Multiple-Objective Manifold Optimization: Stochastic and Deterministic Approaches with Theoretical Insights and Applications

Nima Eslami

In this research proposal, our focus is on advancing manifold optimization techniques, particularly within the context of multiple-objective optimization problems. We aim to develop and analyze trust-region methods for such optimization problems on Riemannian manifolds, both in deterministic and stochastic settings. Here, we address the mathematical preliminaries required for understanding manifold optimization, and optimality conditions for both constrained and unconstrained cases, and propose a study of the convergence properties of the developed methods. Also, we provide instances of practical applications in various fields such as machine learning, finance, and image processing to illustrate the motivation of our research and its potential to innovate in areas where manifold structures naturally emerge. Specifically, our proposed topics of research include investigating optimality conditions and constraint qualifications, and nonsmooth and stochastic trust-region-based methods for multiple-objective optimization problems on manifolds.

Key words: Manifold Optimization, Multiple-Objective Optimization, Trust-Region Methods, Stochastic Optimization, Constraint Qualifications.

Contents

1. Problem Formulation

In the general case, the Riemannian optimization problem is formulated as:

$$\begin{aligned} \min_{p} \quad & F(p) \\ \text{s.t.} \quad & H(p) \leqslant 0, \\ & p \in M, \end{aligned} \tag{1}$$

where M is a Riemannian manifold (RM) and $F(p) = (F^1(p), \ldots, F^m(p))$ and $H(p) = (H^1(p), \ldots, H^m(p))$ for $p \in M$ are functions on M. If $m = 1$, then (1) is called a single-objective Riemannian optimization (SORO) problem, and when $m \geq 2$, (1) is called a multiple-objective Riemannian optimization (MORO) problem. The format of an unconstrained SORO problem would be defined as:

$$\begin{aligned} \min_{p} \quad & f(p) \\ \text{s.t.} \quad & p \in M. \end{aligned} \tag{2}$$

Moreover, an unconstrained version of (1) happens when function H is omitted, otherwise, it is named a constrained optimization problem.

Example 1 (Absil et al. 2004)**.** *In Problem (2), let M be the Stiefel manifold:*

$$St(m, n, \mathbb{R}) = \{X \in M_{m,n}(\mathbb{R}) \mid X^T X = I_n\}. \tag{3}$$

Considering matrices $Q \in M_{d,m}(\mathbb{R})$ and $P \in M_{d,n}(\mathbb{R})$, the following problem is known as the unbalanced Procrustes problem and is defined as:

$$\begin{aligned} \min_{X} \quad & \| QX - P \|_F^2 \\ s.t. \quad & X \in St(m, n, \mathbb{R}). \end{aligned} \tag{4}$$

where $\| . \|_F$ is the Frobenius norm. Note that by taking $M = \mathbb{R}^{m \times n}$, the above problem reduces to the least-squares problem.

Example 2. *In Problem (2), let $M = \mathbb{S}^2$. Consider n arbitrary points $q_1, \ldots, q_n$ on $\mathbb{S}^2$. The unconstrained Riemannian center of mass (or the Riemannian Karcher mean) optimization problem is defined as:*

$$\begin{aligned} \min_{p} \quad & \sum_{i=1}^{n} d_{\mathbb{S}^2}^2(p, q_i) \\ s.t. \quad & p \in \mathbb{S}^2. \end{aligned} \tag{5}$$

where $d_{\mathbb{S}^2}$ is the distance on $\mathbb{S}^2$. A constrained version of (5) is formulated by Bergmann and Herzog (2019) as:

$$\begin{aligned} \min_{p} \quad & \sum_{i=1}^{n} d_{\mathbb{S}^2}^2(p, q_i) \\ s.t. \quad & d_{\mathbb{S}^2}^2(p, c) \leq r^2, \\ & p \in \mathbb{S}^2. \end{aligned} \tag{6}$$

The added constraint to (6) forces the solution of (6) to lie in a closed ball of radius r and center $c \in \mathbb{S}^2$.

Example 3. *Considering the general formulation (1), we can extend Example (2) to the multiple-objective case. Suppose we seek to minimize the distance between a point $p \in \mathbb{S}^2$ and some points $w_i \in \mathbb{S}^2$ for $i \in \{1, \dots, n\}$, while simultaneously maximizing the distance between p and other points $v_j \in \mathbb{S}^2$ for $j \in \{1, \dots, m\}$, then we would need to formulate a multiple-objective optimization problem. The problem can be structured as follows:*

$$\begin{aligned} &\min_{p} \;\; \Big(\sum_{i=1}^{n} d^2_{\mathbb{S}^2}(p, w_i), -\sum_{j=1}^{m} d^2_{\mathbb{S}^2}(p, v_j)\Big) \\ &s.t. \;\; p \in \mathbb{S}^2. \end{aligned} \tag{7}$$

It is also possible to consider the constrained version of Problem (7) as in Problem (6), by restricting the feasible set in (7) to a closed ball.

2. Applications and Motivations

Manifold optimization has found a wide variety of applications in many theoretical and practical areas, particularly in Machine Learning. In the following, we list a number of recent applications of manifold optimization approaches.

- **Manifold-valued Images and Their Applications in Image Processing**. In the context of image processing, the manifold approach is particularly valuable for problems where the data or the parameters of interest naturally reside on a manifold, enabling more efficient and effective optimization than traditional methods that do not consider the manifold structure. In this regard, positive-definite (PD) matrices have become important in medical imaging, notably in diffusion tensor magnetic resonance imaging (DTI), see Bergmann et al. (2016), Yger et al. (2016), and Diepeveen et al. (2021). This technique, which utilizes nuclear magnetic resonance, assumes water molecule diffusion in biological tissues is Gaussian, making each DTI a collection of PD matrices that detail local diffusion processes per voxel (see Pennec (2020) for more details). On the other hand, the set of PD matrices $PD(n, \mathbb{R})$ is an RM, particularly a Hadamard manifold, which is associated with the following Riemannian metric:

$$g_A(P, Q) = \text{trace}(A^{-1} X A^{-1} Y), \quad \forall P, Q \in T_A(PD(n, \mathbb{R})), \tag{8}$$

where $A \in PD(n, \mathbb{R})$. Application of manifold optimization in image processing is not limited to images with values in $PD(n, \mathbb{R})$. Bergmann et al. (2016) and Bacak et al. (2016) study techniques of processing images with values living in a one-dimensional sphere $\mathbb{S}^1$, two-dimensional-sphere $\mathbb{S}^2$, and the space of all rotations in 3-dimensional space $SO(3)$. In the general case, the manifold $SO(n, \mathbb{R})$ is defined as:

$$SO(n, \mathbb{R}) = \{X \in M_{n\times n}(\mathbb{R}) \mid X^T X = I_n, \, \det(X) = 1\}, \tag{9}$$

The Riemannian metric on $SO(n,\mathbb{R})$ would be:

$$g_X(A,B) = \text{trace}(A^T B), \quad \forall A,B \in T_X(SO(n,\mathbb{R})), \tag{10}$$

where the tangent space $T_X(SO(n,\mathbb{R}))$ at a point $X \in SO(n,\mathbb{R})$ is defined as:

$$T_X(SO(n,\mathbb{R})) = \{H \in M_{n\times n}(\mathbb{R}) \mid X^T H + H^T X = 0\}. \tag{11}$$

The application of considering $\mathbb{S}^1$ as the domain of image data is extended to the interferometric synthetic aperture radar or InSAR, see Burgmann et al. (2000) and Bergmann et al. (2016). Images with values in $\mathbb{S}^2$ are applied in 3D directional information (Vese et al. 2002 and Lai et al. 2014) and colored-image processing in the chromaticity-brightness, Chan et al. (2001) and colored-image restoration, Jia et al. (2019). Processing $SO(3,\mathbb{R})$-valued images have applications in tracking robotic 3D rotational data (Drummond et al. 2002 and Weinmann et al. 2014), analysis of backscatter diffraction data (Bachmann et al. 2011) and directional statistics (Moakher 2002 and Axen et al. 2023).

- **Estimation of Covariance Matrix and Its Applications in Finance**. The covariance matrix plays a crucial role in many finance problems such as portfolio optimization (Pantaleo et al. 2011), risk management (Yu et al. 2020), and asset pricing models (Barillas and Shanken 2018). In such problems, the covariance matrix appears in risk measures used in the model, such as the variance risk measure, the standard deviation risk measure, the Value-at-Risk risk measure, and the Conditional Value-at-Risk measure. It is desirable to find a symmetric positive definite estimation to make the mentioned risk measures strictly convex (see Agrawal et al. 2022 and Choi et al. 2019). The problem of finding the best symmetric positive definite approximation of a given matrix, such as the estimated covariance matrix of asset returns in the portfolio optimization problem, can be written as a minimization problem over the RM $PD(n,\mathbb{R})$:

$$\begin{aligned} \min_X \quad & \| A - X \|_F^2 \\ \text{s.t.} \quad & X \in PD(n,\mathbb{R}). \end{aligned} \tag{12}$$

It can be applicable to consider some added constraints to (12), as in Boyd and Xiao (2005). Recently, Han and Park (2022) developed models that leverage either asset returns or realized covariances and introduced a new estimation method based on minimizing the geodesic length between forecasted and observed covariance matrices. Focusing on the manifold optimization technique, Han and Park (2022) offer a more accurate and theoretically sound method for covariance matrix estimation which is applied in portfolio optimization and risk management problems.

