arXiv:2603.04078v1 [math.OC] 4 Mar 2026

文章

带动量的黎曼梯度法

Filippo Leggio[a] 和Diego Scuppa[a]

[a]罗马第一大学计算机、控制与管理工程系，罗马，意大利

文章历史
编制于2026年3月5日

摘要
本文考虑在黎曼流形上最小化光滑函数的问题，并提出一种带动量的黎曼梯度法。所提出的算法是对最近提出的一种无约束优化方法的实质性且非平凡的扩展。我们证明，在保护规则的支持下，该算法可产生一个 ϵ-驻点，其最坏情况复杂度界为 $\mathcal{O}(\epsilon^{-2})$。我们在基准问题上进行了大量计算实验，将所提出的方法与 Manopt 软件包中现有的先进求解器进行比较。结果表明，该方法具有竞争力，且往往表现更优。总体而言，数值证据证实了所提出方法的有效性与稳健性，它将最近提出的基于动量的方法有意义地扩展到了黎曼优化。

关键词
黎曼流形；黎曼优化；非凸优化；非线性优化；动量；全局收敛；复杂度

AMS 分类90C26；90C30；
90C53；90C60

1. 引言

在本文中，我们关注以下问题

$$\min\{f(x) : x \in \mathcal{M}\},$$

其中， $\mathcal{M}$ 是有限维欧几里得空间的黎曼子流形 $\mathcal{E}$[3, 9, 13, 26], ，而 f： $\mathcal{M} \to \mathbb{R}$ 是一个光滑的非凸函数。

黎曼流形经常出现在实际应用中，例如机器学习 [21, 29], 雷达通信 [22],低秩矩阵（及张量）补全[6, 17, 30], 不变子空间计算 [2], 半正定规划 [20], 形状空间分析 [24],以及特殊欧氏群上的同步 [25]。

欧氏空间中的标准无约束优化有几种著名方法已被推广到黎曼设置中 $\mathcal{E} = \mathbb{R}^n$ [5, 14] ，例如共轭梯度法 [6, 27, 30], 信赖域方法 [1, 6, 8],Barzilai–Borwein

联系人：Diego Scuppa。电子邮箱：diego.scuppa@uniroma1.it

方法 [16], 和 L-BFGS 方法 [15]。文献已对这些方法的收敛性进行了广泛分析，其中多项研究给出了全局收敛速率。例如，在满足关于 f的 Lipschitz 型假设时， [10] 给出了黎曼梯度下降法和黎曼信赖域方法的最坏情况复杂度界。最近，在 [32]中证明，采用非单调线搜索时，黎曼共轭梯度法能够以 $\mathcal{O}(\epsilon^{-2})$的迭代复杂度产生一个 ϵ-驻点；该方法还结合了一种重启策略：当某些条件不满足时，以经 Barzilai–Borwein 因子缩放的负黎曼梯度替换计算得到的方向 [4, 23], 。这些条件可视为对 [11]中为无约束优化提出的梯度相关 条件的推广。

在这项工作中，我们研究带动量的黎曼梯度方法，即更新规则将黎曼梯度与前一搜索方向的适当贡献相结合的迭代方案。

带动量的黎曼梯度方法最近已在 [28]中得到研究。该文提出的算法在每次迭代中，求解一个限制在适当选取的切空间子空间上的三次正则化子问题，并结合（可能非单调的）线搜索过程。在涉及二阶信息的温和假设下，作者建立了最坏情况下的迭代复杂度界 $\mathcal{O}(\epsilon^{-2})$。

受近期论文 [18]和 [19], 的启发，我们在本文中提出一种基于线搜索的方法。在每次迭代中，通过求解一个二次二维子问题来获得搜索方向，该子问题涉及切空间中两个方向的线性组合。所提方法的核心在于构造二次子问题时，必须考虑可行集为黎曼流形这一事实所带来的技术和计算问题。在标准假设下，我们证明，借助适当的重启策略，该算法能够生成一个 ϵ-驻点，迭代复杂度为 $\mathcal{O}(\epsilon^{-2})$。

本文的主要贡献是开发了一种用于流形优化的新型一阶算法，该算法具有严格的全局收敛保证和有竞争力的计算复杂度。大量数值实验验证了该算法的实际有效性，结果表明其性能与当前最先进的方法相当，甚至更优。

本文其余部分安排如下。第 2 节概述黎曼流形和黎曼优化的基本概念。第 3 节介绍一种基于梯度相关方向的一般方法的收敛性结果。第 4 节总结 [18], 中所述的方法，其黎曼适配版本将在第 5 节讨论。第 6 节介绍一些算法方面的考量，并提出所述的收敛算法。第 7 节讨论数值结果，第 8 节以总结性评述结束全文。

2. 黎曼流形上的初步技术结果

本节简要回顾黎曼优化中的一些基本概念；例如，参见 [9]。设 $\mathcal{E}$ 为一个维数为 n 的有限维欧几里得空间，配备内积 $\langle \cdot, \cdot \rangle$。典型示例包括 $\mathcal{E} = \mathbb{R}^n$，其具有 $\langle x, y \rangle = x^\top y$；或 $\mathcal{E} = \mathbb{R}^{n_1 \times n_2}$，其具有 $\langle X, Y \rangle = \mathrm{tr}(X^\top Y)$。光滑嵌入流形 $\mathcal{M}$ 是 $\mathcal{E}$ 的一个子集，并且对于每个点 $x \in \mathcal{M}$，都存在 $\mathcal{M}$ 在该点处的局部线性近似

x，称为切空间，记作 $\mathrm{T}_x\mathcal{M}$。若 $\mathcal{M}$ 定义为

$$\mathcal{M} = \{x \in \mathcal{E} : h_1(x) = 0, \ldots, h_m(x) = 0\}, \tag{1}$$

其中，$m < n$ 和 $h_1, \ldots, h_m$：$\mathcal{E} \to \mathbb{R}$ 是满足线性无关约束规范条件（LICQ）的光滑函数，则 x 处的切空间为 $\mathrm{T}_x\mathcal{M} = \{v \in \mathcal{E}: \langle \nabla h_i(x)\ v\rangle = 0,\ i = 1, \ldots, m\}$。

每个切空间 $\mathrm{T}_x\mathcal{M}$ 都可以赋予从$\mathcal{E}$继承的内积。采用这种度量后，$\mathcal{M}$称为 $\mathcal{E}$的黎曼子流形 。

设 f：$\mathcal{M} \to \mathbb{R}$ 是一个连续可微函数。f在 $x \in \mathcal{M}$处的黎曼梯度 定义为唯一向量 $\mathrm{grad}\ f(x) \in \mathrm{T}_x\mathcal{M}$，使得对所有 $v \in \mathrm{T}_x\mathcal{M}$，都有$\langle \mathrm{grad}\ f(x)\ v\rangle = \mathrm{D}f(x)[v]$。可以证明，$\mathrm{grad}\ f(x) = \mathrm{proj}_x(\nabla f(x))$，其中 $\nabla f(x)$表示 f 的欧几里得梯度，而 proj_x：$\mathcal{E} \to \mathrm{T}_x\mathcal{M}$是到切空间上的正交投影。

f 在 x 处的黎曼黑塞矩阵定义为自伴线性算子 $\mathrm{Hess}\ f(x) : \mathrm{T}_x\mathcal{M} \to \mathrm{T}_x\mathcal{M}$，其定义为 $\mathrm{Hess}\ f(x)[v] = \mathrm{proj}_x(\mathrm{D}\ \mathrm{grad}\ f(x)[v])$。

对于每个 $x \in \mathcal{M}$，回缩是一个光滑映射 $\mathrm{R}_x : \mathrm{T}_x\mathcal{M} \to \mathcal{M}$，满足 $\mathrm{R}_x(0) = x$ 且 $\mathrm{D}\ \mathrm{R}_x(0) = \mathrm{Id}$。

此外，对于任意 x，$y \in \mathcal{M}$，向量传输是一个线性映射 $\mathrm{T}_{y\leftarrow x} : \mathrm{T}_x\mathcal{M} \to \mathrm{T}_y\mathcal{M}$，满足 $\mathrm{T}_{x\leftarrow x}$ 在 $\mathrm{T}_x\mathcal{M}$上为恒等映射。就我们的目的而言，一种实用的选择可以是到 $\mathrm{T}_y\mathcal{M}$上的正交投影，并将其限制在 $\mathrm{T}_x\mathcal{M}$上，即 $\mathrm{T}_{y\leftarrow x} := \mathrm{proj}_{y|\mathrm{T}_x\mathcal{M}}$。

