RIEMANNIAN OPTIMIZATION OVER PSD MATRIX CONSTRAINTS AND APPLICATIONS

by

Shixin Zheng

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Mathematics
West Lafayette, Indiana
August 2025

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Xiangxiong Zhang, Chair
Department of Mathematics

Dr. Rongjie Lai
Department of Mathematics

Dr. Guang Lin
Department of Mathematics

Dr. Greg Buzzard
Department of Mathematics

Approved by:
Dr. Erika Birgit Kaufmann

To my mother.

ACKNOWLEDGMENTS

I want to express my sincere thanks to my advisor, Professor Xiangxiong Zhang. His unwavering support and wise guidance have shaped not only my academic pursuits but also my personal development. More than just an advisor, his mentorship, patience, and life wisdom have had a profound influence on me as I made the journey to mentally mature. I will always be grateful for the moments when he shared his life insights through his own experiences and his observations of the world, and encouraged me to become a better thinker and helped me set up the goal for success that is beyond the Ph.D. career.

I am also deeply grateful to every professor that I have met in Purdue, through courses and through research. They have sharpened my knowledge in mathematics and help me become more confident in academic pursuits.

I am also deeply thankful to my girlfriend, Emma, for her unwavering companionship and unconditional support. Her empathy, hands-on help, and her true love have carried me through the most difficult moments of this journey. She has brought balance, comfort, and joy into my life in ways that words can hardly capture. We have come across so many challenges, through which we have both become more mature and love each other more. She has taught me how to love and helped me to grow up like a man.

To my mother, I offer my sincerest thanks and love. Her sacrifices and strong belief in me, both emotionally and financially, have also been the foundation on which I am able to build this pursuit of knowledge. Her resilience and support have taken me further than she knows, and I dedicate this accomplishment to her.

I also want to thank my other family members for always being my steady support from behind the scenes. Their encouragement, trust, and presence, even from afar, have helped me persevere through the challenges of this journey.

I am also grateful to all the friends I have met at Purdue. Whether through tennis, badminton, snowboarding, gym sessions, or simply sharing conversations, they have all brought color, joy, and much more into my life. These memories have made my PhD experience richer and more meaningful. Although some of them I have lost connection to, everyone that has appeared in my life has helped me shape into a better person.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Optimization and sampling over the manifold of fixed-rank positive semidefinite (PSD) matrices arise in a wide range of scientific and engineering applications, including signal processing, machine learning, quantum information, and statistical inference. This thesis develops a comprehensive geometric framework for addressing such problems using Riemannian optimization and Riemannian Langevin Monte Carlo techniques.

We begin by studying the manifold of Hermitian PSD matrices of fixed rank through both embedded and quotient manifold perspectives. A unified framework is proposed to encompass three commonly used geometriesembedded geometry, quotient geometry with the Bures–Wasserstein metric, and quotient geometry with alternative metrics that clarify their relationships and computational trade-offs.

Building on this framework, we design and analyze several Riemannian optimization algorithms, including the Riemannian conjugate gradient method. We prove their equivalence to classical Burer–Monteiro type algorithms and provide new insights into their convergence behavior, especially in the presence of rank-deficiency. A detailed condition number analysis reveals that certain Riemannian metrics lead to ill-conditioning near the boundary of the manifold, impacting algorithm performance.

Extending beyond optimization, we develop Riemannian Langevin Monte Carlo schemes for sampling from distributions defined over fixed-rank PSD manifolds. Two discretizationsbased on embedded and quotient geometriesare proposed, analyzed, and validated through numerical experiments.

Comprehensive numerical results on eigenvalue computation, matrix completion, phase retrieval, and interferometric recovery demonstrate the effectiveness of the proposed algorithms and validate the theoretical findings. This thesis thus offers both theoretical foundations and practical tools for manifold-based optimization and sampling over low-rank PSD matrix constraints.

1. INTRODUCTION

1.1 Optimization over PSD Constraints

Optimization over positive semidefinite (PSD) matrices is a central theme in numerous areas such as covariance estimation [1], kernel learning [2], semidefinite programming [3], etc. In many modern applications, large-scale semidefinite programs and matrix approximation problems demand not only low-rank structure but also efficient numerical schemes that exploit manifold geometry. See [4] and [5] for some of these applications. In mathematical notations, the optimization problem over PSD matrices can be written formally as

$$\begin{array}{ll} \underset{X \in \mathbb{C}^{n \times n}}{\text{minimize}} & f(X) \\ \text{subject to} & X \succcurlyeq 0 \end{array} . \tag{1.1}$$

A particularly compelling structure arises when one restricts attention to the manifold of PSD matrices with fixed rank, which possesses both rich geometric properties and significant practical relevance. For example, real symmetric PSD fixed-rank matrices were used in [6, 7]. When the solution to (1.1) exhibits low-rank structure, we can significantly reduce computational complexity by working directly with the manifold of fixed-rank matrices:

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) \\ \text{subject to} & X \in \mathcal{H}_{+}^{n,p} \end{array} , \tag{1.2}$$

where $\mathcal{H}_{+}^{n,p}$ denotes the set of n-by-n Hermitian PSD matrices of fixed rank $p \ll n$. Optimization over PSD matrices of fixed rank can also be used for solving non-symmetric problems. Suppose the set of non-symmetric matrices of fixed-rank is $\{X = LR^T : \text{rank}(X) = p\}$. Then define the symmetric lifting: $\{\tilde{X} := [L, R][LR]^T\}$, which again becomes a PSD fixed rank constraint. However, the nature of nonconvex optimization problems also makes (1.2) challenging to solve.

Since the elements in the constraint set $\mathcal{H}_{+}^{n,p}$ have a low-rank structure, they can be represented in a low-rank compact form on the order of $\mathcal{O}(np^2)$, which is smaller than the $\mathcal{O}(n^2)$ storage when directly using $X \in \mathbb{C}^{n \times n}$. In many applications, the cost function in

(1.1) takes the form $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_F^2$ where $\mathcal{A}$ is a linear operator and the norm is the Frobenius norm, and $f(X)$ can be evaluated efficiently by $\mathcal{O}(pn \log n)$ flops for $X \in \mathcal{H}_+^{n,p}$, e.g., the PhaseLift problem [8, 9] and the interferometry recovery problem [10, 11]. For these kinds of problems, solving (1.2) with an iterative algorithm that works with low-rank representations for $X \in \mathcal{H}_+^{n,p}$ can lead to a good approximate solution to (1.1) with compact storage and computational cost.

1.2 Sampling over PSD Constraints

Beyond optimization, we can also explore the stochastic counterpart of such problems–namely, Riemannian sampling over fixed-rank PSD manifolds, thus extending the scope of manifold-based computation beyond optimization.

There is an extensive literature on Langevin dynamics in statistics and related areas, with interest in nonconvex optimization [12, 13], as well as machine learning such as generative models [14].

In recent years, there has been interest in studying Langevin diffusion and Monte Carlo Markov Chain (MCMC) schemes on manifolds [15–24]. In this thesis, we focus on the Riemannian Langevin Monte Carlo schemes on $\mathcal{S}_+^{n,p}$ that samples from the Gibbs distribution on the manifold of fixed rank PSD matrices:

$$\text{sample from}\, \rho(X) \propto \mathrm{e}^{-\beta f(X)}, \quad \text{subject to}\, X \in \mathcal{S}_+^{n,p}, \tag{1.3}$$

based on the Riemannian Langevin equation (RLE) on the manifold that generalizes the Langevin dynamics in the Euclidean space.