- **Principal Component Analysis and Its Applications in Machine Learning**. Principal Component Analysis (PCA) is widely regarded as the leading technique for data exploration and analysis across scientific disciplines. The main aim of PCA is to reduce the dataset to a smaller set of features that capture the essence of the original data in a lower-dimensional space, ensuring minimal information loss (see Greenacre et al. 2022 for more details). As in Boumal (2023), the PCA problem can be formulated as a manifold optimization problem over the Stiefel manifold $St(m, n, \mathbb{R})$ described in (3):

$$\begin{aligned} \max_{U} \ & \operatorname{trace}(U^T X^T X U D) \\ \text{s.t.} \ & U \in St(m, n, \mathbb{R}). \end{aligned} \tag{13}$$

Where $D \in M_{n\times n}(\mathbb{R})$ is a diagonal matrix with arbitrary diagonal entries $\alpha_1 > \ldots > \alpha_n > 0$. Here, $X \in M_{m\times k}(\mathbb{R})$ is the matrix whose columns are data points. In the literature, there are variants to the PCA problem, such as the Sparse PCA in Bertsimas and Kitnae (2023), Journee et al. (2010) and its manifold optimization approach in Chen et al. (2020) and Zhang et al. (2022), the Robust PCA in Yi et al. (2016), Maunu et al. (2019) and Zhang et al. (2018) in which the manifold optimization approach is used to for the Robust PCA problem, the streaming kernel PCA in Huang et al. (2021), and Ullah et al. (2018) and a manifold optimization approach to this problem in Tripuraneni et al. (2018), the distributed PCA in Liang et al. (2014), Gang et al. (2022) and its manifold versions in Wang et al. (2023) and in Huang et al. (2020), the Supervised (and Unsupervised) PCA in Barshan et al. (2011), Ghojogh and Crowly (2019), and its manifold optimization version in Ritchie et al. (2019). Moreover, Ritchie et al. (2022) used a multiple-objective optimization approach for the Supervised PCA problem. Up to now, there is no multiple-objective manifold optimization approach for targeting the Supervised PCA problem.

- **Gaussian Mixture Models and Their Applications in Machine Learning**. A Gaussian Mixture Model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of several Gaussian distributions with unknown parameters. The goal of a GMM is to estimate the parameters of the Gaussians (the means, covariances, and the mixture weights) that best fit the data. This is typically done using the Expectation-Maximization (EM) algorithm, a two-step iterative optimization technique that alternates between:

— *Expectation step*: Calculate the probability that each data point belongs to each of the Gaussian distributions, given the current parameter estimates.

— *Maximization step*: Update the parameters of the Gaussians to maximize the likelihood of the data given these probabilities.

See Reynolds (2009) on the GMM model and Dempster et al. (1977) and Xu and Jordan (1996) on the EM algorithm. The GMM optimization problem is as the following:

$$\begin{aligned} \max_{\beta,\mu,\Gamma} \quad & \sum_{i=1}^{m} \big(\log \sum_{j}^{n} (\beta_j P_G(x_i,\mu_j,\Gamma_j)\big) \\ \text{s.t.} \quad & \mu \in \mathbb{R}^n, \\ & \Gamma \in PD(n,\mathbb{R}), \\ & \beta_j \in \Delta^+_{n-1} := \{r_i \in \mathbb{R} \mid \ r_k > r_l > 0, \quad \forall k,l \in \{1,\ldots,n\};\ k>l,\ \sum_{j=1}^{n} r_j = 1.\}, \end{aligned} \tag{14}$$

where $\{x_i\}_{i=1}^m \in$ are sampled independently from a mixture of m Gaussians, meaning that data points sampled from the Gaussian probability distribution of mean $\mu \in \mathbb{R}^m$ and covariance Γ with the density function of:

$$P_G(x,\mu,\Gamma) = \frac{-e^{\dfrac{(x-\mu)^T\Gamma^{-1}(x-\mu)}{2}}}{\det(\Gamma)^{1/2}(2\pi)^{-m/2}}, \tag{15}$$

By solving (14), the goal is to find parameter estimations $(\hat{\beta},\hat{\mu},\hat{\Gamma}) \in \Delta^+_{n-1} \times \mathbb{R}^n \times PD(n,\mathbb{R})$. Problem (14) is solved by the EM algorithm. As Δ^+_{n-1} and $PD(n,\mathbb{R})$ are manifolds, Problem (14) can be solved by manifold optimization techniques. Gaussian Mixture Models are tremendously applicable in machine learning, big data analysis, and pattern recognition (see Bouguila and Fan (2020) and Vanish et al. (2019) for applications of GMM).

- **Geometric Matrix Mean and Its Applications in Machine Learning**. The geometric matrix mean optimization problem is formulated as:

$$\begin{aligned} \min_{X} \quad & \sum_{i=1}^{n} d^2_{PD(n,\mathbb{R})}(X,C_i) \\ \text{s.t.} \quad & X \in PD(n,\mathbb{R}), \end{aligned} \tag{16}$$

which is a variant of the Problem (5). Similar to Problems (5), and (7), and $d_{PD(n,\mathbb{R})}(.,.)$ refers to the distance on the manifold $PD(n,\mathbb{R})$ It is possible to define the constrained and multiple-objective versions of Problem (16) as well. Problem (16) and its variants have found numerous applications in machine learning and image processing such as diffusion tensor imaging (Dryden et al. 2009) and hyperbolic embeddings (Sala et al. 2018).

3. Historical Background

The roots of Riemannian optimization problems go back to Gabay (1982), where a projection gradient method is introduced to minimize a real-valued differentiable function defined over smooth manifolds. Afterward, Rapcsak (1991) studied the notion of geodesic convexity in Riemannian optimization problems, and Udriste (1993) provided a comprehensive study on convex optimization problems and algorithms on RMs. Smith (1995) extended two classical and famous methods, the Newton method and the conjugate gradient method to the Riemannian case and analyzed the convergence of these methods. Later on, Absil et al. (2004) defined a TR method for solving optimization problems on RMs with applications in numerical linear algebra. Subsequently, various versions of line-search algorithms and TR algorithms have been expanded to the aim of solving Riemannian optimization problems (e.g. Wang et al. 2021 and Hu et al. 2018). Successful implementation of optimization algorithms and the general framework that Riemannian geometry provides for the aim of investigating the convergence behavior of optimization methods, besides novel contributions in multiple-objective optimization problems, attracted attention to developing and solving MORO problems. In this regard, Bento et al. (2012) developed a gradient descent method for solving unconstrained MORO problems. Bento et al. (2012) restricted their work to the case of RMs whose sectional curvature is nonnegative. Ferreira et al. (2020) generalized the results of Bento et al. (2012) to the case that the sectional curvature of the RM is greater than any negative constant. Bento et al. (2013) studied an inexact version of the multiple-objective gradient descent method combined with the Armijo rule on RMs. Bento et al. (2013) studied a subgradient method for solving nonsmooth MORO problems. Restricting their work to Hadamard manifolds and locally Lipschitz vectorial functions, Bento et al. (2018) provided a proximal point method for MORO problems. Eslami et al. (2023) studied and analyzed a retraction-based TR method intending to solve MORO problems. Furthermore, Najafi and Hajarian (2023b) developed a Riemannian conjugate gradient method in the multiple-objective case, and also Najafi and Hajarian (2023a) extended the Riemannian BFGS method to the multiple-objective version. More details on each proposed topic are provided in Section 5.

4. Mathematical preliminaries

First of all, we start by considering an unconstrained multiple-objective optimization problem on Riemannian smooth manifold (M, g):

$$\begin{aligned} \min_{p} \quad & F(p) \\ \text{s.t.} \quad & p \in M. \end{aligned} \tag{17}$$

where $F : M \longrightarrow \mathbb{R}^m$ is a smooth function on (M, g), and $F^j : M \longrightarrow \mathbb{R}$ are its component functions. We will talk about the mathematical requirements and then we consider the constrained case of (17). Generally speaking, a smooth manifold M is locally similar to the n-dimensional Euclidean space $\mathbb{R}^n$ and admits a natural frame of differentiation that can be constructed in M. In order to find detailed features of a smooth manifold, see Lee (2012, Chap. 2). Considering smooth structures we are able to address the smoothness of functions defined on a manifold by its image under coordinate charts, and one may prefer to transform problem (17) to its equivalent problem in $\mathbb{R}^n$, then solve it by Euclidean optimization techniques. Since it is hard to choose appropriate coordinate charts from a given atlas, this approach fails and we go for some intrinsic ways. As we can see in Bergmann and Herzog (2019), this intrinsic viewpoint has been a particular strategy for constructing optimization problems on manifolds. Due to the fact that most applicable optimization methods which are based on first or second-order information of functions in the model, such as gradient descent and Newton method, provide a search direction (descent direction) to set up a new point, the tangent space at a point $p \in M$ is vital to be mentioned. The set of all tangent vectors at a point $p \in M$ form a vector space which is called the tangent space at p and is denoted by T_pM.