黎曼优化算法旨在最小化 f 上的光滑函数 $\mathcal{M}$。对 f进行一阶泰勒展开可知，点 $x^* \in \mathcal{M}$ 为局部极小点的必要条件是其为驻点，即 $\mathrm{grad}\ f(x^*) = 0$。此外，如果 R_x 是二阶回缩且 f 是二阶连续可微的，则二阶泰勒展开给出条件 $\langle v, \mathrm{Hess}\ f(x^*)[v]\rangle \geq 0$ 对所有$v \in \mathrm{T}_{x^*}\mathcal{M}$均成立。当 $\mathcal{M}$ 按（1）定义时，这些条件等价于一阶和二阶 Karush-Kuhn-Tucker 最优性条件。

3. 一个具有收敛保证的一阶框架

本节介绍一种具有收敛性保证的通用基于线搜索的框架。理论结果依赖于采用一系列与梯度相关的搜索方向以及 Armijo 线搜索策略。

我们注意到，本节的结果可由 [32]中的更一般分析推得。为完整起见，我们在此给出定理陈述，证明则见附录 A。

设 $\mathcal{M}$ 是有限维欧几里得空间$\mathcal{E}$中的一个光滑黎曼子流形，且 f：$\mathcal{M} \to \mathbb{R}$ 是一个连续可微函数。我们不假设 f是凸的；但假设 f 有下界：

假设 1. 存在 $f_{\mathrm{low}} \in \mathbb{R}$，使得对于所有 $x \in \mathcal{M}$，均有 $f(x) \geq f_{\mathrm{low}}$。

考虑以下优化问题：

$$\min_{x\in\mathcal{M}} f(x).$$

一般的一阶黎曼算法会生成一个点序列 $\{x_k\} \subset \mathcal{M}$

该序列可能收敛于某个驻点 $x^* \in \mathcal{M}$，即满足

$$\operatorname{grad} f(x^*) = 0.$$

一般更新规则可概括如下：给定点 $x_k \in \mathcal{M}$，寻找方向 $d_k \in \mathrm{T}_{x_k}\mathcal{M}$ 和步长 $\eta_k > 0$（通过线搜索），并计算 $\mathcal{M}$ 中的新点：

$$x_{k+1} = \mathrm{R}_{x_k}(\eta_k d_k),$$

其中，R_{x_k} 是一个回缩。若以下假设成立，则称方向序列 $\{d_k\}$ 与序列 $\{x_k\}$梯度相关。

假设 2。 存在常数 $0 < c_1 \le c_2$，使得对于所有 k,

$$\langle \operatorname{grad} f(x_k), d_k \rangle \le -c_1 \| \operatorname{grad} f(x_k) \|^2, \tag{2a}$$

$$\| d_k \| \le c_2 \| \operatorname{grad} f(x_k) \|. \tag{2b}$$

注意，梯度相关方向这一概念扩展了文献 [11] 中针对无约束优化所引入的相应概念。此外，我们作出如下假设。

假设 3。 存在 $L > 0$，并且对于所有 k，存在 $\rho_k > \|d_k\|$ （可能对所有 k均为$\rho_k = \infty$），使得：

（1）回缩 R_{x_k} 至少在球 $B(\rho_k) \cap \mathrm{T}_{x_k}\mathcal{M}$ 上有定义；（2）对于所有 $d \in B(\rho_k) \cap \mathrm{T}_{x_k}\mathcal{M}$,

$$|f(\mathrm{R}_{x_k}(d)) - f(x_k) - \langle \operatorname{grad} f(x_k), d \rangle| \le \frac{L}{2}\|d\|^2. \tag{3}$$

我们注意到，在黎曼优化环境中，可以为黎曼梯度引入利普希茨连续性的概念。然而，正如 [10], 中所讨论的，直接假设利普希茨型条件 (3) 更为方便。

现在我们讨论一种一般一阶算法的理论性质，其中搜索方向序列 $\{d_k\}$与迭代点 $\{x_k\}$梯度相关，如算法 1 所述。此外，我们假设采用初始步长等于 1 的 Armijo线搜索过程（见算法 2）。

命题 3.1。假设假设 *1–2–3* 得到满足（其常数为$f_{\mathrm{low}} \in \mathbb{R}$、 $0 < c_1 \le c_2$和 $L > 0$）。令 $\delta \in (0, 1)$ 且 $\gamma \in (0,1)$。令 $\{x_k\}$ 和$\{d_k\}$ 为算法 *1* 生成的序列。假设 d_k 的计算无需额外的函数或梯度评估。对于任意 $\epsilon > 0$，定义：

- k_ϵ：使得 x_{k_ϵ} 满足 $\| \operatorname{grad} f(x_{k_\epsilon})\| \le \epsilon$的第一次迭代；
- nf_ϵ：计算x_{k_ϵ}所需的函数评估（及回缩）次数；
- ni_ϵ：迭代 k 的总数，使得 $\| \operatorname{grad} f(x_k)\| > \epsilon$。

那么，在最坏情况下，算法 *1* 保证：

算法 1：具有梯度相关方向的一阶算法 输入： $x_0 \in \mathcal{M}$， $\gamma \in (0,1)$， $\delta \in (0,1)$， $\epsilon > 0$， $0 < c_1 \le c_2$输出：近似驻点 x_k**1** $k \leftarrow 0$；**2** 当 $\|$梯度 $f(x_k)\| > \epsilon$执行 **3** 选择 $d_k \in \mathrm{T}_{x_k}\mathcal{M}$使得 $\langle$梯度 $f(x_k)$ $d_k\rangle \le -c_1\|$梯度 $f(x_k)\|^2$且 $\|d_k\| \le c_2\|$梯度 $f(x_k)\|$；**4** $\eta_k \leftarrow$Armijo(x_k，$f(x_k)$，梯度 $f(x_k)$ d_k，1，γ，δ)；**5** $x_{k+1} \leftarrow R_{x_k}(\eta_k d_k)$；**6** $k \leftarrow k+1$；

算法 2： Armijo 线搜索

1 函数 Armijo(x_k, $f(x_k)$, 梯度 $f(x_k)$ d_k, η_k, γ, δ)：2 当 $f\big(R_{x_k}(\eta_k d_k)\big) > f(x_k) + \gamma\,\eta_k\,\langle$**梯度** $f(x_k)$ $d_k\rangle$ **执行**
3 $\eta_k \leftarrow \delta\,\eta_k$**；4 返回** η_k**；**

(1) 若 $L \le 2(1-\gamma)\frac{c_1}{c_2^2}$，则

$$k_\epsilon = nf_\epsilon \le \frac{f(x_0) - f_{\text{low}}}{\gamma c_1}\,\epsilon^{-2} = \mathcal{O}(\epsilon^{-2}); \tag{4}$$

(2) 否则，

$$k_\epsilon \le \frac{Lc_2^2\big(f(x_0) - f_{\text{low}}\big)}{2\delta\gamma(1-\gamma)c_1^2}\,\epsilon^{-2} = \mathcal{O}(\epsilon^{-2}), \tag{5a}$$

$$nf_\epsilon \le \frac{Lc_2^2\big(f(x_0) - f_{\text{low}}\big)}{2\delta\gamma(1-\gamma)c_1^2}\left(\left\lfloor \log_\delta\left(\frac{2}{L}(1-\gamma)\frac{c_1}{c_2^2}\right)\right\rfloor + 2\right)\epsilon^{-2} = \mathcal{O}(\epsilon^{-2}). \tag{5b}$$

此外， ni_ϵ 满足与(4)–(5a)中相同的界。

命题 3.1 的证明见附录 A。

我们注意到，本研究采用的假设比 [28,假设 2.1]，中的假设更宽松；作者在该文中针对一种实用算法，在涉及二阶信息的假设下证明了其最坏情况复杂度为 $\mathcal{O}(\epsilon^{-2})$。