Gibbs distributions originate in statistical physics, while the sampling problem may also be seen as a stochastic variant of the optimization problem in the sense that the sampling problem is related to the optimization problem since in the limit $\beta \to \infty$ the Gibbs distribution concentrates at the global minima of $f(X)$. In general, a Langevin scheme can be used for either optimization [24, 25], or Monte Carlo type numerical integration, which is common in Bayesian statistics. For optimization, stochastic optimization by Langevin dynamics with simulated annealing is an established approach [26]. For sampling, Metropolis-adjusted Langevin algorithm [17] is often used.

1.3 Contributions of This Thesis

This thesis contributes to the field of manifold optimization and sampling in several key ways:

1. Unified Riemannian Optimization Framework: We present and analyze three methodologies for optimization on the manifold of Hermitian PSD matrices of fixed rank, including Burer–Monteiro factorization, embedded geometry, and quotient geometry using the Bures–Wasserstein metric. We show theoretical equivalence between these formulations under appropriate settings and validate them numerically.

2. Condition Number Analysis: We investigate the impact of rank-deficiency on the conditioning of Riemannian Hessians, deriving bounds and demonstrating how they inform the performance of Riemannian optimization algorithms.

3. Convergence Analysis of Orthogonalization-Free Algorithms: We analyze the convergence of orthogonalization-free Riemannian conjugate gradient methods. This contributes to understanding the global behavior of such methods in the Burer–Monteiro setting.

4. Riemannian Langevin Monte Carlo on Fixed-Rank PSD Manifolds: We introduce two numerical schemes for sampling from Gibbs distributions defined on the manifold of fixed-rank PSD matrices using Riemannian Langevin dynamics. These are built upon the same geometric structures used in optimization, allowing a principled transition between deterministic and stochastic methods.

5. Comprehensive Numerical Validation: Each of the proposed frameworks is validated on a range of problems, including eigenvalue computation, matrix completion, PhaseLift, and interferometric inversion. Our experiments demonstrate the practical viability of the theoretical framework developed throughout the thesis.

1.4 Organization of the Thesis

The thesis is organized as follows:

Chapter 2 provides the mathematical preliminaries and geometric foundations of fixed-rank PSD manifolds, including both embedded and quotient geometries.

Chapter 3 presents a unified view of optimization over fixed-rank Hermitian PSD matrices, contrasting three algorithmic approaches and exploring their theoretical connections.

Chapter 4 conducts a condition number analysis of Riemannian Hessians in the presence of rank-deficiency and discusses its implications through illustrative applications.

Chapter 5 studies the convergence properties of orthogonalization-free Riemannian conjugate gradient methods, providing both theoretical guarantees and numerical evidence.

Chapter 6 transitions to Riemannian sampling and introduces Langevin Monte Carlo schemes designed for the fixed-rank PSD manifold, complete with empirical validation against known distributions.

Appendices contain detailed derivations, supplemental mathematical results, and implementation notes that support the main text.

2. PRELIMINARIES AND THE MANIFOLD OF FIXED-RANK PSD MATRICES

In this chapter, we first review some preliminaries and introduce the geometric structure of the manifold of the fixed-rank PSD matrices. We only consider the case of Hermitian PSD matrices since the results of the real symmetric PSD matrices will simply follow.

2.1 Embedded Manifold Geometry of Fixed-rank PSD Matrices

We first show that $\mathcal{H}_+^{n,p}$ is a smooth embedded submanifold of $\mathbb{C}^{n\times n}$.

Theorem 2.1.1. Regard $\mathbb{C}^{n\times n}$ as a real vector space over $\mathbb{R}$ of dimension $2n^2$. Then $\mathcal{H}_+^{n,p}$ is a smooth embedded submanifold of $\mathbb{C}^{n\times n}$ of dimension $2np - p^2$.

Proof. Let

$$E = \begin{bmatrix} I_{p\times p} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & 0_{(n-p)\times(n-p)} \end{bmatrix}$$

and consider the smooth Lie group action

$$\begin{aligned} \Phi : \mathrm{GL}(n,\mathbb{C}) \times \mathbb{C}^{n\times n} &\to \mathbb{C}^{n\times n} \\ (g, N) &\mapsto gNg^* \end{aligned}$$

where

$$\begin{aligned} gNg^* \;=\; & (\mathrm{Re}(g)\,\mathrm{Re}(N) - \mathrm{Im}(g)\,\mathrm{Im}(N))\,\mathrm{Re}(g)^T + (\mathrm{Im}(g)\,\mathrm{Re}(N) + \mathrm{Re}(g)\,\mathrm{Im}(N))\,\mathrm{Im}(g)^T \\ & + \mathbb{i}\left((\mathrm{Im}(g)\,\mathrm{Re}(N) + \mathrm{Re}(g)\,\mathrm{Im}(N))\,\mathrm{Re}(g)^T - (\mathrm{Re}(g)\,\mathrm{Re}(N) - \mathrm{Im}(g)\,\mathrm{Im}(N))\,\mathrm{Im}(g)^T\right). \end{aligned}$$

From the above expression of gNg^*, we see that Φ is a rational mapping. Since $\mathrm{GL}(n,\mathbb{C})$ is a semialgebraic set by Lemma (B.0.1) in the Appendix, we have that $\mathrm{GL}(n,\mathbb{C})\times\mathbb{C}^{n\times n}$ is also a semialgebraic set [27, section 2.1.1]. It follows from (B1) in [28] that Φ is a semialgebraic mapping. Observe that $\mathcal{H}_+^{n,p}$ is the orbit of E through Φ. It therefore follows from (B4) in [28] that $\mathcal{H}_+^{n,p}$ is a smooth submanifold of $\mathbb{C}^{n\times n}$.

Next, we compute the dimension of $\mathcal{H}_+^{n,p}$. Consider the smooth surjective mapping

$$\eta : \mathrm{GL}(n, \mathbb{C}) \to \mathcal{H}_+^{n,p} \quad \gamma \mapsto \gamma E \gamma^*.$$

The differential of η at $\gamma \in \mathrm{GL}(n, \mathbb{C})$ is the linear mapping $\mathrm{D}\eta(\gamma) : T_\gamma \mathrm{GL}(n, \mathbb{C}) = \mathbb{C}^{n\times n} \to T_X \mathcal{H}_+^{n,p}$, where $X = \eta(\gamma) = \gamma E \gamma^*$, by $\mathrm{D}\eta(\gamma)[\Delta] = \Delta E \gamma^* + \gamma E \Delta^*$. Observe that the differential at arbitrary γ is related to the differential at I_n by a full-rank linear transformation:

$$\mathrm{D}\eta(\gamma)[\Delta] = \gamma \mathrm{D}\eta(I_n)[\gamma^{-1}\Delta]\gamma^*. \tag{2.1}$$

