathbb{S}^{q-1} := \{x \in \mathbb{R}^q : \|x\|_2 = 1\}.$$

The optimization problem (6.1) is then reformulated in [35] as:

$$\min_{x\in\mathbb{S}^{q-1}} F(x) = \gamma w(x)^\top \Sigma w(x) - w(x)^\top \mu.$$

This reformulation enables smooth, constraint-free optimization on the sphere by embedding constraints within the geometry.

6.1. Results and analysis

This experiment applied Riemannian conjugate gradient (CG) methods to optimize a portfolio of 10 assets (AAPL, GOOGL, MSFT, AMZN, TSLA, META, NVDA, JPM, JNJ, and XOM) using data from January 4, 2021, to December 30, 2024 (1,003 trading days). The normalization history of these assets prior to optimization is shown in Figure 5, with data sourced from Yahoo Finance (http://finance.yahoo.com/). Optimization is performed on the sphere manifold, enforcing non-negativity of portfolio weights. The methods tested include `RFR`, `HFRPR`, `HHSDY`, `RHZ`, `SMRMIL`, `MRHZ1`, and `MRHZ2`, with convergence and efficiency evaluated for each.

The risk-aversion parameter γ is adaptively computed as:

$$\gamma = \begin{cases} \frac{\text{avg volatility}}{2\times\text{avg return}} & \text{if avg return} > 0, \\ 1.0 & \text{if avg return} \leq 0. \end{cases}$$

This reflects risk (volatility) and reward (return), with a default $\gamma = 1.0$ when the average return is non-positive. The risk-return contributions are computed as:

- **Risk contribution**: $\mathrm{RC}_i = \frac{w_i \cdot (\Sigma w)_i}{\sigma_p^2}$.
- **Return contribution**: $\mathrm{RetC}_i = w_i \max(\mu_i, 0.10)$,

where σ_p^2 is the portfolio variance.

All methods meet the gradient norm tolerance of 10^{-6}. `MRHZ1` converges the fastest, requiring 24 iterations and 0.0073 seconds, while `SMRMIL` is the slowest, taking 101 iterations and 0.189 seconds. These results are summarized in Table 3 and illustrated in Figure 5 (Optimization method).

Table 3. Convergence performance of various Riemannian CG methods.

Method	Iterations	Time (s)	Remarks
`RFR`	57	0.083	Slow convergence
`HFRPR`	42	0.068	Moderate convergence
`HHSDY`	43	0.0674	Moderate convergence
`RHZ`	40	0.055	Fast convergence
`SMRMIL`	101	0.189	Slowest convergence
`MRHZ1`	24	0.0073	Fastest convergence
`MRHZ2`	41	0.073	Fast convergence

6.2. Portfolio allocation and risk analysis

The portfolio achieves a 9.40% annual return, 16.90% risk, and a Sharpe ratio of 0.556 across all methods. Diversification is reflected by a Herfindahl-Hirschman Index (HHI) of 0.232 and an Effective Number of Assets (EffN) of 4.3. The portfolio exhibits strong concentration, with 27.53% allocated to TSLA, 25.31% to NVDA, and 26.11% to XOM, together accounting for nearly 79% of the total portfolio. These results are illustrated in Table 4 and Figure 5, showing the portfolio weights (bar graph) and portfolio allocation (pie graph), respectively.

Table 4. Optimal portfolio allocation.

Ticker	Weight	Expected return	Return contribution
AAPL	13.44%	10.00%	1.34%
MSFT	0.24%	10.00%	0.02%
AMZN	7.36%	1.87%	0.14%
TSLA	27.53%	10.00%	2.75%
NVDA	25.31%	10.00%	2.53%
XOM	26.11%	10.00%	2.61%
Total	100.00%		9.40%

The portfolio achieves a 9.40% return, concentrated in TSLA, NVDA, and XOM, with high return contributions from TSLA (2.75%) and XOM (2.61%).

6.3. Sensitivity analysis

- The marginal risk for each asset is consistent across methods, with `TSLA`, `NVDA`, and `XOM` contributing the highest individual risk (0.178).
- The sensitivity analysis shows that the portfolio's expected return and risk are consistent across different methods, with the Max sharpe approach providing the best balance (9.40% return, 16.90% risk).

Figure 5 reflects the overall optimization result of the portfolio optimization using manifolds in MATLAB with the same specifications as in Examples 2.1–2.2.