4. 一种无约束动量梯度法

在本节中，我们简要总结 [18] 中针对欧几里得空间 $\mathbb{R}^n$提出的动量梯度法。设 $f: \mathbb{R}^n \to \mathbb{R}$ 为连续可微函数，不一定是凸的，并考虑无约束优化问题

$$\min_{x \in \mathbb{R}^n} f(x).$$

该方法生成一列迭代点 $\{x_k\}$，全局收敛到 f的驻点。给定当前迭代点 $x_k \in \mathbb{R}^n$，下一个迭代点计算为

$$x_{k+1} = x_k + \eta_k d_k,$$

其中 $\eta_k > 0$ 由Armijo线搜索过程确定，搜索方向 d_k 定义为

$$d_k = -\alpha_k g_k + \beta_k s_k,$$

其中 $g_k = \nabla f(x_k)$且 $s_k = x_k - x_{k-1}$。系数 α_k 和 β_k 通过最小化一个在 x_k附近近似函数 f 的二次模型得到，即

$$\begin{aligned} \min_{d,\alpha,\beta} \ & f(x_k) + g_k^\top d + \frac{1}{2} d^\top B_k d \\ \text{s.t. } & d = -\alpha g_k + \beta s_k, \end{aligned}$$

其中 $B_k \in \mathbb{R}^{n\times n}$ 是一个对称正定矩阵。该问题等价于最小化如下二维二次模型:

$$\min_{u\in\mathbb{R}^2} \ T_k^\top u + \frac{1}{2} u^\top H_k u,$$

其中

$$T_k = \begin{bmatrix} -\|g_k\|^2 \\ g_k^\top s_k \end{bmatrix}, \qquad u = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

and

$$H_k = \begin{bmatrix} g_k^\top B_k g_k & -g_k^\top B_k s_k \\ -g_k^\top B_k s_k & s_k^\top B_k s_k \end{bmatrix}. \tag{6}$$

[18] 的作者提出了两种备选策略:

- 按(6)式由 $n \times n$ 矩阵 B_k 定义 2×2 矩阵 H_k；
- 直接选择任意对称 2×2 矩阵 H_k，而不依赖于 B_k。

通过要求搜索方向 $\{d_k\}$满足梯度相关条件，即，建立了这两种方法的收敛条件

$$g_k^\top d_k \le -c_1 \|g_k\|^2, \qquad \|d_k\| \le c_2 \|g_k\|, \quad \text{for all } k,$$

对于某些常数 $c_1, c_2 > 0$。特别地，结果表明，若水平集 $\mathcal{L}_0 := \{x \in \mathbb{R}^n : f(x) \le f(x_0)\}$ 是紧集，则该方法要么在有限次迭代内得到 f 的一个驻点 x_ν，要么生成一个无限序列 $\{x_k\}$，该序列存在聚点，且所有聚点都是 f的驻点。此外，若 ∇f 在 $\mathbb{R}^n$上利普希茨连续，则该方法至多经过 $\mathcal{O}(\epsilon^{-2})$次迭代、函数评估和梯度评估，就能找到一个点 x_k，使得 $\|\nabla f(x_k)\| \le \epsilon$。

一个关键的实际问题涉及矩阵 B_k 的选择，或等价地涉及 H_k 的选择。[18]，中讨论了若干方案，包括三个元素的计算

即通过在三个不同点强制二次模型满足插值条件来确定 H_k 的这些元素。然而，该方法需要对目标函数进行额外求值。因此，它不适合推广到流形上的优化问题，因为这将涉及可能计算代价高昂的回缩操作。我们为克服这一问题所提出的策略将在后文给出。

5. 推广到黎曼情形所面临的问题

本节旨在将动量梯度法推广到黎曼情形，并讨论这一推广过程中出现的技术问题。

在此框架下，给定一点 $x_k \in \mathcal{M}$，新点按 $x_{k+1} = \mathrm{R}_{x_k}(\eta_k d_k)$ 生成，其中 $\eta_k > 0$ 由 Armijo 线搜索过程确定，而 d_k 是属于切空间 $\mathrm{T}_{x_k}\mathcal{M}$ 的一个方向。具体而言， d_k 定义为 $\mathrm{T}_{x_k}\mathcal{M}$ 中两个元素的组合：黎曼梯度

$$g_k = \operatorname{grad} f(x_k) = \operatorname{proj}_{x_k} \nabla f(x_k),$$

以及动量项 s_k。为了确保 s_k 属于 $\mathrm{T}_{x_k}\mathcal{M}$，不能像欧几里得情形中那样，将其定义为两个连续迭代点 $x_k - x_{k-1}$ 之差。我们改为定义

$$s_k = \mathrm{T}_{x_k \leftarrow x_{k-1}}(\eta_{k-1} d_{k-1}),$$

也就是说，将之前的搜索方向传输到切空间 $\mathrm{T}_{x_k}\mathcal{M}$ 中。作为向量传输的一种实用选择，我们考虑投影到 $\mathrm{T}_{x_k}\mathcal{M}$ 上的正交投影，即

$$s_k = \operatorname{proj}_{x_k}(\eta_{k-1} d_{k-1}).$$

搜索方向随后由下式得到

$$d_k = -\alpha_k g_k + \beta_k s_k, \tag{7}$$

对于某些参数 $\begin{bmatrix}\alpha_k \\ \beta_k\end{bmatrix} \in \mathbb{R}^2$。

为了计算 α_k 和 β_k，如同第 4 节所总结的欧几里得情形 [18]，我们考虑在点 x_k 附近最小化函数 f 的二次近似：

$$\begin{aligned} &\min_{d \in \mathrm{T}_{x_k}\mathcal{M}} f(x_k) + \langle g_k, d\rangle + \frac{1}{2}\langle d, B_k[d]\rangle \\ &\text{s.t.} \quad d = -\alpha g_k + \beta s_k, \end{aligned} \tag{8}$$

其中， $\langle \cdot, \cdot \rangle$ 表示由 $\mathcal{E}$继承而来的切空间 $\mathrm{T}_{x_k}\mathcal{M}$ 上的内积，而 B_k 是一个线性、自伴且正定的算子，将 $\mathrm{T}_{x_k}\mathcal{M}$ 映射到其自身。注意，若 f 是二阶连续可微的， B_k 取为 $\mathrm{Hess} f(x_k)$，且 R_{x_k} 是二阶回缩，则模型 (8) 对应于 f 在 x_k处的二阶泰勒近似。

与欧几里得情形一样，该问题可以重新表述为 $\mathbb{R}^2$上的一个最小化问题：

$$\min_{u\in\mathbb{R}^2} T_k^\top u + \frac{1}{2}u^\top H_k u, \tag{9}$$

其中

$$T_k = \begin{bmatrix} -\|g_k\|^2 \\ \langle g_k, s_k\rangle \end{bmatrix}, \qquad u = \begin{bmatrix}\alpha \\ \beta\end{bmatrix},$$

and

$$H_k = \begin{bmatrix} \langle g_k, B_k[g_k]\rangle & -\langle g_k, B_k[s_k]\rangle \\ -\langle g_k, B_k[s_k]\rangle & \langle s_k, B_k[s_k]\rangle \end{bmatrix}. \tag{10}$$

一旦求得 (9) 的解 $u_k = \begin{bmatrix}\alpha_k \\ \beta_k\end{bmatrix}$，方向 d_k 便按 (7) 定义。一个主要问题是如何高效地选择算子 B_k（或者等价地，选择 $2\times$2 矩阵 H_k），同时尽可能不需要额外的回缩或函数与梯度求值。例如，对于某个 $v \in \mathrm{T}_x\mathcal{M}$，通过有限差分近似黎曼黑塞矩阵 $\mathrm{Hess} f(x)[v]$，需要额外进行一次回缩和梯度评估。此外，必须确保方向 d_k 具有梯度相关性，以保证收敛性。

注 1. 算子 B_k 在 $\mathrm{T}_{x_k}\mathcal{M}$ 上自伴且正定，这一事实确保了当 g_k 和 s_k 非零时，问题 (8) 存在最优解。证明可参照 [18, 命题 2], ，并作少量调整。此外，若 g_k 和 s_k 线性无关，则 (10) 中定义的 2×2 矩阵 H_k 满秩且正定，因此解是唯一的。此外，如果算子 $\{B_k\}$的特征值一致有界，那么通过 (7) 计算得到的方向序列$\{d_k\}$ 与 $\{x_k\}$ 梯度相关，并且满足假设 2。证明思路与 [18, 命题 3]类似。

6. 所提出的方法

本节讨论带动量的黎曼梯度法的一些实用选择。此外，我们证明所提出方法的全局收敛性。具体而言，我们选择一组特定方向 $\{d_k\}$，使其与点序列 $\{x_k\}$ 梯度相关；由此即可直接根据第 3 节的讨论得出该结论。

带动量的黎曼梯度法的流程与算法 1 相同，其中第 3 行的方向 d_k 按照以下步骤确定：

- 选择一个自伴正定算子 B_k，定义 2×2 对称度量和正定矩阵 H_k（如式 (10) 所示），并考虑二次极小化问题 (9)；
- 求出 (9) 的一个解 $[\alpha_k, \beta_k]^\top$，并将方向确定为 $d_k = -\alpha_k g_k + \beta_k s_k$；
- 如果 d_k 不是梯度相关的，则按照某些规则，定义一个属于 $\mathrm{T}_{x_k}\mathcal{M}$ 的新梯度相关方向 d_k。

两个主要问题是：

(1) 如何正确且高效地定义算子 B_k，从而定义问题 (9) 中的矩阵H_k？ (2) 如果所找到的梯度相关方向不满足该条件，如何选择一个新的梯度相关方向?