Recall that the rank of a differentiable mapping f between two differentiable manifolds is the dimension of the image of the differential of f. So, from equation (2.1) we see that the rank of η is constant. It follows from Theorem 4.14 in [29] that η is a smooth submersion. As a consequence $\mathrm{D}\eta(\gamma)$ maps $T_\gamma \mathrm{GL}(n, \mathbb{C}) = \mathbb{C}^{n\times n}$ surjectively onto $T_X \mathcal{H}_+^{n,p}$ and we obtain

$$T_X \mathcal{H}_+^{n,p} = \left\{ \Delta X + X \Delta^* : \Delta \in \mathbb{C}^{n\times n} \right\}. \tag{2.2}$$

Let $\Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}$ be partitioned according to the partition of $E = \mathrm{diag}(I_{p\times p}) = \begin{bmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{bmatrix}$. Then it can be easily verified that $\Delta \in \mathrm{Ker}\mathrm{D}\eta(I)$ if and only if

$$\Delta_{11} = -\Delta_{11}^*, \quad \Delta_{21} = 0.$$

This implies that Δ_{11} is a skew-Hermitian matrix, hence its diagonal entries are purely imaginary and its off diagonal entries satisfy $a_{\mathrm{ij}} = -\overline{a_{\mathrm{ji}}}$. This gives us $p + 2 \times (1 + 2 + \cdots + (p-1))$ degrees of freedom. For Δ_{12} and Δ_{22} there are $2n(n-p)$ degrees of freedom. So, the dimension of $\mathrm{Ker}(\mathrm{D}\eta(I))$ is $2n(n-p) + p + 2p(p-1)/2 = 2n^2 - 2np + p^2$ and by rank-nullity we get

$$\dim \mathrm{D}\, \eta(I) = 2n^2 - \dim \ker \mathrm{D}\, \eta(I) = 2np - p^2.$$

Since η is of constant rank, the dimension of $T_X\mathcal{H}_+^{n,p}$ is therefore $2np - p^2$. Remember that the dimension of the tangent space at every point of a connected manifold is the same as that of the manifold itself. Let $\mathrm{GL}^+(n,\mathbb{C})$ denote the connected subset of $\mathrm{GL}(n,\mathbb{C})$ with positive determinant, then $\mathcal{H}_+^{n,p}$ is the image of the connected set $\mathrm{GL}^+(n,\mathbb{C})$ under a continuous mapping η, so $\mathcal{H}_+^{n,p}$ is connected. We conclude that the dimension of $\mathcal{H}_+^{n,p}$ is $2np - p^2$.

The next result characterizes the tangent space.

Theorem 2.1.2. Let $X = U\Sigma U^* \in \mathcal{H}_+^{n,p}$. Then the tangent space of $\mathcal{H}_+^{n,p}$ at X is given by

$$T_X\mathcal{H}_+^{n,p} = \left\{ \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} H & K^* \\ K & 0 \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix} \right\}$$

where $H = H^* \in \mathbb{C}^{p\times p}$, $K \in \mathbb{C}^{(n-p)\times p}$.

Proof. Let $t \mapsto U(t)$ be any smooth curve in $\mathrm{St}\,(p,n)$ through U at $t=0$ such that $U(t) \in \mathbb{C}^{n\times p}, U(0) = U$ and $U(t)^*U(t) = I_p$ for all t. Let $t \mapsto \Sigma(t)$ be any smooth curve in $\mathrm{Diag}(p,p)$ through Σ at $t = 0$. Then $X(t) := U(t)\Sigma(t)U(t)^*$ defines a smooth curve in $\mathcal{H}_+^{n,p}$ through X. It follows by differentiating $X(t) := U(t)\Sigma(t)U(t)^*$ that

$$X'(t) = U'(t)\Sigma(t)U(t)^* + U(t)\Sigma'(t)U(t)^* + U(t)\Sigma(t)U'(t)^*.$$

Without loss of generality, since $U'(t)$ is an element of $\mathbb{C}^{n\times p}$ and $U(t)$ has full rank, we can set

$$U'(t) = U(t)A(t) + U_\perp(t)B(t).$$

Hence, we have

$$X'(t) = \begin{bmatrix} U(t) & U_\perp(t) \end{bmatrix} \begin{bmatrix} A(t)\Sigma(t) + \Sigma'(t) + \Sigma(t)A(t)^* & \Sigma(t)B(t)^* \\ B(t)\Sigma(t) & 0 \end{bmatrix} \begin{bmatrix} U(t)^* \\ U_\perp(t)^* \end{bmatrix}.$$

Thus we consider the tangent vectors in the form of $\begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} H & K^* \\ K & 0 \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix}$ with $H = H^*$. For any $H = H^* \in \mathbb{C}^{p\times p}$ and $K \in \mathbb{C}^{(n-p)\times p}$, taking $\Delta = (UH/2 + U_\perp K)\Sigma^{-1}(U^*U)^{-1}U^*$ in (2.2), we see that

$$\left\{ \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} H & K^* \\ K & 0 \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix} \right\} \subseteq T_X\mathcal{H}_+^{n,p}. \tag{2.3}$$

Now counting the real dimension we see that H has $p + 2 \times \frac{p(p-1)}{2} = p^2$ number of freedom and K has $2 \times p(n-p)$ number of freedom. So the LHS of the inclusion (2.3) has freedom $2np - p^2$, which is equal to the dimension of $T_X\mathcal{H}_+^{n,p}$. Hence, the inclusion in (2.3) is an equality.

The *Riemannian metric* of the embedded manifold at $X \in \mathcal{H}_+^{n,p}$ is induced from the Euclidean inner product on $\mathbb{C}^{n\times n}$,

$$g_X(\zeta_1, \zeta_2) = \langle \zeta_1, \zeta_2 \rangle_{\mathbb{C}^{n\times n}} = \mathrm{Re}(\mathrm{tr}(\zeta_1^*\zeta_2)), \quad \zeta_1, \zeta_2 \in T_X\mathcal{H}_+^{n,p}. \tag{2.4}$$

With the Riemannian metric, the angle is defined on a manifold and we can then define the normal space, which is the orthogonal space to the tangent space.

Lemma 2.1.3. The normal space $N_X\mathcal{H}_+^{n,p}$ at $X = U\Sigma U^* \in \mathcal{H}_+^{n,p}$ is given by

$$N_X\mathcal{H}_+^{n,p} = \left\{ \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Omega & -L^* \\ L & M \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix} \right\}, \tag{2.5}$$

where $\Omega = -\Omega^* \in \mathbb{C}^{p\times p}$, $M \in \mathbb{C}^{(n-p)\times(n-p)}$ and $L \in \mathbb{C}^{(n-p)\times p}$.