It is advantageous to note that by $v \in T_pM$, we mean there exists a smooth curve α around p generating v, see Lee (2012) for more details. In addition to smooth structure, there are other important structures on manifolds that make our work easier. Most of the optimization algorithms in manifold settings require manifold M to be a metric space. To do this, we need to take a usage of the Riemannian metric. A Riemannian metric g is an assignment of an inner product $g_p(.,.)$ on T_pM for each $p \in M$ that depends smoothly on p. we recall that g is not a metric, though it induces a natural distance function like d on M. It is also possible to define norm $\| . \|$ associated to Riemannian metric. In this regard, we call the pair (M,g) a smooth RM. To read more about the Riemannian metric, see Lee (2006, Chap. 3). In order to extend the notion of a Line segment into a Riemannian setting, we need to use the Levi-Civita connection denoted by ∇ associated to (M, g) which is an operator to take directional derivative from vector fields among vector fields. Considered as the Riemannian version of line segment, a *geodesic* is a smooth curve α with $\nabla_{\alpha'}\alpha' = 0$. Such a geodesic joining $p, q \in M$ would be named minimal if $\| \alpha' \| = d(p, q)$ and is denoted by $\alpha_{p,q}$. It is an important fact that in a complete RM, any two arbitrary points on M can be joined by a minimal geodesic. Also, based on the existence theorem of geodesics it is possible to state that for any given tangent vector $v \in T_pM$ there would be a unique geodesic α_v which satisfies the initial conditions $\alpha_v(0)$ and $\alpha'_v(0) = v$. Like the classical Euclidean optimization theory, the concept of convexity plays a key role in the Riemannian case. A subset C of (M, g) is called *geodesically convex* or *g-convex* if for any two points $p, q \in C$, there exists a geodesic $\alpha_{p,q}$ joining p and q, and totally contained in C. A function $f : C \subset M \longrightarrow \mathbb{R}$ is called geodesically convex or g-convex, if for any points $p, q \in C$, and for any geodesic $\alpha_{p,q}(\lambda)$ joining $\alpha_{p,q}(0) = p$ to $\alpha_{p,q}(1) = q$ contained in C, one has:

$$f(\alpha_{p,q}(\lambda)) \leq (1-\lambda)f(p) + \lambda f(q), \quad \forall \lambda \in [0,1], \tag{18}$$

Also, a vectorial function $F : M \longrightarrow \mathbb{R}^m$ where $F = (F^1, \ldots, F^m)$ is called $\mathbb{R}^m-$ g-convex if for all $i \in I$, the component function F^i is g- convex. To go further, we need to mention that the directional derivative of a function $f : M \longrightarrow \mathbb{R}$ at a point p with respect to $v \in T_pM$ is denoted by $\mathrm{D}f_p v$. By knowing this, the Riemannian gradient of f at $p \in M$ and for any $v \in T_pM$ is defined as:

$$G\,f(p)v = g_p(G\,f(p), v) = \mathrm{D}f_p v. \tag{19}$$

Also, the Hessian of f at $p \in M$ and for $v \in T_pM$ is defined as $H\,f(p)v = \nabla_v G\,f(p)$. The next theorem brings some famous equivalent conditions for the g-convexity of real-valued functions on RMs.

Theorem 1. *[Udriste 1993] Suppose that $C \subset M$ is a g-convex set and $f : C \subset M \longrightarrow \mathbb{R}^m$ is a smooth function. Then:*

1. *f is convex on C, if and only if for any points $p, q \in C$, and for any geodesic $\alpha_{p,q}(t)$ joining $\alpha_{p,q}(0) = p$ to $\alpha_{p,q}(1) = q$ contained in C and with $\alpha'_{p,q}(0) = v$, we have:*

$$f(q) \geqslant f(p) + g_p(G\,f(p), v), \quad \forall t \in [0,1], \tag{20}$$

2. *f is convex if and only if $H\,f(p)$ is positive definite for all $p \in C$.*

5. Proposed research topics

In this section, we bring the main topics of the research objectives. Parts of these proposed research topics are inspired by and aligned with our work on developing a trust-region method for MORO problems, as detailed in Eslami et al. (2023).

5.1. Constraint Qualifications and optimality conditions for MORO problems

As in the case of classical nonlinear optimization problems optimality conditions are suitable tools for developing optimization algorithms and investigating their convergence behavior in the case of multiple-objective optimization algorithms over RMs . Many research works have been based on defining constraint qualifications (CQ) and finding appropriate assumptions for obtaining optimality conditions in various cases of Euclidean multiple-objective optimization problems such as the smooth problems in Singh et al. (1987) and Maeda et al. (1994), the nonsmooth case in Li et al. (2000), semi-differentiable case in Preda et al. (1999), problems with vanishing constraints in Mishra et al. (2015), problems with switching constraints in Pandey et al. (2021), and problems with equilibrium constraints in Zhang et al. (2018). Recently, Haeser and Ramos (2020) defined the weakest CQs by taking advantage of the multiple-objective extension of normal cones. Also, Stein and Volk (2023) considered other generalizations of normal cones and introduced constraint qualifications in a way that local Pareto points are weak KT points. On the other hand, there exist recent research works on CQs and optimality conditions for optimization problems over RMs. For instance, Bergmann and Herzog (2019), generalized well-known LICQ, MFCQ, ACQ, and GCQ and showed their relations in the Riemannian case, i.e.:

$$\text{LICQ} \to \text{MFCQ} \to \text{ACQ} \to \text{GCQ}, \tag{21}$$

and also provided the connection of the mentioned CQs to the set of Lagrange multipliers for the constrained SORO problems. In a similar context on RMs, Andreani et al. (2023) study constant rank CQ (CRCQ), the constant positive linear independence CQ (CPLD), and their relaxed versions, RCRCQ, and RCPLD respectively. Additionally, Andreani et al. (2023) show that all limit points of a safeguarded augmented Lagrangian algorithm will satisfy the KKT conditions under all proposed CQs. Considering Hadamard manifolds, Upadhyay et al. (2023), studied generalized CQs for the nonsmooth multiple-objective problems. Our research aims to extend generalized multiple-objective normal cones in the case of RMs and study various types of CQs for both smooth and nonsmooth MORO problems.

Before going further, we need to recall some notations. Let $I = \{1,\dots,m\}$ and define:

$$\begin{aligned}
&\mathbb{R}^m_{\geqq} = \{x = (x^1,\dots,x^m) \in \mathbb{R}^m | x^i \geq 0;\ \forall i \in I\},\ \text{(the non-negative orthant of } \mathbb{R}^m\text{)}\\
&\mathbb{R}^m_{\geq} = \{x = (x^1,\dots,x^m) \in \mathbb{R}^m | x^i \geq 0;\ \forall i \in I, x \neq 0\},\\
&\mathbb{R}^m_{>} = \{x = (x^1,\dots,x^m) \in \mathbb{R}^m | x^i > 0;\ \forall i \in I\}.\ \text{(the positive orthant of } \mathbb{R}^m\text{)}
\end{aligned} \tag{22}$$

and for any $x, y \in \mathbb{R}^m$, we have:

$$\begin{aligned}
x \leqq y &\iff \forall i \in I \quad y^i - x^i \geq 0,\\
x \leq y &\iff \forall i \in I \quad y^i - x^i \geq 0,\ x \neq y,\\
x < y &\iff \forall i \in I \quad y^i - x^i > 0.
\end{aligned} \tag{23}$$

Using the above order, now we recall the concept of efficiency in MORO problems. To see comprehensive details of efficiency in the Euclidean framework, see Ehrgot (2005, Chap. 2).

Definition 1. *Consider Problem (17). A point $p^* \in M$ is called*

1. *an* efficient *point, if there is not any $q \in M$ such that:*

$$F(q) \leq F(p^*). \tag{24}$$

2. *a* weak efficient *point, if there is not any $q \in M$ such that:*

$$F(q) < F(p^*). \tag{25}$$

3. *a* local efficient *point, if there exists a neighbourhood $S \subset M$ where there is not any $q \in S$ such that:*

$$F(q) \leq F(p^*). \tag{26}$$

From definition (1) we can simply deduce that an efficient point is also a weak efficient point, but vice versa does not hold. As we mentioned before, it is a normal idea to change a smooth manifold optimization problem into its corresponding Euclidean one. Since at first glance, most of such problems are living in an Euclidean space, it seems essential to show that modeling and considering those problems in manifold version does not drop out efficiency. To see this, we state the next theorem. The single variable one can be found in Yan et al. (2014) and Bergmann and Herzog (2019).

Theorem 2. *Let M be a smooth manifold, and consider $p^* \in M$. If (U, ϕ) is any coordinate chart around p^* (see Lee 2006 for the definition of a chart), then following terms are equivalent:*

1. *$\phi(p^*)$ is a local efficient point of problem below:*

$$\begin{aligned} &\min_{x} \ \ (F\, o\, \phi^{-1})(x) \\ &\text{s.t.} \ \ x \in \phi(U). \end{aligned} \tag{27}$$

Where $(F\, o\, \phi^{-1})(x) = ((F^1\, o\, \phi^{-1})(x), \ldots, (F^m\, o\, \phi^{-1})(x))$.

2. *p^* is a local efficient point of Problem (17).*

For the proof of Theorem (2), see Appendix.

Similar to the Euclidean case, the Jacobian operator, is a crucial part of optimality conditions for vectorial functions. For a function $F : M \longrightarrow \mathbb{R}^m$, where $F(p) = (F^1(p), \ldots, F^m(p))$, we denote the Jacobian operator of F by:

$$J\, F(p) = [G\, F^1(p), \ldots, G\, F^m(p)], \quad \forall p \in M. \tag{28}$$

Where the image set of the operator at each point p, is defined as:

$$R(J\, F(p)) = \{(J\, F(p))(v) = [g_p(G\, F^1(p), v), \ldots, g_p(G\, F^m(p), v)] \mid v \in T_pM\}. \tag{29}$$

In the next definition, we mention the notion of critical points using the Jacobian matrix. For the Euclidean version of this definition, see Carrizo et al. (2016). The Riemannian version can be also found in Bento et al. (2013).

Definition 2. *Consider Problem (17).*

1. *A feasible point p^* is said to be* critical *if:*

$$R(J\, F(p^*)) \cap (-\mathbb{R}^m_{++}) = \emptyset, \tag{30}$$

where the $J\, F(p^*))$ is the Jacobian of function F.