6.1. 算子 B_k 的实用选择

在本小节中，我们给出一种高效方法，用于选择问题 (9) 中的矩阵 H_k ：

$$H_k = \begin{bmatrix} \langle g_k, B_k[g_k]\rangle & -\langle g_k, B_k[s_k]\rangle \\ -\langle g_k, B_k[s_k]\rangle & \langle s_k, B_k[s_k]\rangle \end{bmatrix}.$$

我们的目标是考虑一个自伴且正定的算子 B_k ，使其在 f 二次连续可微时近似黎曼黑塞矩阵 Hess$f(x_k)$。请注意，无需显式计算算子 B_k，只需计算其作用 $B_k[g_k]$ 和 $B_k[s_k]$。我们可以使用海森矩阵的有限差分近似，或使用 [18, 19] 中采用的插值技术；不过，为避免额外的梯度或函数评估，我们采用另一种方法。考虑切空间 $\mathrm{T}_{x_k}\mathcal{M}$ 中定义的元素

$$y_k = g_k - \mathrm{T}_{x_k \leftarrow x_{k-1}}(g_{k-1}).$$

假设自伴且正定的算子 B_k 满足割线方程

$$B_k[s_k] = y_k.$$

那么，定义 $\rho_k = \langle g_k,\ B_k[g_k]\rangle$ 后，我们需要求解以下方程组

$$\begin{bmatrix} \rho_k & -\langle g_k, y_k\rangle \\ -\langle g_k, y_k\rangle & \langle s_k, y_k\rangle \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \|g_k\|^2 \\ -\langle g_k, s_k\rangle \end{bmatrix}. \tag{11}$$

我们注意到，只要算子 B_k 是自伴、正定的，且 g_k 和 s_k 线性无关，上述系统就有解。事实上，我们有 2×2 矩阵是正定的（参见注释 1）。

对于 B_k的实际选取，我们采用袁亚湘与Stoer针对无约束优化 [31], 提出的方法，其中假设 $s_k^\top y_k > 0$，并考虑 $n\times n$ 矩阵 B_k 的缩放无记忆BFGS更新，即

$$B_k = \frac{1}{\lambda_k}\left(I - \frac{s_k s_k^\top}{s_k^\top s_k}\right) + \frac{y_k y_k^\top}{s_k^\top y_k}, \tag{12}$$

其中标量 $\lambda_k > 0$ 根据 Barzilai–Borwein谱梯度法 [4, 23], 选取，例如,

$$\lambda_k = \frac{\|s_k\|^2}{s_k^\top y_k}.$$

由 (12) 定义的矩阵 B_k 是对称正定的，满足割线方程 $B_k s_k = y_k$，且构造时不需要额外的函数或梯度评估。

推广到黎曼优化情形是直接的。假设
that

$$\langle s_k, y_k \rangle > 0. \tag{13}$$

考虑到 (12)，定义算子 B_k： $\mathsf{T}_{x_k}\mathcal{M} \to \mathsf{T}_{x_k}\mathcal{M}$ 如下：对每个 $d \in \mathsf{T}_{x_k}\mathcal{M}$,

$$B_k[d] = \frac{1}{\lambda_k}\left(d - \frac{\langle s_k, d\rangle}{\|s_k\|^2} s_k\right) + \frac{\langle y_k, d\rangle}{\langle s_k, y_k\rangle} y_k, \tag{14}$$

其中 $\lambda_k > 0$。容易验证，(14) 中定义的算子 B_k 是自伴的、正定的，并且满足割线方程：$B_k[s_k] = y_k$。

我们可以按照 Barzilai–Borwein 方法选取标量 $\lambda_k > 0$。具体而言，令 $0 < \lambda_{\min} \le \lambda_{\max}$；我们可以将 λ_k 定义为：

$$\lambda_k = \min\left\{\lambda_{\max}, \max\{\lambda_{\min}, \tilde{\lambda}_k\}\right\}, \tag{15}$$

其中

$$\tilde{\lambda}_k = \lambda_k^{BB1} = \frac{\|s_k\|^2}{\langle s_k, y_k\rangle}, \quad \text{or} \quad \tilde{\lambda}_k = \lambda_k^{BB2} = \frac{\langle s_k, y_k\rangle}{\|y_k\|^2}. \tag{16}$$

6.2. 重启策略

首先，我们注意到，式 (13) 中定义的条件 $\langle s_k, y_k\rangle > 0$ 确保算子 B_k 是正定的。然而，如果不施加额外条件，就无法保证正定算子序列 $\{B_k\}$ 的特征值有界。因此，我们无法应用 [18,命题 3] 的黎曼对应形式（参见注释 1），并据此得出所获得的方向序列 $\{d_k\}$ 是梯度相关的；我们只能断定这些方向是下降方向，即对所有 k， $\langle d_k, g_k\rangle < 0$。

为解决这一问题，我们提议采用 [32]中使用的相同重启策略。选取两个常数 $0 < c_1 \le c_2$，并检验 d_k 是否满足梯度相关条件 (2)。否则，考虑一个新的方向 d_k^{new}，定义为

$$d_k^{new} = -\lambda_k g_k,$$

也就是说，取负梯度并乘以 Barzilai–Borwein 因子 λ_k进行缩放，如式 (15)–(16) 所定义。因此，在这种情况下，我们的方法退化为 Iannazzo & Porcelli [16], 提出的方法，但采用单调线搜索。容易验证，若所选常数满足 $c_1 \le \lambda_{\min}$ 和 $\lambda_{\max} \le c_2$，则新方向是梯度相关的。

注 2. 条件 $\langle s_k,\ y_k\rangle > 0$ 可通过采用基于强 Wolfe 条件的线搜索来满足。然而，这种线搜索的开销可能显著高于 Armijo 线搜索。实践中，我们建议不强制满足此条件。在

未验证 (13) 的情况下，我们建议如 [16], 一样直接取该方向为

$$d_k = -\lambda_{\max} g_k.$$

6.3. 收敛算法

基于前文讨论的策略，并通过求解系统 (11) 得到 α_k 和 β_k 的显式公式，所提出的算法在算法 3 中进行了形式化描述。其收敛性质将在下面的命题中给出。

命题 6.1. 假设假设 *1* 和 *3* 成立。那么，算法 *3* 至多需要 $\mathcal{O}(\epsilon^{-2})$ 次迭代、回缩以及函数与梯度求值，即可达到

$$\| \operatorname{grad} f(x_k) \| \leq \epsilon.$$

证明**.** 根据构造，序列 $\{d_k\}$ 与 $\{x_k\}$ 梯度相关；因此，假设 2 成立。此外，计算方向 d_k 时无需额外进行函数或梯度求值。因此，可以应用命题 3.1，得到最坏情况复杂度界 $\mathcal{O}(\epsilon^{-2})$。 □

7. 实验结果

按照算法 3 中的伪代码，我们使用 MATLAB R2025b 实现了所提出的带动量的黎曼梯度法 (RGMM)。实现代码及实验所用脚本公开发布于 https://github.com/DiegoScuppa/RGMM。