Proof. First we show that every vector in (2.5) is orthogonal to $T_X\mathcal{H}_+^{n,p}$. Since U is orthonormal, we only need to show that $\left\langle \begin{bmatrix} H & K^* \\ K & 0 \end{bmatrix}, \begin{bmatrix} \Omega & -L^* \\ L & M \end{bmatrix} \right\rangle_{\mathbb{C}^{n\times n}} = 0$ for all H, K, Ω, L and M defined in Theorem 2.1.2 and Lemma 2.1.3. Indeed we have

$$\left\langle \begin{bmatrix} H & K^* \\ K & 0 \end{bmatrix}, \begin{bmatrix} \Omega & -L^* \\ L & M \end{bmatrix} \right\rangle_{\mathbb{C}^{n\times n}} = \langle \Omega, H \rangle_{\mathbb{C}^{n\times n}} - \langle L^*, K^* \rangle_{\mathbb{C}^{n\times n}} + \langle L, K \rangle_{\mathbb{C}^{n\times n}}$$

$$= \quad \langle \Omega, H \rangle_{\mathbb{C}^{n\times n}} = 0.$$

Next, we count the real dimension of $N_X\mathcal{H}_+^{n,p}$. Remember that a skew-Hermitian matrix has purely imaginary numbers on its diagonal entries, and $\omega_{\mathrm{ij}} = -\overline{\omega}_{\mathrm{ji}}$ on its off diagonal entries. So the number of degree of freedoms in Ω is $p + 2 \times \frac{p(p-1)}{2} = p^2$. The number of degree of freedoms in L is $2 \times p(n-p)$, and the number of degree of freedoms in M is $2 \times (n-p)^2$. So, the dimension of $N_X\mathcal{H}_+^{n,p}$ is $2n^2 + p^2 - 2np$. This gives us the desired dimension since the sum of the dimension of the tangent space and its normal space should be $2n^2$.

The orthogonal projection from $\mathbb{C}^{n\times n}$ onto $T_X\mathcal{H}_+^{n,p}$ can also be calculated based on the Riemannian metric, which is given in the following theorem.

Theorem 2.1.4. Let $X = YY^* = U\Sigma U^*$ be the compact SVD for $X \in \mathcal{H}_+^{n,p}$ with $Y \in \mathbb{C}_*^{n\times p}$. Let $Z \in \mathbb{C}^{n\times n}$. Then the operator P_X^t defined below is the orthogonal projection onto $T_X\mathcal{H}_+^{n,p}$:

$$\begin{aligned} P_X^t(Z) &= \frac{1}{2}\left(P_Y(Z+Z^*)P_Y + P_Y^{\perp}(Z+Z^*)P_Y + P_Y(Z+Z^*)P_Y^{\perp}\right) \\ &= \frac{1}{2}\left(P_U(Z+Z^*)P_U + P_U^{\perp}(Z+Z^*)P_U + P_U(Z+Z^*)P_U^{\perp}\right) \qquad (2.6) \\ &= \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} U^*\frac{(Z+Z^*)}{2}U & U^*\frac{(Z+Z^*)}{2}U_\perp \\ U_\perp^*\frac{(Z+Z^*)}{2}U & 0 \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix}, \end{aligned}$$

where $P_Y = Y(Y^*Y)^{-1}Y^*$, $P_Y^{\perp} = I - P_Y = P_{Y_\perp}$, $P_U = UU^*$ and $P_U^{\perp} = I - P_U = P_{U_\perp}$.

Proof. First, observe that

$$\begin{aligned} P_X^t(Z) &= \begin{bmatrix} P_Y & P_{Y_\perp} \end{bmatrix} \begin{bmatrix} \frac{Z+Z^*}{2} & \frac{Z+Z^*}{2} \\ \frac{Z+Z^*}{2} & 0 \end{bmatrix} \begin{bmatrix} P_Y \\ P_{Y_\perp} \end{bmatrix} \\ &= \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} U^*\frac{(Z+Z^*)}{2}U & U^*\frac{(Z+Z^*)}{2}U_\perp \\ U_\perp^*\frac{(Z+Z^*)}{2}U & 0 \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix} \end{aligned}$$

is a tangent vector at X. So it suffices to show that $Z - P_X^t(Z)$ is a normal vector. Write Z as $Z = P_Y Z P_Y + P_Y Z P_{Y_\perp} + P_{Y_\perp} Z P_Y + P_{Y_\perp} Z P_{Y_\perp} = \begin{bmatrix} P_Y & P_{Y_\perp} \end{bmatrix} \begin{bmatrix} Z & Z \\ Z & Z \end{bmatrix} \begin{bmatrix} P_Y \\ P_{Y_\perp} \end{bmatrix}$. Then we have

$$\begin{aligned} Z - P_X^t(Z) &= \begin{bmatrix} P_Y & P_{Y_\perp} \end{bmatrix} \begin{bmatrix} \frac{Z-Z^*}{2} & \frac{Z-Z^*}{2} \\ \frac{Z-Z^*}{2} & Z \end{bmatrix} \begin{bmatrix} P_Y \\ P_{Y_\perp} \end{bmatrix} \\ &= \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} U^* \frac{(Z-Z^*)}{2} U & U^* \frac{(Z-Z^*)}{2} U_\perp \\ U_\perp^* \frac{(Z-Z^*)}{2} U & U_\perp^* Z U_\perp \end{bmatrix} \begin{bmatrix} U^* \\ U_\perp^* \end{bmatrix} \end{aligned}$$

Hence, $Z - P_X^t(Z)$ is a normal vector, which completes the proof.

Remark 2.1.5. We can write $P_X^t = P_X^s + P_X^p$ by introducing the two operators

$$P_X^s : Z \mapsto P_U \frac{Z + Z^*}{2} P_U \tag{2.7a}$$

$$P_X^p : Z \mapsto P_{U_\perp} \frac{Z + Z^*}{2} P_U + P_U \frac{Z + Z^*}{2} P_{U_\perp} \tag{2.7b}$$

2.2 Quotient Manifold Geometry of Fixed-rank PSD Matrices

Besides being regarded as an embedded manifold in $\mathbb{C}^{n\times n}$, $\mathcal{H}_+^{n,p}$ can also be viewed as a quotient set $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$, where $\mathcal{O}_p = \{O \in \mathbb{C}^{p\times p} : U^*U = I\}$ denotes the unitary group, since any $X \in \mathcal{H}_+^{n,p}$ can be written as $X = YY^*$ with $Y \in \mathbb{C}_*^{n\times p}$. But there is an ambiguity in Y because such an Y is not uniquely determined by X. We define an equivalence relation on $\mathbb{C}_*^{n\times p}$ through the smooth Lie group action of $\mathcal{O}_p$ on the manifold $\mathbb{C}_*^{n\times p}$:

$$\begin{aligned} \mathbb{C}_*^{n\times p} \times \mathcal{O}_p \to &\quad \mathbb{C}_*^{n\times p} \\ (Y, O) \mapsto &\quad YO. \end{aligned}$$

This action defines an equivalence relation on $\mathbb{C}_*^{n\times p}$ by setting $Y_1 \sim Y_2$ if there exists an $O \in \mathcal{O}_p$ such that $Y_1 = Y_2 O$. Hence we have constructed a quotient space $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ that removes this ambiguity. The set $\mathbb{C}_*^{n\times p}$ is called the *total space* of $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$.