2. *At an arbitrary point $p \in M$, the tangent vector $v_p \in T_pM$ is called a* descent direction *if:*

$$g_p(G\, F^i, v_p) \leq 0; \ \ \forall i \in I. \tag{31}$$

By applying the above definition, the next result is concluded:

Corollary 1. *Consider Problem (17), the point $p^* \in M$ is critical if and only if, there is no $v_{p^*} \in T_{p^*}M$ such that:*

$$g_{p^*}(G\,F^i, v_{p^*}) \leq 0;\ \ \forall i \in I. \tag{32}$$

In the next theorem, we bring the relation between efficient points and critical points in general MORO problems. See Carrizo et al. (2016) for the Euclidean version. To read similar results with different points of view and ways of proof, see Bento et al. (2013).

Theorem 3. *Consider Problem (17):*

(i) If p^* is a local weak efficient point, then p^* is a critical point for F.

(ii) If F is $\mathbb{R}^m-$ geodesically convex and p^* is critical for F, then p^* is a weak efficient point.

(iii) Suppose F is a smooth function on M, and the Hessians of F^j are positive definite for all p and for all j. If $p^* \in M$ is critical point for F, then p^* is an efficient point.

For the proof of Theorem (3), see Appendix.

In this part, we bring the definition of a properly efficient point for the constrained MORO problem formulated as follows:

$$\begin{aligned} \min_{p}\ \ & F(p) = (F^1(p), \ldots, F^m(p)) \\ \text{s.t.}\ \ & H(p) = (H^1(p), \ldots, H^m(p)) \leqslant 0, \\ & p \in M. \end{aligned} \tag{33}$$

Where $F: M \longrightarrow \mathbb{R}^m$, and $H: M \longrightarrow \mathbb{R}^m$ are smooth functions on manifold M. The feasible set of Problem (33) is:

$$\Omega = \{p \in M \mid H^j(p) \leq 0,\ \forall j \in I\}. \tag{34}$$

Also $I_0(q) = \{j \in I \mid H^j(q) = 0\}$ is defined as the active set of (33). The Euclidean version of the definition of properly efficient points for (33) can be found in Ehrgot (2005, Chap. 2), and the Riemannian version is in Chen et al. (2014).

Definition 3. *Consider Problem (33). A feasible point $p^* \in \Omega$ is called properly efficient in the sense of Geoffrion if it satisfies the condition of efficiency in Definition (1), and if there is a scalar $M > 0$ such that for all i and $p \in \Omega$ with $F^i(p) < F^i(p^*)$ there exists an index j such that $F^j(p^*) < F^j(p)$ in a way that*

$$\frac{F^i(p^*) - F^i(p)}{F^j(p) - F^j(p^*)} \leq M. \tag{35}$$

In the next, we recall Kuhn and Tucker's definition of Kuhn (2013) for the Problem (33).

Definition 4. *Consider Problem (33). A feasible point $p^* \in \Omega$ is called properly efficient in the sense of Kuhn and Tucker if it satisfies condition (1) of efficiency and if there is no $v \in T_{p^*}M$ with following properties*

$$g_{p^*}(G\,F^k(p^*), v) \leq 0, \quad \forall k \in I, \tag{36}$$

$$g_{p^*}(G\,F^i(p^*), v) < 0, \quad \text{for some } i \in I, \tag{37}$$

$$g_{p^*}(G\,H^j(p^*), v) \leq 0, \quad \forall j \in I_0(p^*). \tag{38}$$

In the following, we bring the concept of Kuhn and Tucker's constraint qualification. The Euclidean version of this definition can be found in Ehrgot (2005, Def. 2.49).

Definition 5. *The MORO Problem (33) satisfies the Kuhn and Tucker's constraint qualification at point $p^* \in \Omega$ if for all $v \in T_{p^*}M$ with $g_p(G\,H^j(p^*), v) \leq 0$ for all $j \in I_0(p^*)$, there exists a scalar $\epsilon > 0$, a smooth curve $\beta : [0, \epsilon] \to M$, and $\mu > 0$ such that $\beta(0) = p^*$, and $H(\beta(t)) \leqslant 0$ for all $t \in [0, \epsilon]$, and $\beta(0) = \mu v$.*

Now, by considering definitions (3), and (4) and using KTCQ (5), we show that a properly efficient point p^* for the Problem (33) is properly efficient in the sense of (4). For the Euclidean version of this theorem, see Geoffrian (1968).

Theorem 4. *Consider Problem (33). If this problem satisfies KTCQ conditions (5) at p^* and p^* is properly efficient by (3), then it would be properly efficient in the meaning of (4).*

For the proof of Theorem (4), see Appendix.

5.2. A retraction-based nonsmooth TR method for solving nonsmooth MORO problems

By using the notion of retractions on RMs, this research aims to develop a nonsmooth version of the TR algorithm for solving MORO problems. TR methods are a type of well-known and classical method for solving nonlinear unconstrained optimization problems in the Euclidean case (see Conn et al. (2000) for the single-objective problems). Regarding multiple-objective optimization problems, Carrizo et al. (2016) developed the TR method as a parameter-free scalarization approach to solve such problems in the Euclidean case, and Qu et al. (2013) studies a similar scheme for the nonsmooth vectorial problems. As an extension to the Riemannian case, the TR method for SORO problems has been successfully studied by Absil et al. (2004). Moreover, inspired by Carrizo et al. (2016) and Absil et al. (2004), Eslami et al. (2023) studied the TR method for the smooth MORO problems. Here, we provide an overview of the smooth multiple-objective Riemannian TR (MORTR) by Eslami et al. (2023).

5.2.1. Smooth TR method on RMs The classical TR method for unconstrained minimization of a smooth single-objective function f defined on $\mathbb{R}^n$, as in Conn et al. (2000), is based on updating current iteration $x \in \mathbb{R}^n$ by adding a new vector $v \in \mathbb{R}^n$. This new vector v is obtained by solving the following optimization problem called the TR subproblem:

$$\begin{aligned} \min_{v} \quad & T(v) = f(x) + \nabla^T v + \frac{1}{2} v \nabla^2 v \\ s.t. \quad & \| v \| \leq \Delta, \\ & v \in \mathbb{R}^n, \end{aligned} \tag{39}$$

The TR subproblem (39) minimizes the second-order Taylor expansion of function f as an approximation to the original function f and is constrained to vectors falling within $\| v \| \leq \Delta$, which is called a trust region and Δ is called the corresponding TR radius. The parameter Δ is initially set at the beginning and then adjusted at each iteration because it affects the quality of the method. The quality of the method is examined using the following criteria:

$$\rho = \frac{f(x) - f(x+v)}{T(0) - T(v)}, \tag{40}$$

Based on the value of ρ in (40), the new point is going to be accepted or rejected, and then the TR radius is updated. In the case of rejection, the Subproblem (39) would be solved again by considering the new TR radius as in Conn et al. (2000). In the Riemannian context, it is not possible to perform the operation $x + v$ on a general manifold (M, g), as they might not have a linear structure combined with an additive operation. To deal with the challenge, the concept of exponential mapping or its generalization which is called a retraction mappings is used instead. Indeed, at each step of solving the Riemannian subproblem at the current point $p \in M$, an optimal vector v is obtained which lies on the tangent space $p \in M$, not on the manifold M. This is where the difference between a classical Euclidean optimization algorithm and a Riemannian optimization algorithm appears. In the Euclidean context, at the current point $x \in \mathbb{R}^n$, the direction vector to the next iteration lies in $\mathbb{R}^n$. In the Riemannian setting, such a vector lives on the tangent space and is required to be projected to a manifold in order to find the next iteration. The mentioned projection can be done using exponential or retraction mappings (see Picture (1)). The definition of a retraction mapping is as follows and also can be found in Absil et al. (2008, Chap. 4).

Definition 6. *At any point $p \in M$, a mapping $Ret_p : T_pM \to M$ is called a retraction mapping if it satisfies the following conditions:*

1. *For $0_p \in T_pM$, $Ret_p(0_p) = p$.*
2. *$DRet_p(0_p) = id_{T_pM}$, where id_{T_pM} is the identity mapping on T_pM.*

Example 4. *[Boumal et al. 2023, Chap. 5] Consider manifolds $\mathbb{S}^2$, $PD(n,\mathbb{R})$, $St(m,n,\mathbb{R})$, and $SO(n,\mathbb{R})$. A possible choice for $\mathbb{S}^2$ is:*

$$\mathrm{Ret}_p(v) := \frac{p+v}{\| p+v \|}. \tag{41}$$

For $PD(n,\mathbb{R})$, a possible choice of retraction is:

$$\mathrm{Ret}_U(W) = U + W + \frac{1}{2} W U^{-1} W, \quad \forall U \in \mathbb{U}^n_{++},\ \forall W \in T_U \mathbb{U}^n_{++}. \tag{42}$$

For $St(m,n,\mathbb{R})$, the polar retraction is defined as:

$$\mathrm{Ret}_U(W) = (U+W)\big((I_n + V^T V)^{-1}\big)^{1/2} \tag{43}$$

Moreover, since $SO(n,\mathbb{R})$ is submanifold of $St(n,n,\mathbb{R})$, then the retraction (43) is working for $SO(n,\mathbb{R})$, too.