如第 6.1 节所述，算子 B_k 选用无记忆 BFGS 更新（14），由此得到第 15–16 行所报告的 α_k 和 β_k 的显式表达式。步长 λ_k 根据（15）选取，可采用以下策略选项之一：直接公式（（16）中的 λ_k^{BB1}）、逆公式（λ_k^{BB2}）或 [16]中提出的交替方案。此外，正如第 6.2 节所讨论的，当梯度相关条件不满足时，该方法将回退到 Barzilai–Borwein 步长，如 [16]所示。

我们将 RGMM 与 Manopt 软件包 [7]中的多个求解器进行比较：黎曼 Barzilai–Borwein（RBB）、黎曼共轭梯度法（RCG）、黎曼信赖域（RTR）和黎曼 L-BFGS（RLBFGS）。为确保公平比较，我们通过修改 Manopt 的 barzilaiborwein.m 例程实现了 RGMM，从而 [16],确保与 Manopt 接口完全兼容。

所有实验均在配备 16 GB 内存的 Intel Core Ultra 7 155H 机器上运行。我们从 Manopt 发行包中测试了 15 个问题，每个问题使用五种参数组合，共计 75 个问题实例。表 1 总结了问题集及其对应的流形结构。前八个问题与 [16], 中考虑的问题相同，但当时仅使用了前三种参数组合。

除非另有说明，否则均使用求解器的默认参数。对于 RCG，我们采用了 th适用于 β_k 的修正 Hestenes–Stiefel 更新，这是默认选项，且已被证明

算法 3： 带动量的黎曼梯度法

输入： $x_0 \in \mathcal{M}$, $\gamma \in (0,1)$, $\delta \in (0,1)$, $\epsilon > 0$, $0 < c_1 \le \lambda_{\min} \le \lambda_0 \le \lambda_{\max} \le c_2$

输出： 近似驻点 x_k

$k \leftarrow 0$

$f_k \leftarrow f(x_k)$

$g_k \leftarrow$ 梯度 $f(x_k)$

当 $\|g_k\| > \epsilon$ **时**

  如果 $k = 0$ **那么**

    $d_k \leftarrow -\lambda_0 g_k$

  else

    $s_k \leftarrow \mathrm{T}_{x_k \leftarrow x_{k-1}}(\eta_{k-1} d_{k-1})$

    $y_k \leftarrow g_k - \mathrm{T}_{x_k \leftarrow x_{k-1}}(g_{k-1})$

    # 检查曲率条件

    如果 $\langle s_k,\ y_k \rangle \le 0$ **那么**

      $d_k \leftarrow -\lambda_{\max} g_k$

    else

      # 计算系数和方向

      $\lambda_k \leftarrow$ 最小值 $\{\lambda_{\max}, \text{最大值}\{\lambda_{\min},\ \frac{\|s_k\|^2}{\langle s_k,\ y_k \rangle}\}\}$

      $\alpha_k \leftarrow \lambda_k \frac{\|g_k\|^2 \langle s_k,\ y_k \rangle}{\langle s_k,\ y_k \rangle \left(\|g_k\|^2 - \frac{\langle y_k \rangle - \langle g_k, \cdot \rangle \langle g_k,\ s_k \rangle}{\cdot\ \ s_k)^2} \right)}$

      $\beta_k \leftarrow \frac{\alpha_k \langle g_k,\ y_k \rangle - \langle g_k,\ s_k \rangle}{\langle s_k,\ y_k \rangle}$

      $d_k \leftarrow -\alpha_k g_k + \beta_k s_k$

      # 检查梯度相关方向

      如果 $\langle g_k, d_k \rangle > -c_1 \|g_k\|^2$ **或** $\|d_k\| > c_2 \|g_k\|$ **则**

        $d_k \leftarrow -\lambda_k g_k$

  # Armijo 线搜索

  $\eta \leftarrow 1$

  当 $f\big(R_{x_k}(\eta d_k)\big) > f_k + \gamma\, \eta\, \langle g_k, d_k \rangle$ **执行**

    $\eta \leftarrow \delta\, \eta$;

  $\eta_k \leftarrow \eta$

  $x_{k+1} \leftarrow R_{x_k}(\eta_k d_k)$

  $k \leftarrow k + 1$

  $f_k \leftarrow f\ (x_k)$

  $g_k \leftarrow$ 梯度 $f\ (x_k)$

表 1. 测试问题与流形结构。前八个问题也在 [16]中进行了研究。

Name	**Size**	**描述**	**流形**
KM	(a) $n = 50$ (b) $n = 100$ (c) $n = 200$ (d) $n = 500$ (e) $n = 1000$	计算一组数据的 Karcher 均值 $n \times n$ 正定矩阵。	SPD
SPCA	(a) $n = 100,\ p = 10,\ m = 2$ (b) $n = 500,\ p = 15,\ m = 5$ (c) $n = 1000,\ p = 30,\ m = 10$ (d) $n = 2000,\ p = 60,\ m = 20$ (e) $n = 5000,\ p = 1500,\ m = 50$	计算 m 稀疏主成分 $p \times n$ 矩阵编码的组成部分 p n 变量的样本。	Stiefel
SP	(a) $n = 8$ (b) $n = 12$ (c) $n = 24$ (d) $n = 50$ (e) $n = 100$	求 n 等圆的最大直径 可放置在球面上而不 重叠（球面堆积问题）。	球面的乘积 球面的乘积
DIS	(a) $n = 128$ (b) $n = 500$ (c) $n = 1000$ (d) $n = 2000$ (e) $n = 5000$	寻找以下对象的标准正交基 $n \times n$的主不变三维子空间 矩阵。	格拉斯曼
LRMC	(a) $n = 100,\ p = 10,\ m = 20$ (b) $n = 500,\ p = 15,\ m = 50$ (c) $n = 1000,\ p = 30,\ m = 100$ (d) $n = 2000,\ p = 60,\ m = 200$ (e) $n = 5000,\ p = 150,\ m = 500$	低秩矩阵补全：给定部分 秩为 p 的 $m \times n$ 矩阵的观测值, 尝试将其补全。	固定秩 矩阵
MC	(a) $n = 20$ (b) $n = 100$ (c) $n = 300$ (d) $n = 1000$ (e) $n = 5000$	最大割：给定一个图的 $n \times n$ 拉普拉斯矩阵, 求最大割或其上界, 即最大割值的上界。	对称正定矩阵 秩为 2
TSVD	(a) $n = 60,\ m = 42,\ p = 5$ (b) $n = 100,\ m = 60,\ p = 7$ (c) $n = 200,\ m = 70,\ p = 8$ (d) $n = 500,\ m = 80,\ p = 10$ (e) $n = 1000,\ m = 100,\ p = 15$	计算一个矩阵的奇异值分解 $m \times n$ 矩阵截断至秩 p。	格拉斯曼
GP	(a) $m = 10,\ N = 50$ (b) $m = 50,\ N = 100$ (c) $m = 50,\ N = 500$ (d) $m = 100,\ N = 1000$ (e) $m = 500,\ N = 5000$	广义普罗克鲁斯特斯：旋转对齐 点云。数据：矩阵 $A \in \mathbb{R}^{3\times m\times N}$，每个切片都是由 m 个点组成的点云，位于 $\mathbb{R}^3$ 中。	旋转的乘积 与 欧几里得 A 的空间
RIC	(a) $N = 10,\ K = 2$ (b) $N = 20,\ K = 3$ (c) $N = 50,\ K = 5$ (d) $N = 100,\ K = 10$ (e) $N = 200,\ K = 20$	射电干涉测量校准：计算 N 个站的 K 个接收器的增益矩阵。	对称 固定秩 复 矩阵
ESVD	(a) $N = 10$ (b) $N = 100$ (c) $N = 1000$ (d) $N = 10000$ (e) $N = 100000$	本质奇异值分解：解决该问题 $\sum_{i=1}^{N} \|\|E_i - A_i\|\|^2$（$E_i$ 本质矩阵）。	本质 矩阵
DSD	(a) $n = 50$ (b) $n = 100$ (c) $n = 200$ (d) $n = 500$ (e) $n = 1000$	双随机去噪：寻找一个双随机 矩阵，使其在 弗罗贝尼乌斯范数。	多项式 对称 矩阵
ESDP	(a) $n = 100$ (b) $n = 200$ (c) $n = 500$ (d) $n = 1000$ (e) $n = 2000$	椭圆体半正定规划：半定规划 （SDP），带有单位对角线约束。	球面的 乘积
PDIM	(a) $n = 2,\ m = 20$ (b) $n = 5,\ m = 50$ (c) $n = 10,\ m = 100$ (d) $n = 20,\ m = 200$ (e) $n = 50,\ m = 500$	正定内蕴均值：计算 一组 m 的内蕴均值 大小为 $n \times n$ 的正定矩阵。	SDP矩阵
GEC	(a) $n = 32$ (b) $n = 64$ (c) $n = 128$ (d) $n = 256$ (e) $n = 512$	求取.....的标准正交基 $B^{-1}A$的主不变三维子空间, 其中 A 和 B 是 $n \times n$ 矩阵。	格拉斯曼 广义
NE	(a) $n = 50,\ k = 5$ (b) $n = 100,\ k = 10$ (c) $n = 200,\ k = 20$ (d) $n = 500,\ k = 50$ (e) $n = 1000,\ k = 100$	非线性特征空间：求解非线性 特征值问题。	格拉斯曼