Denote the natural projection as

$$\pi : \mathbb{C}_*^{n\times p} \to \mathbb{C}_*^{n\times p}/\mathcal{O}_p.$$

For any $Y \in \mathbb{C}_*^{n\times p}$, $\pi(Y)$ is an element in $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. We denote the equivalence class containing Y as

$$[Y] = \pi^{-1}(\pi(Y)) = \{YO | O \in \mathcal{O}_p\}.$$

Define

$$\begin{aligned} \beta : \mathbb{C}_*^{n\times p} &\to \mathcal{H}_+^{n,p} \\ Y &\mapsto YY^*. \end{aligned}$$

Then β is invariant under the equivalence relation $\sim$ and induces a unique function $\tilde{\beta}$ on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$, called the projection of β, such that $\beta = \tilde{\beta} \circ \pi$ [30, section 3.4.2]. One can easily check that $\tilde{\beta}$ is a bijection. This is summarized in the diagram below:

$$\begin{tikzcd}
\mathbb{C}_*^{n\times p} \arrow[d, "\pi"] \arrow[dr, dashed, "\beta := \tilde{\beta}\circ\pi"] & \\
\mathbb{C}_*^{n\times p}/\mathcal{O}_p \arrow[r, leftrightarrow, "\tilde{\beta}"] & \mathcal{H}_+^{n,p}
\end{tikzcd}$$

The next theorem shows that $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is a smooth manifold.

Theorem 2.2.1. The quotient space $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is a quotient manifold over $\mathbb{R}$ of dimension $2np - p^2$ and has a unique smooth structure such that the natural projection π is a smooth submersion.

Proof. The proof follows from Corollary 21.6 and Theorem 21.10 of [29].

The next theorem shows that $\mathcal{H}_+^{n,p}$ and $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ are essentially the same in the sense that there is a diffeomorphism between them. The proof uses the same technique in [4, Prop. A.7]

Theorem 2.2.2. The quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is diffeomorphic to $\mathcal{H}_+^{n,p}$ under $\tilde{\beta}$.

Proof. Recall from Theorem 2.1.2, any tangent vector in $T_{\beta(Y)}\mathcal{H}_+^{n,p}$ can be written as

$$\zeta_{\beta(Y)} = YHY^* + Y_\perp KY^* + YK^*Y_\perp^*,$$

where $Y_\perp$ has orthonormal columns. Let $V = YH/2 + Y_\perp K$, then $D\beta(Y)[V] = \zeta_{\beta(Y)}$. This implies that β is a submersion.

Now notice that $\pi = \tilde{\beta}^{-1} \circ \beta$ and $\beta = \tilde{\beta} \circ \pi$. By [31, Prop. 6.1.2], we conclude that $\tilde{\beta}^{-1}$ and $\tilde{\beta}$ are both differentiable. So $\tilde{\beta}$ is a diffeomorphism between $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{H}_+^{n,p}$.

The equivalence class $[Y]$ is an embedded submanifold of $\mathbb{C}_*^{n\times p}$([30, Prop. 3.4.4]). The tangent space of $[Y]$ at Y is therefore a subspace of $T_Y\mathbb{C}_*^{n\times p}$ called the *vertical space* at Y and is denoted by $\mathcal{V}_Y$. The following proposition characterizes $\mathcal{V}_Y$.

Proposition 2.2.3. The vertical space at $Y \in [Y] = \{YO|O \in \mathcal{O}_p\}$, which is the tangent space of $[Y]$ at Y is

$$\mathcal{V}_Y = \left\{Y\Omega|\Omega^* = -\Omega, \Omega \in \mathbb{C}^{p\times p}\right\}.$$

Proof. The tangent space of $\mathcal{O}_p$ at I_p is $T_{I_p}\mathcal{O}_p = \{\Omega : \Omega^* = -\Omega, \Omega \in \mathbb{C}^{p\times p}\}$, which is also the set $\{\gamma'(0) : \gamma \text{ is a curve in } \mathcal{O}_p, \gamma(0) = I_p\}$. Hence $T_Y\{YO|O \in \mathcal{O}_p\} = \{Y\gamma'(0) : \gamma \text{ is a curve in } \mathcal{O}_p, \gamma(0) = I_p\} = \{Y\Omega|\Omega^* = -\Omega, \Omega \in \mathbb{C}^{p\times p}\}$.

2.2.1 Choices of Riemannian Metric and Horizontal Space

A *Riemannian metric* can be defined on the total space $\mathbb{C}_*^{n\times p}$. That is, $g_Y(\cdot,\cdot)$ is an inner product on $T_Y\mathbb{C}_*^{n\times p}$. Once we choose a Riemannian metric g for $\mathbb{C}_*^{n\times p}$, we can obtain the orthogonal complement in $T_Y\mathbb{C}_*^{n\times p}$ of $\mathcal{V}_Y$ with respect to the metric since $\mathcal{V}_Y$ is an embedded submanifold of the total space. In other words, we choose the *horizontal distribution* as orthogonal complement w.r.t. Riemannian metric, see [30, Section 3.5.8]. This orthogonal complement to $\mathcal{V}_Y$ is called *horizontal space* at Y and is denoted by $\mathcal{H}_Y$. We thus have

$$T_Y\mathbb{C}_*^{n\times p} = \mathcal{H}_Y \oplus \mathcal{V}_Y. \tag{2.8}$$

Once we have the horizontal space, there exists a unique vector $\bar{\xi}_Y \in \mathcal{H}_Y$ that satisfies $\mathrm{D}\,\pi(Y)[\bar{\xi}_Y] = \xi_{\pi(Y)}$ for each $\xi_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. This $\bar{\xi}_Y$ is called the *horizontal lift* of $\xi_{\pi(Y)}$ at Y.

There exist more than one choice of Riemannian metric on $\mathbb{C}_*^{n\times p}$. Different Riemannian metrics do not affect the vertical space, but generally result in different horizontal spaces.

Now, we will introduce several metric choices for the total space $\mathbb{C}_*^{n\times p}$. Then, we will show that these metrics also induce Riemannian metrics for the quotient manifold, such that the quotient manifold becomes a Riemannian manifold.

The Bures-Wasserstein Metric

The most straightforward choice of a Riemannian metric on $\mathbb{C}_*^{n\times p}$ is the canonical Euclidean inner product on $\mathbb{C}^{n\times p}$ defined by

$$g_Y^1(A, B) := \langle A, B\rangle_{\mathbb{C}^{n\times p}} = \mathrm{Re}(\mathrm{tr}(A^*B)), \quad \forall A, B \in T_Y\mathbb{C}_*^{n\times p} = \mathbb{C}^{n\times p}.$$

The metric g^1 is also called the Bures-Wasserstein metric [32] for the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. On the other hand, the following metric for Hermitian positive-definite matrices $\mathcal{H}_+^{n,n}$ [33–35] is also called the Bures-Wasserstein metric.

Definition 2.2.1 (The Bures-Wasserstein metric for $\mathcal{H}_+^{n,n}$). Let $X \in \mathcal{H}_+^{n,n}$ and $A, B \in T_X\mathcal{H}_+^{n,n}$. Then

$$g_X^{\mathrm{BW}}(A, B) := \frac{1}{2}\left\langle \mathcal{L}_X(A), B\right\rangle,$$

where $\mathcal{L}_X(A) = M$ solves the following Lyapunov equation

$$XM + MX = A \tag{2.9}$$

which has a unique solution provided X is Hermitian positive-definite.