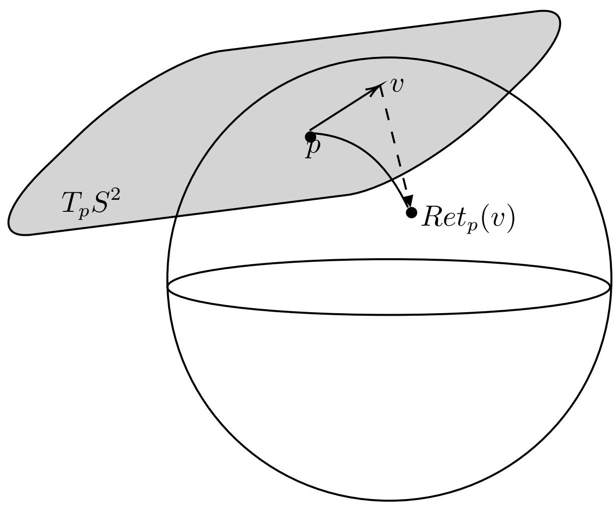

Figure 1 **A retraction on $\mathbb{S}^2$ mapping vector $v \in T_p\mathbb{S}^2$ to $\mathbb{S}^2$.**

In addition to finding the direction to the next iteration at each step of the Riemannian TR method, it is possible to transform a function $f : M \to \mathbb{R}$ by using retractions. For any point $p \in M$, the function $f\, o\, Ret_p$ is defined on T_pM and makes it appropriate to use the advantages of the vector space structure of T_pM. A Riemannian TR method for solving the SORO problem of minimizing the smooth function $f : M \to \mathbb{R}$, involves a Riemannian TR subproblem, which at current iteration $\in M$ is defined as (see in Absil et al. 2004):

$$\begin{aligned} \min_v \quad & M(v) = f(p) + G\, f(p) v + \frac{1}{2} g_p(H\, f(p) v, v) \\ s.t. \quad & g_p(v,v) \le \Delta^2, \\ & v \in T_P M. \end{aligned} \tag{44}$$

The other steps of the Riemannian TR method are similar to the Euclidean one. It is possible to extend the TR method to the multiple-objective optimization problems. Qu et al. (2013), developed a TR method to solve such problems in the Euclidean setting. Eslami et al. (2023) studied a retraction-based TR method for solving MORO problems. we describe a scheme of the method developed By Eslami et al. (2023). They consider a retraction Ret_p and can derive the TR subproblem at the kth iteration for the MORO Problem (17) on (M, g). The approximation of the component function F^i would be:

$$\begin{aligned} M_k^i(v) = {} & F^i(\mathrm{Ret}_{p_k}(0_{p_k})) + g_{p_k}(\mathrm{grad}\, F^i(\mathrm{Ret}_{p_k}(0_{p_k})), v) \\ & + \frac{g_{p_k}(\mathrm{Hess}\, F^i(\mathrm{Ret}_{p_k}(0_{p_k}))v, v)}{2}. \end{aligned} \tag{45}$$

So, the objective function of the TR subproblem is:

$$\begin{aligned} M_k^{\max}(v) & := \max_{i \in I} M_k^i(v) \\ & = \max_{i \in I} \Big\{ F^i(\mathrm{Ret}_{p_k}(0_{p_k})) \\ & \quad + g_{p_k}(\mathrm{grad}\, F^i(\mathrm{Ret}_{p_k}(0_{p_k})), v) + \frac{g_{p_k}(\mathrm{Hess}\, F^i(\mathrm{Ret}_{p_k}(0_{p_k}))v, v)}{2} \Big\}. \end{aligned} \tag{46}$$

Then, as in Eslami et al. (2023) the Riemannian TR subproblem at iteration k becomes:

$$\begin{aligned} \min_{v} \quad & M_k^{\max}(v) \\ \text{s.t.} \quad & g_{p_k}(v, v) \leq \Delta_k^2, \\ & v \in T_{p_k} M. \end{aligned} \tag{47}$$

Similarly to the $\mathbb{R}^n$ case, instead of solving (47), it is preferred to solve the equivalent problem in the Euclidean vector space $\mathbb{R} \times T_{p_k} M$, which is:

$$\begin{aligned} \min_{r, v} \quad & A(r, v) := r \\ \text{s.t.} \quad & g_{p_k}(v, v) \leq \Delta_k^2, \\ & M_k^i(v) \leq r, \quad \forall i \in I, \\ & g_{p_k}(\mathrm{grad} F^i(\mathrm{Ret}_{p_k}(0_{p_k})), v) \leq r, \quad \forall i \in I. \end{aligned} \tag{48}$$

Eventually, the MORTR algorithm for solving a MORO problem over (M, g) would be (see Eslami et al. 2023):

Algorithm 1 The MORTR algorithm by Eslami et al. (2023)

Step 1 . Set initial parameters: An initial point p_0, an initial TR radius Δ_0, and parameters $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$ with conditions $0 < \alpha_1 < \alpha_2 < 1 < \alpha_3$, $0 < \beta_1 \leq \beta_2 < 1$, and $\xi > 0$. Also, set $k = 0$.

Step 2 . Solve the TR Subproblem (48). **If** $|A_k| < \xi$, **then** stop.

Step 3 . Compute the quotient (criteria) for any $j \in I$:

$$\rho_k^j = \frac{F^j(\mathrm{Ret}_{p_k}(0)) - F^j(\mathrm{Ret}_{p_k}(v_k))}{M_k^j(0) - M_k^j(v_k)}. \tag{49}$$

Step 4 . Check the quality of the model, based on the criteria from Step 2. **If** $\min_j\{\rho_k^j\} \geq \beta_2$, **then** update:

$$p_k = \mathrm{Ret}_{p_k}(v_k), \quad \Delta_{k+1} = \alpha_3 \Delta_k.$$

Else if $\rho_k^j \geq \beta_1$, for all $j \in I$, and there exists a $j_0 \in I$, with $\rho_k^{j_0} \leq \beta_2$, **then**:

$$p_{k+1} = \mathrm{Ret}_{p_k}(v_k), \quad \Delta_{k+1} = \alpha_2 \Delta_k.$$

Otherwise update $\Delta_{k+1} = \alpha_1 \Delta_k$, and $p_{k+1} = p_k$. Update k as $k+1$ and return to Step 1.

To the best of our knowledge, the nonsmooth version of the TR method for the nonsmooth MORO problem has not been studied yet. Motivating this research topic, Grohs et al. (2016) developed a nonsmooth TR method for SORO problems. Considering the nonsmooth version of the TR method for SORO problems in Grohs and Hosseini (2016) and the nonsmooth multiple-objective TR method in the Euclidean setting (Qu et al. 2013), our target in this proposal is to expand the method introduced in Eslami et al. (2023) to a nonsmooth version and examine its numerical performance and theoretical convergence analysis.

5.3. Stochastic TR method on RMs and stochastic MORO problems

The TR method for solving SORO problems is one of the classical methods that was studied first by Absil et al. (2004). Many interesting variants of this method have been extended for solving SORO problems, such as quasi-newton TR methods (Wei et al. 2023, Huang et al. 2015, Wei et al. 2016, and Huang et al. 2022), the nonsmooth TR method (Grohs et al. 2016, and the inexact versions of TR method (Zhao et al. 2023a, and Kasai et al. 2018). All the mentioned types of TR methods are used for solving deterministic, and it is our aim to consider the stochastic version of TR methods over RMs. A stochastic optimization problem on RMs is defined as the following problem:

$$\begin{aligned} \min_{p} \quad & \mathbb{E}_{\varepsilon}[\tilde{f}(p, \varepsilon)] \\ s.t. \quad & p \in M, \end{aligned} \tag{50}$$

where $\tilde{f}$ is a noisy version of a smooth function f with noise ε which follows a normal distribution. In the Euclidean case $M = \mathbb{R}^n$, various methods have been developed to solve problem (50). As one of the first works in this area, the Riemannian version of stochastic gradient descent is studied by Bonnabel (2013). Furthermore, Li et al. (2023) generalized a randomized stochastic gradient descent for solving (50). Due to the importance of the TR method and its wide applications in machine learning and nonconvex optimization, the stochastic version of the TR method has been considered to be implemented on Euclidean variants of problem (50), i.e. when $M = \mathbb{R}^n$. In this regard, Gratton et al. (2018) and Xu et al. (2020) considered a deterministic case of (50) and developed the TR method using stochastic estimation of gradient and Hessian. For the purpose of stochastic optimization, Chang et al. (2013), and Shashaani et al. (2018) designed stochastic TR methods. Chen et al. (2018) studied a variance-reduced TR method and analyzed its convergence to stationary points. Larson et al. (2016) slightly modified the method in Chen et al. (2018). Blanchet et al. (2019) provided a new stochastic variant of the TR method which is similar to Chen et al. (2018), and Larson et al. (2016) introduced a framework that considers the underlying stochastic process of the stochastic algorithm and found bounds on the stopping time of the process. We aim to expand the stochastic TR method in Blanchet et al. (2019) to the Riemannian Problem (50). In the next step, our goal is to extend such stochastic TR method to the multiple-objective case and to take advantage of randomization in the stochastic method for stochastic optimization problems over RMs. The general formulation of a stochastic multiple-objective optimization problem in $\mathbb{R}^n$ is defined as (see Fliege and Xu 2011):

$$\begin{aligned} \min_{x} \quad & (\mathbb{E}_{\varepsilon}[\tilde{f}_1(x,\varepsilon)], \ldots, \mathbb{E}_{\varepsilon}[\tilde{f}_n(x,\varepsilon)]) \\ s.t. \quad & x \in \mathbb{R}^n, \end{aligned} \tag{51}$$

where $f_i : \mathbb{R}^n \times \mathbb{R}^L \longrightarrow \mathbb{R}$ is a smooth function, $\varepsilon \in \mathbb{R}^L$ is a random vector, and $x \in \mathbb{R}^n$ is the decision variable. The stochastic multiple-objective optimization problems have been widely studied in the Euclidean setting due to their vast applications (see Fliege and Xu 2011, Bonnel et al. 2014, Gutjahr et al. 2016, Poirion et al. 2017, Liu et al. 2022, Zhou et al. 2022, and Zhao et al. 2023b). Recently, Zhao et al. (2023b) introduced a novel algorithm for stochastic multiobjective optimization. At each iteration of the algorithm (see Zhao et al. 2023b, Section 2), it approximates the true gradients of the objective functions using a sample average gradient, calculated from a set of random samples. We aim to first define the stochastic version of MORO problems and then develop and investigate the numerical and analytical behavior of existing Euclidean methods in the Riemannian case. Moreover, our final aim is to study the multiple-objective stochastic TR method in the Riemannian case. As the Euclidean space $\mathbb{R}^n$ is an RM itself, in addition to Riemannian optimization problems, such an extension can be applied to the problem in the Euclidean setting.