表 2. 在最佳 CPU 时间内求解的问题百分比（$\pi_S(1)$），以及成功求解的问题百分比（$\pi_S(\infty)$）。

求解器	$\pi_S(1)$	$\pi_S(\infty)$
RGMM	33.4%	98.1%
RBB	28.9%	98.5%
RCG	13。1%	95.3%
RTR	16.7%	100.0%
RLBFG	7.9%	98。8%

在我们的初步测试中最为有效。对于 RTR，即使已有解析海森矩阵，我们仍使用有限差分近似，以便将比较重点放在函数评估和梯度评估以及 CPU 时间和迭代次数上。

对于 RGMM，我们设置 $c_1 = 10^{-9}$ 和 $c_2 = 10^{9}$。除将 λ_0 固定为 1 外，其余所有参数均设为 RBB 的默认值：$\lambda_{\min} = 10^{-3}$、$\lambda_{\max}$ =10₃、$\gamma = 10^{-4}$、$\delta = 0.5$，以及策略=direct。与 RBB 不同，我们的方法采用单调线搜索，因为初步实验表明，非单调策略并不能获得更好的性能。

为完整起见，我们指出：我们并非总是在第 22 行用$\eta = 1$ 初始化线搜索，而是当步长 $\|d_k\|$ 显著大于先前迭代时，允许采用更小的初始值。引入这一保护措施，是因为初步测试表明，过大的初始步长会大幅增加线搜索所需的函数评估次数和回缩次数。

每个问题均从 10 个随机初始点 x_0 开始求解，共计 750 个实例。所有求解器均从相同的 x_0 初始化，并在黎曼梯度的范数降低至原来的 10^{-6} 倍时终止，即当 tolgradnorm= $10^{-6}\|\operatorname{grad} f(x_0)\|$ 时。其他停止准则包括 maxiter=$5 \cdot 10^4$、maxtime= 600 秒和 minstepsize= 10^{-10}。如果求解器因任一其他停止准则而终止，则视为该求解器在相应实例上失败。对于每个求解器，我们计算了 10 次运行的 CPU 时间、迭代次数、函数评估次数、梯度评估次数和失败次数的平均值；详细结果见 GitHub 仓库。

我们还报告了性能剖面 [12]。对于求解器 S 和问题实例 P，令 t_P^S 表示求解器 S 在 P 上针对所选度量（例如 CPU 时间）的计算成本；若 S 在 P 上失败，则令 $t_P^S = \infty$。令 $t_P = \min_S t_P^S$。对于 $\tau \geq 1$，求解器 S 的性能剖面定义为

$$\pi_S(\tau) = \frac{\#\{\text{instances } P : t_P^S / t_P \leq \tau\}}{\#\{\text{instances } P\}}.$$

这里，$\pi_S(1)$ 表示 S 是最佳求解器的实例所占比例，而对于较大的 τ 值，$\pi_S(\tau)$ 衡量 S 的表现与最佳求解器相差不超过 τ 倍的频率。

图 1 显示了 $\tau \in [1, 10^2]$ 的 CPU 时间性能剖面。如表 2 所报告，RGMM 在大多数实例上表现最佳（即具有最大的 $\pi_S(1)$）。此外，对于介于 1 与 8 之间的 τ，RGMM 达到最高的剖面，表明其在广泛的容差范围内具有更优的稳健性。对于 $\tau \geq 10^2$，性能剖面呈现常数值，这些值代表由相应

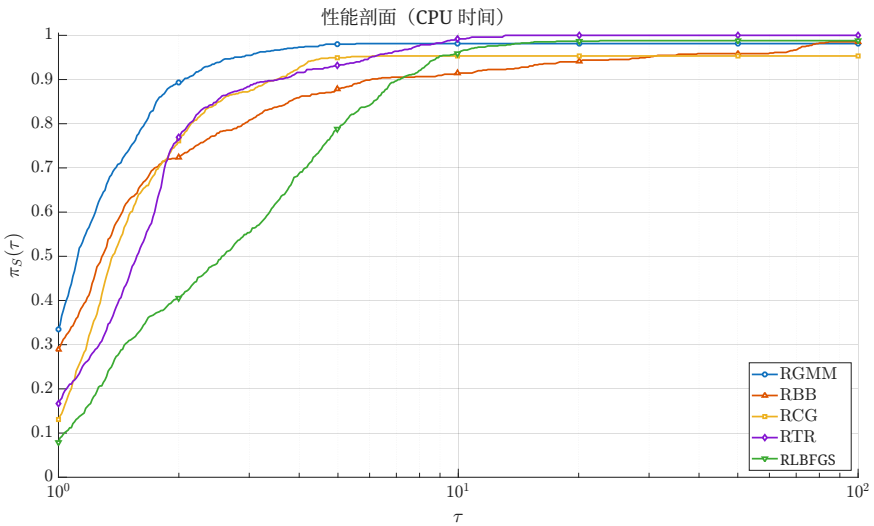


图 1. CPU 时间性能剖面。

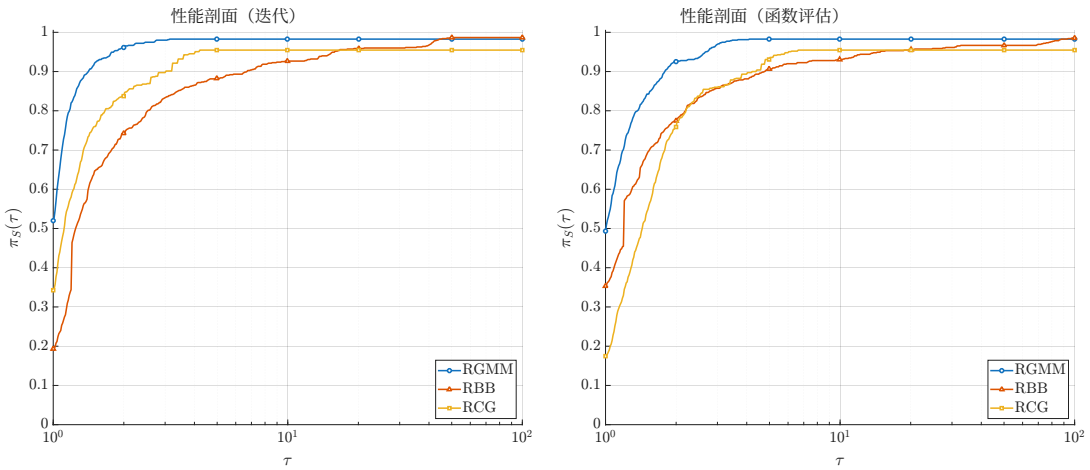
求解器求解的实例数量，如表 2 所报告。具体而言，

- RGMM 在 75 个问题中的一个问题（GEC (e)）的所有实例上均失败；此外，它在 GEC (c) 的两个实例和 GEC (d) 的两个实例上失败；
- RBB 在一个问题（RIC (e)）的所有实例上均失败；此外，它在 GEC (e) 的一个实例上失败；
- RCG 在两个问题（GEC (e) 和 KM (e)）的所有实例上均失败；此外，它在 GEC (c) 的八个实例、GEC (d) 的五个实例、SP (a) 的一个实例和 SP (b) 的一个实例上失败；
- RTR 求解了所有问题的全部实例；
- RLBFGS 求解了所有问题的至少一个实例，但在 GEC(c) 的一个实例、GEC(e) 的八个实例以及 KM(e) 的一个实例上未能求解。