Notice that it is not clear whether Definition 2.2.1 can also apply to a low-rank matrix $X \in \mathcal{H}_+^{n,p}$. In this subsection, we show how the metric g^1 can be used to generalize Definition 2.2.1 to Definition 2.2.2, which defines the Bures-Wasserstein metric in the low-rank case $\mathcal{H}_+^{n,p}$. This non-trivial generalization is presented as Theorem 2.2.6. It is of interest to see how g^1 connects the Bures-Wasserstein metric on the quotient manifold to its counterpart on the embedded manifold.

Definition 2.2.2 (The Bures-Wasserstein metric on $\mathcal{H}_+^{n,p}$). Let $A, B \in T_{YY^*}\mathcal{H}_+^{n,p}$, then by the 1-to-1 correspondence between $T_{YY^*}\mathcal{H}_+^{n,p}$ and the horizontal space $\mathcal{H}_Y^1$, there exist unique $\xi_Y, \eta_Y \in \mathcal{H}_Y^1$ such that $A = Y\xi_Y^* + \xi_Y Y^*$ and $B = Y\eta_Y^* + \eta_Y Y^*$. We define the Bures-Wasserstein metric at the low-rank $X = YY^*$ as

$$g_{YY^*}^{\text{BW}}(A, B) := g_Y^1(\xi_Y, \eta_Y).$$

Lemma 2.2.4. For any $A, B \in T_X\mathcal{H}_+^{n,p}$ with $X = YY^*$, there is a unique solution $M \in T_X\mathcal{H}_+^{n,p}$ satisfying both

$$Y^*XMY + Y^*MXY = Y^*AY \tag{2.10}$$

and

$$g_{YY^*}^{\text{BW}}(A, B) = \frac{1}{2}\langle M, B\rangle_{\mathbb{C}^{n\times n}}. \tag{2.11}$$

Proof. Let $\xi_Y = Y(Y^*Y)^{-1}S + Y_\perp K \in \mathcal{H}_Y^1$ with $S^* = S$ be the unique horizontal vector such that $A = Y\xi_Y^* + \xi_Y Y^*$. Let $Y = UR$ where U has size n-by-p with orthonormal columns and R is an p-by-p invertible matrix. Thus (2.10) is equivalent to

$$RR^*(U^*MU) + (U^*MU)RR^* = RSR^{-1} + (R^*)^{-1}SR^*. \tag{2.12}$$

Since RR^* is positive definite, (2.12) has a unique solution in U^*MU; see Remark 2.2.5 below, which can be written explicitly:

$$U^*MU = (R^*)^{-1}SR^{-1}. \tag{2.13}$$

Thus $M = \begin{bmatrix} U & Y_\perp \end{bmatrix} \begin{bmatrix} (R^*)^{-1}SR^{-1} & K_M^* \\ K_M & 0 \end{bmatrix} \begin{bmatrix} U^* \\ Y_\perp^* \end{bmatrix}$, where K_M is to be determined by the additional equation (2.11). With $B = Y\eta_Y^* + \eta_Y Y^*$ we have,

$$\frac{1}{2} \langle M, B \rangle_{\mathbb{C}^{n\times n}} = \frac{1}{2} \langle M, Y\eta_Y^* \rangle_{\mathbb{C}^{n\times n}} + \frac{1}{2} \langle M, \eta_Y Y^* \rangle_{\mathbb{C}^{n\times n}} = \langle MY, \eta_Y \rangle_{\mathbb{C}^{n\times p}} .$$

Thus in order for (2.11) to hold, M needs to satisfy $MY = \xi_Y$. Recall that $\xi_Y = Y(Y^*Y)^{-1}S + Y_\perp K = U(R^*)^{-1}S + Y_\perp K$. Thus K_M needs to satisfy $Y_\perp K_M R = Y_\perp K$, which gives the unique $K_M = KR^{-1}$.

Remark 2.2.5. The solution X to the Lyapunov equation $XE + EX = Z$ for a Hermitian E is unique if E is Hermitian positive-definite [4, Section 2.2]. Let $E = U\Lambda U^*$ be the SVD, then the Lyapunov equation $XE + EX = Z$ becomes

$$(U^*XU)\Lambda + \Lambda(U^*XU) = U^*ZU,$$

which gives the solution

$$(U^*XU)_{\mathrm{i,j}} = (U^*ZU)_{\mathrm{i,j}} / (\Lambda_{\mathrm{i,i}} + \Lambda_{\mathrm{j,j}}).$$

Now we can show that Definition 2.2.1 generalizes the Definition 2.2.2 and defines the Bures-Wasserstein metric in the low-rank case $\mathcal{H}_+^{n,p}$ in the following theorem.

Theorem 2.2.6 (Equivalence of the two Bures-Wasserstein metrics). If $p = n$, then the Definition 2.2.2 reduces to the Definition 2.2.1.

Proof. For the case $p = n$, Y is invertible, thus (2.10) is equivalent to the Lyapunov equation (2.9). Therefore, the Definition 2.2.2 indeed reduces to the Definition 2.2.1 when $p = n$.

The next proposition characterizes the horizontal space for metric g^1.

Proposition 2.2.7. Under metric g^1, the horizontal space at Y satisfies

$$\mathcal{H}_Y^1 \quad = \quad \left\{ Z \in \mathbb{C}^{n\times p} : Y^*Z = Z^*Y \right\}$$

$$= \left\{ Y(Y^*Y)^{-1}S + Y_\perp K | S^* = S, S \in \mathbb{C}^{p\times p}, K \in \mathbb{C}^{(n-p)\times p} \right\},$$

where $Y_\perp$ has orthonormal columns.

Proof. The result of real case can be found in [32] but the proof was omitted. For completeness, we outline the proof here. $Z \in \mathbb{C}^{n\times p}$ belongs to $\mathcal{H}_Y^1$ if and only if Z is orthogonal to $\mathcal{V}_Y$ under the metric g_Y^1, i.e., $g_Y^1(Z, Y\Omega) = \langle Z, Y\Omega\rangle_{\mathbb{C}^{n\times p}} = \langle Y^*Z, \Omega\rangle_{\mathbb{C}^{n\times p}} = 0, \forall \Omega = -\Omega^*$. This is equivalent to $Y^*Z = Z^*Y$. The second equality can be obtained by writing any $Z \in \mathcal{H}_Y^1$ as $Z = Y(Y^*Y)^{-1}S + Y_\perp K$ as $Y(Y^*Y)^{-1}$ and $Y_\perp$ forms a basis for the column space of $\mathbb{C}^{n\times p}$, and verify that $S = S^*$

Proposition 2.2.8. If we use g^1 as our Riemannian metric on $\mathbb{C}_*^{n\times p}$, then the orthogonal projections of any $A \in \mathbb{C}^{n\times p}$ to $\mathcal{V}_Y$ and $\mathcal{H}_Y^1$ are

$$P_Y^{\mathcal{V}}(A) = Y\Omega, \quad P_Y^{\mathcal{H}^1}(A) = A - Y\Omega,$$

where Ω is the skew-symmetric matrix that solves the Lyapunov equation

$$\Omega Y^*Y + Y^*Y\Omega = Y^*A - A^*Y.$$

The Second Quotient Metric

Another Riemannian metric used in [36, 37] is

$$g_Y^2(A, B) := \langle AY^*, BY^*\rangle_{\mathbb{C}^{n\times n}} = \mathrm{Re}(\mathrm{tr}((Y^*Y)A^*B)), \quad \forall A, B \in T_Y\mathbb{C}_*^{n\times p} = \mathbb{C}^{n\times p}.$$

Proposition 2.2.9. Under the metric g^2, the horizontal space at Y is characterized by

$$\begin{aligned} \mathcal{H}_Y^2 &= \left\{ Z \in \mathbb{C}^{n\times p} : (Y^*Y)^{-1}Y^*Z = Z^*Y(Y^*Y)^{-1} \right\} \\ &= \left\{ YS + Y_\perp K | S^* = S, S \in \mathbb{C}^{p\times p}, K \in \mathbb{C}^{(n-p)\times p} \right\}. \end{aligned}$$

Proof. The proof follows the same idea used in proving Proposition 2.2.7.