Appendix

Here, we bring the proof of Theorems (2),(3), (4), respectively.

Proof of Theorem (2)

Proof. Firstly, suppose that (U, ϕ) is a coordinate chart of M containing p^* in which $x^* = \phi(p^*)$ and p^* is a local efficient point for Problem (17). So there exists an open subset $V \subset M$ such that there is no $q \in V \cap U$ in which $F(q) \leq F(p^*)$. Equivalently, it means that for all $q \in V \cap U$, there exists $j_0^q \in I$ such that $F_{j_0^q}(q) \geq F_{j_0^q}(p^*)$ and also there exists $j_1^q \in I$ such that $F_{j_1^q}(q) \neq F_{j_1^q}(p^*)$. Since ϕ is a homeomorphism on U, then for $q \in V \cap U$ and $j_0^q, j_1^q \in I$ we have:

$$\begin{aligned} &(F^q_{j_0}\, o\, \phi^{-1})(\phi(q)) \geq (F^q_{j_0}\, o\, \phi^{-1})(\phi(p^*)), \\ &(F^q_{j_1}\, o\, \phi^{-1})(\phi(q)) \neq (F^q_{j_1}\, o\, \phi^{-1})(\phi(p^*)). \end{aligned} \tag{52}$$

Where $\phi(q), \phi(p^*) \in \phi(V \cap U) \subset \mathbb{R}^m$. From (52) we can say that there is no $y \in \phi(V \cap U)$ such that $(F\, o\, \phi^{-1})(y) \leq (F\, o\, \phi^{-1})(x^*)$. it means that $x^* = \phi(p^*)$ is a local efficient point for Problem (27). We can use a similar way to prove vice versa. □

Proof of Theorem (3)

Proof. (i) By contradiction, suppose that p^* is not a critical point. Then, there exists a tangent vector $v_p \in T_{p^*}M$ in which for every $j \in I$ we have $G\, F^j(p^*)v < 0$. By definition of the directional derivative, we have:

$$G\, F^j(p^*)v_{p^*} = \lim_{t \to 0} \frac{F^j(\alpha(t)) - F^j(\alpha(o))}{t} < 0. \tag{53}$$

Where $\alpha(t)$ is a smooth curve around p^* that satisfies initial conditions $\alpha(0) = p^*$ and $\alpha'(0) = v_{p^*}$. So from (53) a $t_0 > 0$ can be found such that for all $t \in (0, t_0]$ we have:

$$F^j(\alpha(t)) - F^j(\alpha(0)) < 0,$$

In conclusion, if for a $t \in (0, t_0]$ we put $q := \alpha(t)$ then we have $F(q) < F(p^*)$, which contradicts the weak efficiency assumption of p^*.

(ii) Let F be a $\mathbb{R}^m-$ geodesically convex function, and $p^* \in M$ a critical point of F. Consider any point $q \in M$. Due to the completeness of M, there exists a minimal geodesic $\alpha_{p^*,q}(t)$ around p^* joining p^* to q. It means that by a parametrization we can say $\alpha_{p^*,q}(0) = p^*$, $\alpha_{p,q}(1) = q$, and $\alpha'_{p^*,q}(0) = v_{p^*}$. Because p^* is a critical point, then by definition (2) we have:

$$R(J\, F(p^*)) \cap (-\mathbb{R}^m_{++}) = \emptyset. \tag{54}$$

Then, for any vector $u \in T_{p^*}M$ there exists an index $j \in I$ such that $g_p(G\, F^j(p), u) \geqslant 0$. So for $v_{p^*} \in T_{p^*}M$ there exists an index $j^{p^*} \in I$ such that $g_{p^*}(G\, F^{j^{p^*}}(p^*), v_{p^*}) \geqslant 0$. Thus, by first-order theorem of g-convex functions (1), we have:

$$F^{j^{p^*}}(q) \geqslant F^{j^{p^*}}(p^*) + g_{p^*}(G\, F^{j^{p^*}}(p^*), v_{p^*}) \geqslant F^{j^{p^*}}(p), \tag{55}$$

So for any $q \in M$, we are able to find $j^{p^*} \in I$ with $F^{j^{p^*}}(q) \geq F^{j^{p^*}}(p)$. Equivalently, there is no $q \in M$ with $F(q) < F(p^*)$. It means that p^* is a weakly efficient point.

(iii) Because of the positive definiteness of Hessians of F^j for any $j \in I$, we can say that they are strictly geodesically convex. Thus, there exists a minimal geodesic $\alpha_{p,q}(t)$ around a critical point $p \in M$ joining it to any point $q \in M$, with $\alpha_{p,q}(0) = p$, $\alpha_{p,q}(1) = q$, and $\alpha'_{p,q}(0) = v_p$. Similar to part (ii) there exists an index $j^p \in I$ such that:

$$F^{j^p}(q) > F^{j^p}(p^*) + g_p(G\,F^{j^p}(p), v_p) \geqslant F^{j^p}(p),$$

So for any $q \in M$ there exists a $j \in I$ such that $F^j(q) > F^J(p)$. It means that there is no q with $F(q) \leq F(p)$. Consequently, p would be an efficient point. $\square$

Proof of Theorem (4)

Proof. We prove this theorem by the contraposition technique. Suppose that p^* is properly efficient in the sense of definition (3). Also, suppose that it is not properly efficient in the meaning of definition (4) by Kuhn and Tucker. It gives that there exists a tangent vector $v \in T_pM$, and index $s \in I$ such that:

$$g_{p^*}(G\,F^s(p^*), v) < 0, \tag{56}$$

$$g_{p^*}(G\,F^k(p^*), v) \leq 0, \quad \forall k \in I - \{s\}, \tag{57}$$

$$g_{p^*}(G\,H^j(p^*), v) \leq 0, \quad \forall j \in I_0(p^*). \tag{58}$$

Since it is assumed Problem (33) satisfies KTCQ conditions in (5), there can be found a curve β with KTCQ's properties. Consider a sequence $\epsilon_k \to 0$ and define:

$$\Phi = \{j \in I \mid F^j(\beta(\epsilon_k)) > F^j(p^*)\},$$

By applying the Taylor expansion theorem on F^j at $\beta(\epsilon_k)$ for $j \in I$, we have:

$$F^j(\beta(\epsilon_k)) = F^j(p^*) + t_k g_{p^*}(G\,F^j(p^*), \epsilon v) + O(\epsilon_k).$$

Because $F^j(\beta(\epsilon_k) = F^j(p^*) > 0$ for $j \in \Phi$ and $g_{p^*}(G\,F^j(p^*), v) \leq 0$, we will have $g_{p^*}(G\,F^j(p^*), v) = 0$, for all $j \in \Phi$. However, due to the fact that $g_p(G\,F^s(p^*), v) < 0$, it gives:

$$\frac{F^s(p^*) - F^s(\beta(\epsilon_k))}{F^j(\beta(\epsilon_k)) - F^j(p^*)} \to \infty, \quad for \;\; j \in \Phi, \tag{59}$$

It means that p^* is not properly efficient based on definition (3) and it contradicts the hypothesis. $\square$

References

Absil, P-A, Robert Mahony, Rodolphe Sepulchre. 2008. *Optimization algorithms on matrix manifolds*. Princeton University Press.

Absil, Pierre-Antoine, Christopher G Baker, Kyle A Gallivan. 2004. Trust-region methods on riemannian manifolds with applications in numerical linear algebra. *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Leuven, Belgium*. Citeseer, 5–9.

Agrawal, Raj, Uma Roy, Caroline Uhler. 2022. Covariance matrix estimation under total positivity for portfolio selection. *Journal of Financial Econometrics* **20**(2) 367–389.

Andreani, Roberto, Kelvin Rodrigues Couto, Orizon Pereira Ferreira, Gabriel Haeser. 2023. Constraint qualifications and strong global convergence properties of an augmented lagrangian method on riemannian manifolds. *arXiv preprint arXiv:2306.14345* .

Axen, Seth D, Mateusz Baran, Ronny Bergmann, Krzysztof Rzecki. 2023. Manifolds. jl: an extensible julia framework for data analysis on manifolds. *ACM Transactions on Mathematical Software* **49**(4) 1–23.

Bacák, Miroslav, Ronny Bergmann, Gabriele Steidl, Andreas Weinmann. 2016. A second order nonsmooth variational model for restoring manifold-valued images. *SIAM Journal on Scientific Computing* **38**(1) A567–A597.

Bachmann, Florian, Ralf Hielscher, Helmut Schaeben. 2011. Grain detection from 2d and 3d ebsd data—specification of the mtex algorithm. *Ultramicroscopy* **111**(12) 1720–1733.

Barillas, Francisco, Jay Shanken. 2018. Comparing asset pricing models. *The Journal of Finance* **73**(2) 715–754.