图2给出了 RGMM、RBB 和 RCG 的迭代次数与函数评估次数性能剖面。RGMM 在 52.0% 的实例中迭代次数最少，在 49.3% 的实例中函数评估次数最少。此外，除 τ 足够大时以外，在所有取值下，RGMM 在这两项度量上的表现均最佳，这反映出 RBB 求解的实例略多于 RGMM。

总体而言，实验结果表明 RGMM 是一种稳健且具有竞争力的方法。在大多数测试实例中，它通常所需的迭代次数和函数评估次数最少，并展现出优异的 CPU 时间性能。具体而言，RGMM 在约 33.4% 的实例中是最快的求解器，并在 $\tau \in [1, 8]$ 下取得最高的性能剖面，凸显了它在广泛容差范围内的稳定性。其失败率可忽略不计，与其他竞争求解器相当。

最后，在所有测试运行中，除问题 SP (a) 的三个实例中各有一次迭代外，$c_1 = 10^{-9}$ 和 $c_2 = 10^{9}$ 均满足梯度相关条件。直观而言，选择足够小的 c_1 和足够大的 c_2，会使这些条件近似等同于标准下降要求，而我们的方法始终满足这一要求。此外，曲率条件 $\langle s_k,\ y_k \rangle > 0$ 很少被违反：在总计 107951 次迭代中仅有 514 次（不到 0.5%）；因此，第 6.2 节所述的保护措施主要起到理论上的保障作用。



（a）迭代性能剖面。　　（b）函数评估性能剖面。

图2. 迭代与函数评估性能剖面（RGMM, RBB, RCG）。

8. 结论

本文提出了一种新的带动量的黎曼梯度方法，该方法将 Lapucci 等人近期为无约束情形 [18]开发的方法加以改编。我们研究了在合理假设下收敛至驻点的全局收敛性，并给出了 $\mathcal{O}(\epsilon^{-2})$ 的最坏情况复杂度界；这些假设比 [28] 中提出的假设更宽松。我们提出了一种实用且有效的算子 B_k 选取方法，无需显式了解海森矩阵，也无需额外计算梯度、函数值或回缩，因此这种选取方式行之有效；不过，未来仍可研究其他方案。我们在 Manopt 软件包中的一系列问题上测试了该方法，并将其与现有求解器进行了比较。我们的算法在大多数问题上是速度最快的求解器，同时也是最稳健的，展现出良好的可靠性，因此适合用于求解黎曼流形上的优化问题。

致谢

作者谨向Marco Sciandrone教授致以诚挚谢意，感谢他提出的宝贵建议、对稿件的细致审阅，以及在本研究开展过程中给予的鼓励。

利益披露声明

作者声明不存在需要披露的利益冲突。

资金支持

开展本研究未获得任何资助。

代码可用性声明

本文实验部分所开发的代码已在 https://github.com/DiegoScuppa/RGMM 公开提供。

参考文献

[1] P.A. Absil、C.G. Baker 和 K.A. Gallivan，黎曼流形上的信赖域方法，《计算数学基础》7（2007），第 303–330 页。可访问 https://doi.org/10.1007/s10208-005-0179-9。[2] P.A. Absil、R. Mahony 和 R. Sepulchre，格拉斯曼流形的黎曼几何：兼论算法计算，《应用数学学报》80（2004），第 199–220 页。可访问 https://doi.org/10.1023/B:ACAP.0000013855.14971.91。[3] P.A. Absil、R. Mahony 和 R. Sepulchre，矩阵流形上的优化算法，普林斯顿大学出版社，普林斯顿，新泽西州，2008 年。可访问 https://doi.org/ 10.1515/9781400830244。[4] J. Barzilai 和 J.M. Borwein，两点步长梯度方法，《IMA数值分析杂志》8（1988），第 141–148 页。可访问 https://doi.org/10.1093/imanum/ 8.1.141。[5] D. Bertsekas，非线性规划，第 3 版，Athena Scientific 优化与计算系列，Athena Scientific，2016 年。可访问 https://books.google.it/books? id=TwOujgEACAAJ。[6] N. Boumal 和 P.A. Absil，通过黎曼优化在格拉斯曼流形上进行低秩矩阵补全，《线性代数及其应用》475（2015），第 200–239 页。可访问 https://doi.org/10.1016/j.laa.2015.02.027。[7] N. Boumal、B. Mishra、P.A. Absil 和 R. Sepulchre，*Manopt*：用于流形上优化的 *Matlab*工具箱，《机器学习研究杂志》15（2014），第 1455–1459 页。可访问 https://www.manopt.org。[8] N. Boumal，采用有限差分海森近似的黎曼信赖域方法具有全局收敛性，载于信息几何科学，01. 2015，第 467–475 页。可访问 https://doi.org/10.1007/978-3-319-25040-3_50。[9] N. Boumal，光滑流形上的优化导论，剑桥大学出版社，2023 年。可访问 https://www.nicolasboumal.net/book。[10] N. Boumal、P.A. Absil 和 C. Cartis，流形上非凸优化的全局收敛速率，《IMA数值分析杂志》39（2018），第 1–33 页。可访问 https://doi.org/10.1093/imanum/drx080。[11] C. Cartis、P. Sampaio 和 P. Toint，无约束优化中非单调梯度相关算法的最坏情况评估复杂度，《优化》64（2015），第 1349–1361 页。可访问 https://doi.org/10.1080/02331934.2013.869809。[12]E. Dolan 和 J. Mor´e，通过性能剖面对优化软件进行基准测试，《数学规划》91（2001）。可访问 https://doi.org/10.1007/ s101070100263。[13] D. Gabay，在微分流形上极小化可微函数，《优化理论与应用杂志》37（1982），第 177–219 页。可访问 https://doi.org/10.1007/BF00934767。[14]L. Grippo 和 M. Sciandrone，非线性优化方法导论，Springer，2023 年。可访问 https://doi.org/10.1007/978-3-031-26790-1。[15] W. Huang、P.A. Absil 和 K.A. Gallivan，一种无需微分回缩的非凸优化黎曼*BFGS*方法，《SIAM优化杂志》28（2018），第 470–495 页。可访问 https://doi.org/10.1137/17M1127582。[16] B. Iannazzo 和 M. Porcelli，采用非单调线搜索的黎曼 *Barzilai–Borwein* 方法及矩阵几何平均计算，《IMA数值分析杂志》