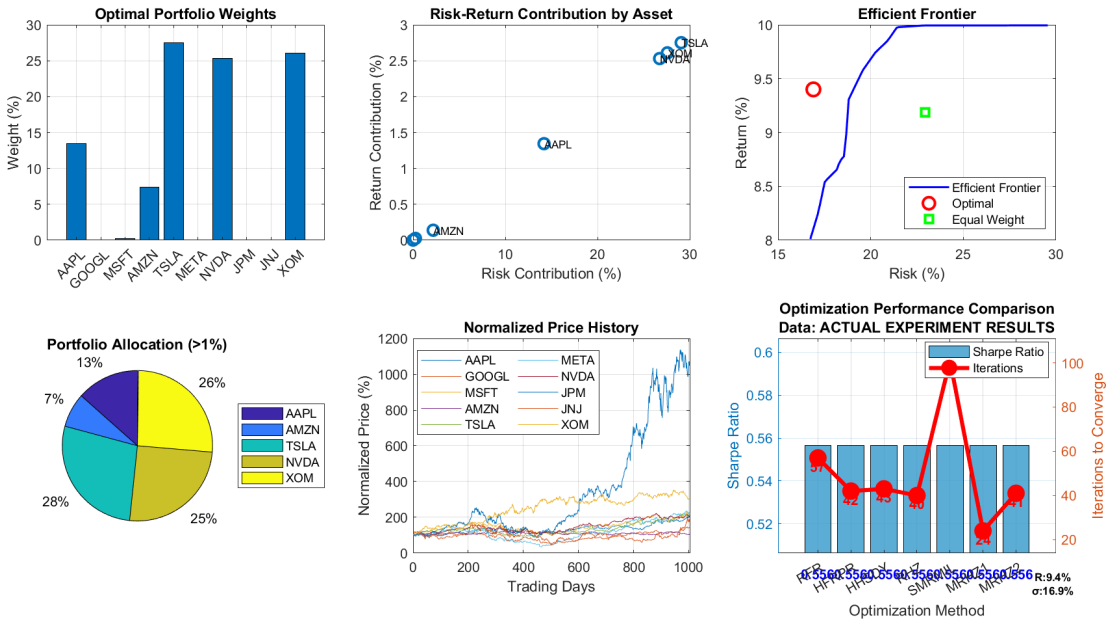

Figure 5. Evaluation of the methods' performance.

Table 5 demonstrates the effectiveness of Riemannian optimization for portfolio selection, with all methods converging to the same optimal portfolio. `MRHZ1` and `MRHZ2` are the most efficient, offering

fast convergence and stability. The portfolio shows consistent return, risk, and Sharpe ratio across all methods, indicating robustness. The sphere manifold formulation captures the Markowitz optimum, focusing on assets with the best risk-return characteristics. Moreover, while institutional constraints can be added, the geometric approach keeps the core optimization clean and separate from practical considerations.

Table 5. Performance comparison with benchmark strategies. The Sharpe ratio indicates risk-adjusted return; higher values correspond to better return per unit of risk. The Max Sharpe strategy provides the best balance between expected return and portfolio risk. HHI and Effective N reflect diversification.

Strategy	Return	Risk	Sharpe	HHI	Effective N
Equal Weight	0.0919	0.2293	0.4007	0.1000	10.0
Max Sharpe	0.0940	0.1690	0.5564	0.2316	4.32
Min Variance	0.0527	0.1324	0.3976	0.3845	2.60

7. Conclusions

In this work, we build upon and significantly extend the research outlined in [15], which introduced an exponential retraction-based Riemannian HZ method. We deliver this generalization and push the boundaries in several novel directions. Our major contributions are threefold. First, we introduce a new family of HZ-type conjugate gradient methods, namely `MRHZ1` and `MRHZ2`, within a fully general retraction-based framework. Our approach relaxes the need for strong convexity by avoiding reliance on the Gauss lemma, and we establish their global convergence for non-convex problems through a norm-decreasing property. Second, to enhance practical stability, we propose a robust *min(max)* parameter regulation strategy, which guarantees sufficient descent without requiring strong Wolfe conditions. This simplifies implementation and improves the reliability of the methods. Third, we demonstrate the practical utility of our framework through a novel application: Riemannian portfolio optimization on the sphere. By reformulating the problem geometrically, we naturally enforce simplex constraints and extend Euclidean methods. Our numerical experiments confirm that all Riemannian CG variants converge to a consistent, efficient portfolio, with `MRHZ1`, `MRHZ2`, and `RHZ` being the fastest and most stable. Their robustness is further underscored by an identical optimal Sharpe ratio in all algorithms. In conclusion, our work not only realizes the theoretical promise of a general Riemannian HZ method, but also establishes a new, practical state-of-the-art for complex optimization tasks, such as constrained portfolio selection.