Barshan, Elnaz, Ali Ghodsi, Zohreh Azimifar, Mansoor Zolghadri Jahromi. 2011. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition* **44**(7) 1357–1371.

Bento, GC, JX Cruz Neto. 2013. A subgradient method for multiobjective optimization on riemannian manifolds. *Journal of Optimization Theory and Applications* **159** 125–137.

Bento, GC, João Xavier da Cruz Neto, PSM Santos. 2013. An inexact steepest descent method for multicriteria optimization on riemannian manifolds. *Journal of Optimization Theory and Applications* **159** 108–124.

Bento, GC, Orizon Pereira Ferreira, P Roberto Oliveira. 2012. Unconstrained steepest descent method for multicriteria optimization on riemannian manifolds. *Journal of Optimization Theory and Applications* **154** 88–107.

Bento, Glaydston de C, João Xavier da Cruz Neto, Lucas V de Meireles. 2018. Proximal point method for locally lipschitz functions in multiobjective optimization of hadamard manifolds. *Journal of Optimization Theory and Applications* **179** 37–52.

Bergmann, Ronny, Raymond H Chan, Ralf Hielscher, Johannes Persch, Gabriele Steidl. 2016. Restoration of manifold-valued images by half-quadratic minimization. *Inverse Problems and Imaging* **10**(2) 281–304.

Bergmann, Ronny, Roland Herzog. 2019. Intrinsic formulation of kkt conditions and constraint qualifications on smooth manifolds. *SIAM Journal on Optimization* **29**(4) 2423–2444.

Bertsimas, Dimitris, Driss Lahlou Kitane. 2023. Sparse pca: A geometric approach. *Journal of Machine Learning Research* **24**(32) 1–33.

Blanchet, Jose, Coralia Cartis, Matt Menickelly, Katya Scheinberg. 2019. Convergence rate analysis of a stochastic trust-region method via supermartingales. *INFORMS journal on optimization* **1**(2) 92–119.

Bonnabel, Silvere. 2013. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control* **58**(9) 2217–2229.

Bonnel, Henri, Julien Collonge. 2014. Stochastic optimization over a pareto set associated with a stochastic multi-objective optimization problem. *Journal of Optimization Theory and Applications* **162** 405–427.

Bouguila, Nizar, Wentao Fan. 2020. *Mixture models and applications*. Springer.

Boumal, Nicolas. 2023. *An introduction to optimization on smooth manifolds*. Cambridge University Press. doi:10.1017/9781009166164. URL https://www.nicolasboumal.net/book.

Boyd, Stephen, Lin Xiao. 2005. Least-squares covariance matrix adjustment. *SIAM Journal on Matrix Analysis and Applications* **27**(2) 532–546.

Bürgmann, Roland, Paul A Rosen, Eric J Fielding. 2000. Synthetic aperture radar interferometry to measure earth's surface topography and its deformation. *Annual review of earth and planetary sciences* **28**(1) 169–209.

Carrizo, Gabriel A, Pablo A Lotito, María C Maciel. 2016. Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Mathematical Programming* **159** 339–369.

Chan, Tony F, Sung Ha Kang, Jianhong Shen. 2001. Total variation denoising and enhancement of color images based on the cb and hsv color models. *Journal of Visual Communication and Image Representation* **12**(4) 422–435.

Chang, Kuo-Hao, L Jeff Hong, Hong Wan. 2013. Stochastic trust-region response-surface method (strong)—a new response-surface framework for simulation optimization. *INFORMS Journal on Computing* **25**(2) 230–243.

Chen, Ruobing, Matt Menickelly, Katya Scheinberg. 2018. Stochastic optimization using a trust-region method and random models. *Mathematical Programming* **169** 447–487.

Chen, Sheng-lan, Nan-Jing Huang, Donal O'Regan, et al. 2014. Geodesic b-preinvex functions and multiobjective optimization problems on riemannian manifolds. *Journal of Applied Mathematics* **2014**.

Chen, Shixiang, Shiqian Ma, Lingzhou Xue, Hui Zou. 2020. An alternating manifold proximal gradient method for sparse principal component analysis and sparse canonical correlation analysis. *INFORMS Journal on Optimization* **2**(3) 192–208.

Choi, Young-Geun, Johan Lim, Anindya Roy, Junyong Park. 2019. Fixed support positive-definite modification of covariance matrix estimators via linear shrinkage. *Journal of Multivariate Analysis* **171** 234–249.

Conn, Andrew R, Nicholas IM Gould, Philippe L Toint. 2000. *Trust region methods*. SIAM.

Dempster, Arthur P, Nan M Laird, Donald B Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)* **39**(1) 1–22.

Diepeveen, Willem, Jan Lellmann. 2021. An inexact semismooth newton method on riemannian manifolds with application to duality-based total variation denoising. *SIAM Journal on Imaging Sciences* **14**(4) 1565–1600.

Drummond, Tom, Roberto Cipolla. 2002. Real-time visual tracking of complex structures. *IEEE Transactions on pattern analysis and machine intelligence* **24**(7) 932–946.

Dryden, Ian L., Alexey Koloydenko, Diwei Zhou. 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics* **3**(3) 1102 – 1123. doi:10.1214/09-AOAS249. URL https://doi.org/10.1214/09-AOAS249.

Ehrgott, Matthias. 2005. *Multicriteria optimization*, vol. 491. Springer Science & Business Media.

Eslami, N, B Najafi, SM Vaezpour. 2023. A trust region method for solving multicriteria optimization problems on riemannian manifolds. *Journal of Optimization Theory and Applications* **196**(1) 212–239.

Ferreira, Orizon P, Mauricio S Louzeiro, Leandro F Prudente. 2020. Iteration-complexity and asymptotic analysis of steepest descent method for multiobjective optimization on riemannian manifolds. *Journal of Optimization Theory and Applications* **184** 507–533.

Fliege, Jörg, Huifu Xu. 2011. Stochastic multiobjective optimization: sample average approximation and applications. *Journal of optimization theory and applications* **151** 135–162.

Gabay, Daniel. 1982. Minimizing a differentiable function over a differential manifold. *Journal of Optimization Theory and Applications* **37** 177–219.

Gang, Arpita, Waheed U Bajwa. 2022. Fast-pca: A fast and exact algorithm for distributed principal component analysis. *IEEE Transactions on Signal Processing* **70** 6080–6095.

Geoffrion, Arthur M. 1968. Proper efficiency and the theory of vector maximization. *Journal of mathematical analysis and applications* **22**(3) 618–630.

Ghojogh, Benyamin, Mark Crowley. 2019. Unsupervised and supervised principal component analysis: Tutorial. *arXiv preprint arXiv:1906.03148* .

Gratton, Serge, Clément W Royer, Luís N Vicente, Zaikun Zhang. 2018. Complexity and global rates of trust-region methods based on probabilistic models. *IMA Journal of Numerical Analysis* **38**(3) 1579–1597.

Greenacre, Michael, Patrick JF Groenen, Trevor Hastie, Alfonso Iodice d'Enza, Angelos Markos, Elena Tuzhilina. 2022. Principal component analysis. *Nature Reviews Methods Primers* **2**(1) 100.

Grohs, Philipp, Seyedehsomayeh Hosseini. 2016. Nonsmooth trust region algorithms for locally lipschitz functions on riemannian manifolds. *IMA Journal of Numerical Analysis* **36**(3) 1167–1192.

Gutjahr, Walter J, Alois Pichler. 2016. Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research* **236** 475–499.

Haeser, Gabriel, Alberto Ramos. 2020. Constraint qualifications for karush–kuhn–tucker conditions in multiobjective optimization. *Journal of Optimization Theory and Applications* **187** 469–487.

Han, Chulwoo, Frank C Park. 2022. A geometric framework for covariance dynamics. *Journal of Banking & Finance* **134** 106319.

Hu, Jiang, Andre Milzarek, Zaiwen Wen, Yaxiang Yuan. 2018. Adaptive quadratically regularized newton method for riemannian optimization. *SIAM Journal on Matrix Analysis and Applications* **39**(3) 1181–1207.

Huang, De, Jonathan Niles-Weed, Rachel Ward. 2021. Streaming k-pca: Efficient guarantees for oja's algorithm, beyond rank-one updates. *Conference on Learning Theory*. PMLR, 2463–2498.

Huang, Long-Kai, Sinno Pan. 2020. Communication-efficient distributed pca by riemannian optimization. *International Conference on Machine Learning*. PMLR, 4465–4474.

Huang, Wen, Kyle A Gallivan. 2022. A limited-memory riemannian symmetric rank-one trust-region method with a restart strategy. *Journal of Scientific Computing* **93**(1) 1.

Huang, Wen, Kyle A Gallivan, P-A Absil. 2015. A broyden class of quasi-newton methods for riemannian optimization. *SIAM Journal on Optimization* **25**(3) 1660–1685.

Jia, Zhigang, Michael K Ng, Wei Wang. 2019. Color image restoration by saturation-value total variation. *SIAM Journal on Imaging Sciences* **12**(2) 972–1000.

Journée, Michel, Yurii Nesterov, Peter Richtárik, Rodolphe Sepulchre. 2010. Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research* **11**(2).

Kasai, Hiroyuki, Bamdev Mishra. 2018. Inexact trust-region algorithms on riemannian manifolds. *Advances in neural information processing systems* **31**.

Kuhn, Harold W, Albert W Tucker. 2013. Nonlinear programming. *Traces and emergence of nonlinear programming*. Springer, 247–258.

Lai, Rongjie, Stanley Osher. 2014. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing* **58** 431–449.