Proposition 2.2.10. If we use g^2 as our Riemannian metric on $\mathbb{C}_*^{n\times p}$, then the orthogonal projection of any $A \in \mathbb{C}^{n\times p}$ to vertical space $\mathcal{V}_Y$ satisfies

$$P_Y^{\mathcal{V}}(A) = Y\left(\frac{(Y^*Y)^{-1}Y^*A - A^*Y(Y^*Y)^{-1}}{2}\right) = YSkew\left((Y^*Y)^{-1}Y^*A\right),$$

and the orthogonal projection of any $A \in \mathbb{C}^{n\times p}$ to the horizontal space $\mathcal{H}_Y^2$ is

$$\begin{aligned} P_Y^{\mathcal{H}^2}(A) &= A - P_Y^{\mathcal{V}}(A) \\ &= Y\left(\frac{(Y^*Y)^{-1}Y^*A + A^*Y(Y^*Y)^{-1}}{2}\right) + Y_\perp Y_\perp^* A \\ &= YHerm\left((Y^*Y)^{-1}Y^*A\right) + Y_\perp Y_\perp^* A. \end{aligned}$$

The Third Quotient Metric

The third Riemannian metric for $\mathbb{C}_*^{n\times p}$ is motivated by the Riemannian metric of $\mathcal{H}_+^{n,p}$ and the diffeomorphism between $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{H}_+^{n,p}$. We know that β is a submersion. Every tangent vector of $\mathcal{H}_+^{n,p}$ corresponds to a tangent vector of $\mathbb{C}_*^{n\times p}$. We can use the Riemannian metric of $\mathcal{H}_+^{n,p}$ and the correspondence of tangent vectors between $\mathcal{H}_+^{n,p}$ and $\mathbb{C}_*^{n\times p}$ to define a Riemannian metric for $\mathbb{C}_*^{n\times p}$. A natural first attempt would be to use

$$g_Y(A,B) := \langle \mathrm{D}\,\beta(Y)[A], \mathrm{D}\,\beta(Y)[B]\rangle_{\mathbb{C}^{n\times n}} = \langle YA^* + AY^*, YB^* + BY^*\rangle_{\mathbb{C}^{n\times n}},$$

which is however not a Riemannian metric because it is not positive-definite. To see this, notice that $\ker(\mathrm{D}\,\beta(Y)[\cdot]) = \mathcal{V}_Y$. Consider $C \neq 0 \in \mathcal{V}_Y$, then $g_Y^3(C,C) = 0$. To modify this definition for g^3, we can use the Riemannian metric g^2 and the decomposition $T_Y\mathbb{C}_*^{n\times p} = \mathcal{H}_Y^2 \oplus \mathcal{V}_Y$, by which $A \in T_Y\mathbb{C}_*^{n\times p}$ can be uniquely decomposed as

$$A = A^{\mathcal{V}} + A^{\mathcal{H}^2},$$

where $A^{\mathcal{V}} \in \mathcal{V}_Y$ and $A^{\mathcal{H}^2} \in \mathcal{H}_Y^2$. Now define g^3 as

$$g_Y^3(A,B) \;:=\; \left\langle \mathrm{D}\,\beta(Y)[A^{\mathcal{H}^2}], \mathrm{D}\,\beta(Y)[B^{\mathcal{H}^2}]\right\rangle_{\mathbb{C}^{n\times n}} + g_Y^2\left(A^{\mathcal{V}}, B^{\mathcal{V}}\right)$$

$$
\begin{aligned}
&= \langle \mathrm{D}\,\beta(Y)[A], \mathrm{D}\,\beta(Y)[B]\rangle_{\mathbb{C}^{n\times n}} + \left\langle P_Y^{\mathcal{V}}(A)Y^*, P_Y^{\mathcal{V}}(B)Y^*\right\rangle_{\mathbb{C}^{n\times n}}, \\
&= \langle YA^* + AY^*, YB^* + BY^*\rangle_{\mathbb{C}^{n\times n}} + \left\langle P_Y^{\mathcal{V}}(A)Y^*, P_Y^{\mathcal{V}}(B)Y^*\right\rangle_{\mathbb{C}^{n\times n}}
\end{aligned}
$$

where $P_Y^{\mathcal{V}}$ is the projection of any tangent vector of $\mathbb{C}_*^{n\times p}$ to the vertical space $\mathcal{V}_Y$. It is straightforward to verify that g^3 defined above is now a Riemannian metric. With the definition (2.17), the properties $\operatorname{tr}(UV) = \operatorname{tr}(VU)$ for two matrices U, V and $\operatorname{Re}(\operatorname{tr}(C+C^*)) = 2\operatorname{Re}(\operatorname{tr}(C))$, we have

$$
\forall A, B \in \mathcal{H}_Y^2, \quad g_Y^3(A,B) = \langle YA^* + AY^*, YB^* + BY^*\rangle_{\mathbb{C}^{n\times n}} = 2\,\langle AY^*Y + YA^*Y, B\rangle_{\mathbb{C}^{n\times p}}. \tag{2.14}
$$

Proposition 2.2.11. Under metric g^3, the horizontal space at Y is the same set as $\mathcal{H}_Y^2$. That is,

$$
\begin{aligned}
\mathcal{H}_Y^3 &= \left\{Z \in \mathbb{C}^{n\times p} : (Y^*Y)^{-1}Y^*Z = Z^*Y(Y^*Y)^{-1}\right\} \\
&= \left\{YS + Y_\perp K | S^* = S, S \in \mathbb{C}^{p\times p}, K \in \mathbb{C}^{(n-p)\times p}\right\}.
\end{aligned}
$$

Proof. $Z \in \mathcal{H}_Y^3$ if and only if $g_Y^3(Z, Y\Omega) = 0$ for all $\Omega = \Omega^*$. That is, $\forall \Omega = \Omega^*$,

$$
\langle YZ^* + ZY^*, 2Y\Omega Y^*\rangle_{\mathbb{C}^{n\times n}} + \left\langle P_Y^{\mathcal{V}}(Z)Y^*, Y\Omega Y^*\right\rangle_{\mathbb{C}^{n\times n}} = 0.
$$

Hence we must have

$$
\langle YZ^* + ZY^*, 2Y\Omega Y^*\rangle_{\mathbb{C}^{n\times n}} = 0 \tag{2.15a}
$$

and

$$
\left\langle P_Y^{\mathcal{V}}(Z)Y^*, Y\Omega Y^*\right\rangle_{\mathbb{C}^{n\times n}} = 0. \tag{2.15b}
$$

(2.15a) is equivalent to

$$
\langle YZY^*Y, \Omega\rangle_{\mathbb{C}^{n\times n}} = 0 \quad \forall \Omega = \Omega^*.
$$

Hence YZY^*Y must be Hermitian since Ω is an arbitrary skew Hermitian matrix. Therefore Z is in $\mathcal{H}_Y^2$ as well and hence (2.15b) is satisfied.