Author contributions

Nasiru Salihu: Conceptualization, methodology, coding, writing-original draft, visualization and investigation, editing the manuscript; Seyed Yaser Mousavi Siamakani: Supervision, validating the experiment; Auwal Bala Abubakar: Validating the experiment, editing the manuscript; Also Mohammed Saleh: Supervision. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that no generative-AI tools were used in the preparation of this manuscript.

Acknowledgments

This research was supported by Department of Mathematics and Applied Mathematics of Sefako Makgatho Health Sciences University, Ga-Rankuwa, Pretoria, Medunsa, South Africa.

Conflict of interest

The authors declare no conflicts of interest.

References

1. P. A. Absil, R. Mahony, R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton: Princeton University Press, 2009. https://doi.org/10.1515/9781400830244
2. N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge: Cambridge University Press, 2023. https://doi.org/10.1017/9781009166164
3. N. Andrei, *Modern numerical nonlinear optimization*, Cham: Springer, 2022. https://doi.org/10.1007/978-3-031-08720-2
4. W. Ring, B. Wirth, Optimization methods on Riemannian manifolds and their application to shape space, *SIAM J. Optim.*, **22** (2012), 596–627. https://doi.org/10.1137/11082885X
5. H. Sato, T. Iwai, A new, globally convergent Riemannian conjugate gradient method, *Optimization*, **64** (2015), 1011–1031. https://doi.org/10.1080/02331934.2013.836650
6. Y. H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.*, **10** (1999), 177–182. https://doi.org/10.1137/S1052623497318992
7. H. Sakai, H. Iiduka, Hybrid Riemannian conjugate gradient methods with global convergence properties, *Comput. Optim. Appl.*, **77** (2020), 811–830. https://doi.org/10.1007/s10589-020-00224-9
8. H. Sakai, H. Iiduka, Sufficient descent Riemannian conjugate gradient methods, *J. Optim. Theory Appl.*, **190** (2021), 130–150. https://doi.org/10.1007/s10957-021-01874-3
9. C. Tang, W. Tan, S. Xing, H. Zheng, A class of spectral conjugate gradient methods for Riemannian optimization, *Numer. Algor.*, **94** (2023), 131–147. https://doi.org/10.1007/s11075-022-01495-5
10. S. Nasiru, P. Kumam, S. Salisu, L. Wang, T. Seangwattana, Spectral conjugate gradient for Riemannian optimization: Application to the Gough–Stewart platform, *Int. J. Dyn. Control*, **13** (2025), 288. https://doi.org/10.1007/s40435-025-01788-2
11. N. Salihu, P. Kumam, S. M. Ibrahim, W. Kumam, Some combined techniques of spectral conjugate gradient methods with applications to robotic and image restoration models, *Numer. Algor.*, **100** (2025), 553–593. https://doi.org/10.1007/s11075-024-01970-1

12. W. W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.*, **16** (2005), 170–192. https://doi.org/10.1137/030601880

13. W. W. Hager, H. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, *ACM Trans. Math. Software*, **32** (2006), 113–137. https://doi.org/10.1145/1132973.1132979

14. M. Rivaie, M. Mamat, L. W. June, I. Mohd, A new class of nonlinear conjugate gradient coefficients with global convergence properties, *Appl. Math. Comput.*, **218** (2012), 11323–11332. https://doi.org/10.1016/j.amc.2012.05.030

15. H. Sakai, H. Sato, H. Iiduka, Global convergence of Hager–Zhang type Riemannian conjugate gradient method, *Appl. Math. Comput.*, **441** (2023), 127685. https://doi.org/10.1016/j.amc.2022.127685

16. N. Salihu, P. Kumam, S. Salisu, Two efficient nonlinear conjugate gradient methods for Riemannian manifolds, *Comput. Appl. Math.*, **43** (2024), 415. https://doi.org/10.1007/s40314-024-02920-2

17. N. Salihu, P. Kumam, S. Salisu, L. Wang, K. Sitthithakerngkiet, A revised MRMIL Riemannian conjugate gradient method with simplified global convergence properties, *Appl. Numer. Math.*, **214** (2025), 110–126. https://doi.org/10.1016/j.apnum.2025.03.007