Larson, Jeffrey, Stephen C Billups. 2016. Stochastic derivative-free optimization using a trust region framework. *Computational Optimization and applications* **64** 619–645.

Lee, John M. 2006. *Riemannian manifolds: an introduction to curvature*, vol. 176. Springer Science & Business Media.

Lee, John M, John M Lee. 2012. *Smooth manifolds*. Springer.

Li, Jiaxiang, Krishnakumar Balasubramanian, Shiqian Ma. 2023. Stochastic zeroth-order riemannian derivative estimation and optimization. *Mathematics of Operations Research* **48**(2) 1183–1211.

Li, XF. 2000. Constraint qualifications in nonsmooth multiobjective optimization. *Journal of optimization theory and applications* **106** 373–398.

Liang, Yingyu, Maria-Florina F Balcan, Vandana Kanchanapally, David Woodruff. 2014. Improved distributed principal component analysis. *Advances in neural information processing systems* **27**.

Liu, Suyun, Luis Nunes Vicente. 2022. Accuracy and fairness trade-offs in machine learning: A stochastic multi-objective approach. *Computational Management Science* **19**(3) 513–537.

Maeda, T. 1994. Constraint qualifications in multiobjective optimization problems: differentiable case. *Journal of Optimization Theory and Applications* **80** 483–500.

Maunu, Tyler, Teng Zhang, Gilad Lerman. 2019. A well-tempered landscape for non-convex robust subspace recovery. *Journal of Machine Learning Research* **20**(37).

Mishra, SK, Vinay Singh, Vivek Laha, RN Mohapatra. 2015. On constraint qualifications for multiobjective optimization problems with vanishing constraints. *Optimization methods, theory and applications* 95–135.

Moakher, Maher. 2002. Means and averaging in the group of rotations. *SIAM journal on matrix analysis and applications* **24**(1) 1–16.

Najafi, Shahabeddin, Masoud Hajarian. 2023a. Multiobjective bfgs method for optimization on riemannian manifolds. *Computational Optimization and Applications* 1–18.

Najafi, Shahabeddin, Masoud Hajarian. 2023b. Multiobjective conjugate gradient methods on riemannian manifolds. *Journal of Optimization Theory and Applications* **197**(3) 1229–1248.

Pandey, Yogendra, Vinay Singh. 2021. On constraint qualifications for multiobjective optimization problems with switching constraints. *Optimization, Variational Analysis and Applications: IFSOVAA-2020, Varanasi, India, February 2–4*. Springer, 283–306.

Pantaleo, Ester, Michele Tumminello, Fabrizio Lillo, Rosario N Mantegna. 2011. When do improved covariance matrix estimators enhance portfolio optimization? an empirical comparative study of nine estimators. *Quantitative Finance* **11**(7) 1067–1080.

Pennec, Xavier. 2020. Manifold-valued image processing with spd matrices. *Riemannian geometric statistics in medical image analysis*. Elsevier, 75–134.

Poirion, Fabrice, Quentin Mercier, Jean-Antoine Désidéri. 2017. Descent algorithm for nonsmooth stochastic multiobjective optimization. *Computational Optimization and Applications* **68** 317–331.

Preda, V., I. Chiţescu. 1999. On constraint qualification in multiobjective optimization problems: semidifferentiable case. *Optimization methods, theory and applications* **100** 417–433.

Qu, Shaojian, Mark Goh, Bing Liang. 2013. Trust region methods for solving multiobjective optimisation. *Optimization Methods and Software* **28**(4) 796–811.

Rapcsak, Tamas. 1991. Geodesic convexity in nonlinear optimization. *Journal of Optimization Theory and Applications* **69**(1) 169–183.

Reynolds, Douglas A, et al. 2009. Gaussian mixture models. *Encyclopedia of biometrics* **741**(659-663).

Ritchie, Alexander, Laura Balzano, Daniel Kessler, Chandra S Sripada, Clayton Scott. 2022. Supervised pca: A multiobjective approach. *arXiv preprint arXiv:2011.05309* .

Ritchie, Alexander, Clayton Scott, Laura Balzano, Daniel Kessler, Chandra S Sripada. 2019. Supervised principal component analysis via manifold optimization. *2019 IEEE Data Science Workshop (DSW)*. IEEE, 6–10.

Sala, Frederic, Chris De Sa, Albert Gu, Christopher Ré. 2018. Representation tradeoffs for hyperbolic embeddings. *International conference on machine learning*. PMLR, 4460–4469.

Shashaani, Sara, Fatemeh S Hashemi, Raghu Pasupathy. 2018. Astro-df: A class of adaptive sampling trust-region algorithms for derivative-free stochastic optimization. *SIAM Journal on Optimization* **28**(4) 3145–3176.

Singh, C. 1987. Optimality conditions in multiobjective differentiable programming. *Journal of Optimization Theory and Applications* **53** 115–123.

Smith, Steven. 1995. Optimization techniques on riemannian manifolds. *Hamiltonian and Gradient Flows, Algorithms and Control* 113–136.

Stein, Oliver, Maximilian Volk. 2023. Generalized polarity and weakest constraint qualifications in multiobjective optimization. *Journal of Optimization Theory and Applications* **198**(3) 1156–1190.

Tripuraneni, Nilesh, Nicolas Flammarion, Francis Bach, Michael I Jordan. 2018. Averaging stochastic gradient descent on riemannian manifolds. *Conference on Learning Theory*. PMLR, 650–687.

Udriste, Constantin. 1993. *Convex functions and optimization methods on Riemannian manifolds*, vol. 297. Springer Science & Business Media.

Ullah, Enayat, Poorya Mianjy, Teodor Vanislavov Marinov, Raman Arora. 2018. Streaming kernel pca with $\tilde{O}(\sqrt{n})$ random features. *Advances in Neural Information Processing Systems* **31**.

Upadhyay, Balendu Bhooshan, Arnav Ghosh, Savin Treanţă. 2023. Constraint qualifications and optimality criteria for nonsmooth multiobjective programming problems on hadamard manifolds. *Journal of Optimization Theory and Applications* 1–26.

Vashishth, Vidushi, Anshuman Chhabra, Deepak Kumar Sharma. 2019. Gmmr: A gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic iot networks. *Computer Communications* **134** 138–148.

Vese, Luminita A, Stanley J Osher. 2002. Numerical methods for p-harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis* **40**(6) 2085–2104.

Wang, Jinhua, Xiangmei Wang, Chong Li, Jen-Chih Yao. 2021. Convergence analysis of gradient algorithms on riemannian manifolds without curvature constraints and application to riemannian mass. *SIAM Journal on Optimization* **31**(1) 172–199.

Wang, Xiaolu, Yuchen Jiao, Hoi-To Wai, Yuantao Gu. 2023. Incremental aggregated riemannian gradient method for distributed pca. *International Conference on Artificial Intelligence and Statistics*. PMLR, 7492–7510.

Wei, Hejie, Wei Hong Yang. 2016. A riemannian subspace limited-memory sr1 trust region method. *Optimization Letters* **10**(8) 1705–1723.

Wei, Hejie, Wei Hong Yang, Yinsheng Chai. 2023. A riemannian subspace bfgs trust region method. *Optimization Letters* 1–26.

Weinmann, Andreas, Laurent Demaret, Martin Storath. 2014. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences* **7**(4) 2226–2257.

Xu, Lei, Michael I Jordan. 1996. On convergence properties of the em algorithm for gaussian mixtures. *Neural computation* **8**(1) 129–151.

Xu, Peng, Fred Roosta, Michael W Mahoney. 2020. Newton-type methods for non-convex optimization under inexact hessian information. *Mathematical Programming* **184**(1-2) 35–70.

Yang, Wei Hong, Lei-Hong Zhang, Ruyi Song. 2014. Optimality conditions for the nonlinear programming problems on riemannian manifolds. *Pacific Journal of Optimization* **10**(2) 415–434.

Yger, Florian, Maxime Berar, Fabien Lotte. 2016. Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(10) 1753–1762.

Yi, Xinyang, Dohyung Park, Yudong Chen, Constantine Caramanis. 2016. Fast algorithms for robust pca via gradient descent. *Advances in neural information processing systems* **29**.

Yu, Philip LH, FC Ng, Jessica KW Ting. 2020. Adjusting covariance matrix for risk management. *Quantitative Finance* **20**(10) 1681–1699.

Zhang, Peng, Jin Zhang, Gui-Hua Lin, Xinmin Yang. 2018. Constraint qualifications and proper pareto optimality conditions for multiobjective problems with equilibrium constraints. *Journal of Optimization Theory and Applications* **176** 763–782.

Zhang, Teng, Yi Yang. 2018. Robust pca by manifold optimization. *The Journal of Machine Learning Research* **19**(1) 3101–3139.

Zhang, Zhuan, Shuisheng Zhou, Dong Li, Ting Yang. 2022. Riemannian proximal stochastic gradient descent for sparse 2dpca. *Digital Signal Processing* **122** 103320.

Zhao, Shimin, Tao Yan, Yuanguo Zhu. 2023a. Proximal gradient algorithm with trust region scheme on riemannian manifold. *Journal of Global Optimization* 1–26.

Zhao, Yong, Wang Chen, Xinmin Yang. 2023b. Adaptive sampling stochastic multigradient algorithm for stochastic multiobjective optimization. *Journal of Optimization Theory and Applications* 1–27.

Zhou, Shiji, Wenpeng Zhang, Jiyan Jiang, Wenliang Zhong, Jinjie Gu, Wenwu Zhu. 2022. On the convergence of stochastic multi-objective gradient manipulation and beyond. *Advances in Neural Information Processing Systems* **35** 38103–38115.