析》38（2017），第 495–517 页。可访问 https://doi.org/10.1093/imanum/drx015。[17]D. Kressner、M. Steinlechner 和 B. Vandereycken，通过黎曼优化进行低秩张量补全，《BIT 数学》54（2014），第 447–468 页。可访问 https://doi.org/10.1007/s10543-013-0455-z。[18] M. Lapucci、G. Liuzzi、S. Lucidi、D. Pucci 和 M. Sciandrone，一种具有动量的全局收敛梯度方法，《计算优化与应用》93（2025），第 795–820 页。可访问 https://doi.org/10.1007/s10589-025-00741-5。[19] M. Lapucci、G. Liuzzi、S. Lucidi、M. Sciandrone 和 D. Scuppa，具有动量的投影梯度方法（2026）。可访问 https://arxiv.org/abs/2601.16683。[20] M.B. Li 和 M.A. Erdogdu，用于求解半定规划的黎曼朗之万算法，《Bernoulli》29（2023），第 3093–3113 页。可访问 https://doi.org/10.3150/ 22-BEJ1576。[21] T. Lin 和 H. Zha，黎曼流形学习，《IEEE模式分析与机器智能汇刊》30（2008），第 796–809 页。可访问 https://doi.org/10.1109/TPAMI.2007.70735。[22] F. Liu、C. Masouros、A. Li、H. Sun 和 L. Hanzo，*MU-MIMO*通信与*MIMO*雷达：从共存到联合传输，《IEEE无线通信汇刊》PP（2018），第 1–1 页。可访问 https://doi.org/10.1109/TWC.2018.2803045。[23] M. Raydan，大规模无约束极小化问题的 *Barzilai-Borwein*梯度法，《SIAM优化杂志》7（1997），第 26–33 页。可访问 https://doi.org/10.1137/S1052623494266365。[24] W. Ring 和 B. Wirth，黎曼流形上的优化方法及其在形状空间中的应用，《SIAM优化杂志》22（2012），第 596–627 页。可访问 https://doi.org/10.1137/11082885X。[25] D. Rosen、L. Carlone、A. Bandeira 和 J. Leonard，*SE-Sync*：一种用于特殊欧氏群上同步的可认证正确算法，《国际机器人研究杂志》38（2016）。可访问 https://doi.org/10.1177/0278364918784361。[26] H. Sato，黎曼优化及其应用，Springer，2021 年。可访问 https://doi.org/10.1007/978-3-030-62391-3。[27] H. Sato，黎曼共轭梯度法：一般框架及具有收敛性分析的具体算法，《SIAM优化杂志》32（2022），第 2690–2717 页。可访问 http://dx.doi.org/10.1137/21M1464178。[28]T. Tang、K.C. Toh、N. Xiao 和 Y. Ye，一种黎曼降维二阶方法及其在传感器网络定位中的应用，《SIAM科学计算杂志》46（2024），第 A2025–A2046 页。可访问 https://doi.org/10.1137/ 23M1567229。[29] O. Tuzel、F. Porikli 和 P. Meer，通过在黎曼流形上分类进行行人检测，《IEEE模式分析与机器智能汇刊》30（2008），第 1713–1727 页。可访问 https://doi.org/10.1109/TPAMI.2008.75。[30] B. Vandereycken，通过黎曼优化进行低秩矩阵补全，《SIAM优化杂志》23（2013），第 1214–1236 页。可访问 https://doi.org/10.1137/ 110845768。[31] Y.X. Yuan 和 J. Stoer，共轭梯度算法的子空间研究，《ZAMM——应用数学与力学杂志 / Zeitschrift f¨ur Angewandte Mathematik und Mechanik》75（1995），第 69–77 页。可访问 https://onlinelibrary.wiley.com/doi/abs/10.1002/zamm.19950750118。[32] J. Zhou、K. Deng、H. Wang 和 Z. Peng，一种具有迭代复杂度保证的新型黎曼共轭梯度法（2026）。可访问 https://doi.org/10.11948/20240564。

附录 A. 证明

在本附录中，我们给出命题 3.1 的证明。首先，我们证明 Armijo 线搜索过程的一个关键性质。

引理 A.1. 假设假设 *2–3* 成立（其中常数为 $0 < c_1 \le c_2$ 和$L > 0$）。令 $\delta \in (0,1)$， $\gamma \in (0,1)$。那么，*Armijo* 线搜索会在有限次迭代内终止，并找到一个参数 $\eta_k = \delta^{j_k} \in (0,1]$，其中 $j_k \in \mathbb{N}$。此外，

(1) 若 $L \le 2(1-\gamma)\frac{c_1}{c_2^2}$，则 $j_k = 0$（即 $\eta_k = 1$）；

(2) 否则，

$$j_k \le \left\lfloor \log_\delta \left(\frac{2}{L}(1-\gamma)\frac{c_1}{c_2^2} \right) \right\rfloor + 1.$$

证明。 根据假设 3，对于所有 k 和所有 $j \in \mathbb{N}$，

$$f(\mathrm{R}_{x_k}(\delta^j d_k)) \le f(x_k) + \delta^j \langle \operatorname{grad} f(x_k), d_k \rangle + \frac{L}{2}\delta^{2j}\|d_k\|^2. \tag{A1}$$

对于所有 k，定义 Armijo 线搜索中失败试验的集合（该集合可能为空）：

$$J_k := \left\{ j \in \mathbb{N} \text{ such that } f(\mathrm{R}_{x_k}(\delta^j d_k)) > f(x_k) + \gamma\delta^j \langle \operatorname{grad} f(x_k), d_k \rangle \right\}.$$

然后，再利用 (A1)，对于所有 $j \in J_k$，我们有

$$(\gamma - 1)\delta^j \langle \operatorname{grad} f(x_k), d_k \rangle < \frac{L}{2}\delta^{2j}\|d_k\|^2. \tag{A2}$$

因此，由 (2) 和 (A2) 可得：

$$(1-\gamma)c_1\|g_k\|^2 < \frac{L}{2}\delta^j c_2^2 \|g_k\|^2.$$

特别地，

$$\delta^j > \frac{2}{L}(1-\gamma)\frac{c_1}{c_2^2}.$$

若 $L \le 2(1-\gamma)\frac{c_1}{c_2^2}$，则有 $\delta^j > 1$，即 $j < 0$。因此，集合 J_k为空。由此，取 $\eta_k = \delta^0 = 1$即可满足 Armijo 条件。否则，每个$j \in J_k$都必须满足 $0 \le j <$log $_\delta \frac{2}{L}(1-\gamma)\frac{c_1}{c_2^2}$。因此，取 $\eta_k = \delta^{j_k}$即可满足 Armijo 条件，其中 j_k至多为 $\lfloor$log$_\delta(\frac{2}{L}($ $1-\gamma)\frac{c_1}{c_2^2})\rfloor + 1$。 □

命题 ***3.1*** 的证明。令 $\epsilon > 0$。利用 (2a)、假设 1 和 Armijo充分下降，对于所有 $K \in \mathbb{N}$, 有:

$$\begin{aligned}\sum_{k=0}^{K-1} \|\operatorname{grad} f(x_k)\|^2 \leq \sum_{k=0}^{K-1} \big(-\frac{1}{c_1}\langle \operatorname{grad} f(x_k), d_k\rangle\big) \leq \\ \leq \sum_{k=0}^{K-1} \frac{1}{\gamma c_1 \eta_k}\big(f(x_k) - f(x_{k+1})\big) \leq \frac{1}{\gamma c_1 \tilde{\eta}}\big(f(x_0) - f_{low}\big),\end{aligned} \tag{A3}$$

其中， $\tilde{\eta}$ 是一个满足以下条件的量：对所有 k， $0 < \tilde{\eta} \leq \eta_k$ （该量由引理 A.1 保证存在）。于是，取 $K = k_\epsilon$，可得对所有 $k = 0$、.....、$k_\epsilon - 1$， $\|\operatorname{grad} f(x_k)\| > \epsilon$；因此:

$$k_\epsilon \cdot \epsilon^2 < \sum_{k=0}^{k_\epsilon - 1} \|\operatorname{grad} f(x_k)\|^2. \tag{A4}$$

由 (A3) 和 (A4) 可得:

$$k_\epsilon \leq \frac{\big(f(x_0) - f_{low}\big)}{\gamma c_1 \tilde{\eta}}\, \epsilon^{-2}.$$

由引理 A.1，若 $L \leq 2(1-\gamma)\frac{c_1}{c_2^2}$，则对所有 k， $\eta_k = 1$；因此， $\tilde{\eta}$ 可取为 1。此外，每次迭代只进行一次函数评估和一次回缩。因此，(4) 得证。否则，再由引理 A.1，

$$\eta_k \geq 2\frac{1-\gamma}{L}\frac{c_1}{c_2^2}\delta, \tag{A5}$$

那么，我们可以将 $\tilde{\eta}$ 取为 (A5) 的右端。于是，(5a) 得证。此外，在每次迭代中，至多

$$\left\lfloor \log_\delta\left(\frac{2}{L}(1-\gamma)\frac{c_1}{c_2^2}\right)\right\rfloor + 2.$$

进行函数评估和回缩。因此，(5b) 得证。此外，对 $K \to \infty$ 取极限，也满足 (A3) 中的不等式。因此，回忆 ni_ϵ 的定义，

$$ni_\epsilon \cdot \epsilon^2 < \sum_{k:\|\operatorname{grad} f(x_k)\|>\epsilon} \|\operatorname{grad} f(x_k)\|^2 \leq \sum_{k=0}^{\infty} \|\operatorname{grad} f(x_k)\|^2 \leq \frac{1}{\gamma c_1 \tilde{\eta}}\big(f(x_0) - f_{low}\big).$$

因此,

$$ni_\epsilon \leq \frac{\big(f(x_0) - f_{low}\big)}{\gamma c_1 \tilde{\eta}}\, \epsilon^{-2}.$$

因此，对于 ni_ϵ，(4) 和 (5a) 也成立。 □