18. N. Salihu, P. Kumam, I. M. Sulaiman, I. Arzuka, W. Kumam, An efficient Newton-like conjugate gradient method with restart strategy and its application, *Math. Comput. Simulat.*, **226** (2024), 354–372. https://doi.org/10.1016/j.matcom.2024.07.008

19. N. Salihu, P. Kumam, L. Wang, M. M. Yahaya, An improved Dai–Kou conjugate gradient method with spectral search direction and applications, *Eng. Optim.*, **58** (2026), 547–584. https://doi.org/10.1080/0305215X.2025.2475004

20. X. Wang, A class of spectral three-term descent Hestenes–Stiefel conjugate gradient algorithms for large-scale unconstrained optimization and image restoration problems, *Appl. Numer. Math.*, **192** (2023), 41–56. https://doi.org/10.1016/j.apnum.2023.05.024

21. A. M. Awwal, I. M. Sulaiman, M. Malik, M. Mamat, P. Kumam, K. Sitthithakerngkiet, A spectral RMIL+ conjugate gradient method for unconstrained optimization with applications in portfolio selection and motion control, *IEEE Access*, **9** (2021), 75398–75414. https://doi.org/10.1109/ACCESS.2021.3081570

22. M. Malik, I. M. Sulaiman, A. B. Abubakar, G. Ardaneswari, Sukono, A new family of hybrid three-term conjugate gradient method for unconstrained optimization with application to image restoration and portfolio selection, *AIMS Mathematics*, **8** (2023), 1–28. https://doi.org/10.3934/math.2023001

23. N. Salihu, S. M. Ibrahim, P. Kaelo, I. A. Moghrabi, E. N. Madi, A spectral Fletcher–Reeves conjugate gradient method with integrated strategy for unconstrained optimization and portfolio selection, *PloS One*, **20** (2025), e0320416. https://doi.org/10.1371/journal.pone.0320416

24. X. Zhu, A Riemannian conjugate gradient method for optimization on the Stiefel manifold, *Comput. Optim. Appl.*, **67** (2017), 73–110. https://doi.org/10.1007/s10589-016-9883-4

25. B. Vandereycken, Low-rank matrix completion by Riemannian optimization, *SIAM J. Optim.*, **23** (2013), 1214–1236. https://doi.org/10.1137/110845768
26. H. Sakai, H. Iiduka, Riemannian adaptive optimization algorithm and its application to natural language processing, *IEEE Trans. Cybern.*, **52** (2021), 7328–7339. https://doi.org/10.1109/TCYB.2021.3049845
27. B. McCann, M. Nazari, Riemannian optimization techniques on Lie groups describing rigid body attitude and pose, *J. Optim. Theory Appl.*, **207** (2025), 63. https://doi.org/10.1007/s10957-025-02810-5
28. H. Sato, Riemannian conjugate gradient methods: General framework and specific algorithms with convergence analyses, *SIAM J. Optim.*, **32** (2022), 2690–2717. https://doi.org/10.1137/21M1464178
29. J. K. Liu, S. J. Li, New hybrid conjugate gradient method for unconstrained optimization, *Appl. Math. Comput.*, **245** (2014), 36–43. https://doi.org/10.1016/j.amc.2014.07.096
30. J. Nocedal, S. J. Wright, *Numerical optimization*, 2 Eds., New York: Springer, 2006. https://doi.org/10.1007/978-0-387-40065-5
31. N. Boumal, B. Mishra, P. A. Absil, R. Sepulchre, Manopt: A MATLAB toolbox for optimization on manifolds, *J. Mach. Learn. Res.*, **15** (2014), 1455–1459.
32. E. D. Dolan, J. J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.*, **91** (2002), 201–213. https://doi.org/10.1007/s101070100263
33. M. Bartholomew-Biggs, *Nonlinear optimization with financial applications*, New York: Springer, 2006. https://doi.org/10.1007/b102601
34. H. Markowitz, Portfolio selection, *J. Finance*, **7** (1952), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
35. N. Salihu, P. Kumam, K. Sitthithakerngkiet, T. Seangwattana, A Riemannian conjugate gradient framework based on Dai–Liao principles for matrix approximation, *IEEE Access*, **13** (2025), 214813–214834. https://doi.org/10.1109/ACCESS.2025.3645695

AIMS Press