Thus we have shown that

$$\mathcal{H}_Y^3 = \mathcal{H}_Y^2 = \left\{Z \in \mathbb{C}^{n\times p} : (Y^*Y)^{-1}Y^*Z = Z^*Y(Y^*Y)^{-1}\right\}.$$

Proposition 2.2.12. If we use g^3 as our Riemannian metric on $\mathbb{C}_*^{n\times p}$, then the orthogonal projection of any $A \in \mathbb{C}^{n\times p}$ to vertical space $\mathcal{V}_Y$ satisfies

$$P_Y^{\mathcal{V}}(A) = Y\left(\frac{(Y^*Y)^{-1}Y^*A - A^*Y(Y^*Y)^{-1}}{2}\right) = Y skew((Y^*Y)^{-1}Y^*A),$$

and the orthogonal projection of any $A \in \mathbb{C}^{n\times p}$ to the horizontal space $\mathcal{H}_Y^3$ is

$$\begin{aligned} P_Y^{\mathcal{H}^3}(A) &= A - P_Y^{\mathcal{V}}(A) \\ &= Y\left(\frac{(Y^*Y)^{-1}Y^*A + A^*Y(Y^*Y)^{-1}}{2}\right) + Y_\perp Y_\perp^* A \\ &= Y Herm\left((Y^*Y)^{-1}Y^*A\right) + Y_\perp Y_\perp^* A. \end{aligned}$$

2.2.2 The Riemannian Quotient Manifold

Now, we will show that the quotient manifold becomes a Riemannian quotient manifold with the Riemannian metrics induced from the total space.

First we show in the following lemma the relationship between the horizontal lifts of the quotient tangent vector $\xi_{\pi(Y)}$ lifted at different representatives in $[Y]$.

Lemma 2.2.13. Let η be a vector field on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$, and let $\bar{\eta}$ be the horizontal lift of η. Then for each $Y \in \mathbb{C}_*^{n\times p}$, we have

$$\bar{\eta}_{YO} = \bar{\eta}_Y O$$

for all $O \in \mathcal{O}_p$.

Proof. [4, Prop. A.8] gives a proof based on metric g^1 for the real case, and [36, Lemma 5.1] proves the result for metric g^2. For completeness, we will provide a proof applying to all three metrics g^{i}.

By the definition of horizontal lift, we have

$$\eta_{\mathrm{\pi}(Y)} = \eta_{\mathrm{\pi}(YO)} = \mathrm{D}\,\mathrm{\pi}(YO)[\overline{\eta}_{YO}].$$

On the other hand, notice that $\mathrm{\pi}(Y) = \mathrm{\pi}(YO)$. Taking the differential w.r.t. Y we have

$$\mathrm{D}\,\mathrm{\pi}(Y)[A] = \mathrm{D}\,\mathrm{\pi}(YO)[AO] \quad \forall A \in \mathbb{C}^{n\times p}.$$

In particular, let $A = \overline{\eta}_Y \in \mathcal{H}_Y$ we have

$$\eta_{\mathrm{\pi}(Y)} = \mathrm{D}\,\mathrm{\pi}(Y)[\overline{\eta}_Y] = \mathrm{D}\,\mathrm{\pi}(YO)[\overline{\eta}_Y O].$$

Thus, we have

$$\mathrm{D}\,\mathrm{\pi}(YO)[\overline{\eta}_Y O] = \mathrm{D}\,\mathrm{\pi}(YO)[\overline{\eta}_{YO}]$$

So

$$\overline{\eta}_Y O - \overline{\eta}_{YO} \in \ker(\mathrm{D}\,\mathrm{\pi}(YO)[\cdot]) = \mathcal{V}_{YO}.$$

Now, one can verify that for each g^{i} and $YO\Omega \in \mathcal{V}_{YO}$, $g^{\mathrm{i}}_{YO}(\overline{\eta}_Y O, YO\Omega) = 0$. So $\overline{\eta}_Y O$ is orthogonal to $\mathcal{V}_Y$ and hence $\overline{\eta}_Y O \in \mathcal{H}^{\mathrm{i}}_{YO}$. So we have

$$\overline{\eta}_Y O - \overline{\eta}_{YO} \in \mathcal{H}^{\mathrm{i}}_{YO}.$$

Therefore $\overline{\eta}_Y O - \overline{\eta}_{YO} \in \mathcal{V}_{YO} \cap \mathcal{H}^{\mathrm{i}}_{YO} = \{0\}$ and we complete the proof.

Recall from [30, Section 3.6.2] that if the expression $g_Y(\bar{\xi}_Y, \bar{\zeta}_Y)$ does not depend on the choice of $Y \in [Y]$ for every $\mathrm{\pi}(Y) \in \mathbb{C}^{n\times p}_*/\mathcal{O}_p$ and every $\xi_{\mathrm{\pi}(Y)}, \zeta_{\mathrm{\pi}(Y)} \in T_{\mathrm{\pi}(Y)}\mathbb{C}^{n\times p}_*/\mathcal{O}_p$, then

$$g_{\mathrm{\pi}(Y)}(\xi_{\mathrm{\pi}(Y)}, \zeta_{\mathrm{\pi}(Y)}) := g_Y(\bar{\xi}_Y, \bar{\zeta}_Y) \tag{2.16}$$

defines a Riemannian metric on the quotient manifold $\mathbb{C}^{n\times p}_*/\mathcal{O}_p$. By Lemma 2.2.13, it is straightforward to verify that each Riemannian metric g^{i} on $\mathbb{C}^{n\times p}_*$ induces a Riemannian metric on $\mathbb{C}^{n\times p}_*/\mathcal{O}_p$. The quotient manifold $\mathbb{C}^{n\times p}_*/\mathcal{O}_p$ endowed with a Riemannian metric defined in (2.16) is called a *Riemannian quotient manifold.* By abuse of notation, we use g^{i} for denoting Riemannian metrics on both total space $\mathbb{C}_*^{n\times p}$ and quotient space $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$.

2.3 Riemannian Gradients

In this section, we tackle functions defined on the manifold of PSD fixed-rank matrices and their gradients. First, let us define the Fréchet gradient of a real-valued function f.

2.3.1 The Fréchet Gradient of a Real-valued Function

A real-valued function $f(X)$ defined on complex matrices is not holomorphic, thus $f(X)$ does not have a complex derivative with respect to $X \in \mathbb{C}^{n\times n}$.

For any real vector space $\mathcal{E}$, the inner product on $\mathcal{E}$ is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{E}}$. For real matrices $A, B \in \mathbb{R}^{m\times n}$, the HilbertSchmidt inner product is $\langle A, B\rangle_{\mathbb{R}^{m\times n}} = \operatorname{tr}(A^T B)$.

The linear spaces of complex matrices can be regarded as vector spaces over $\mathbb{R}$. Let $\operatorname{Re}(A)$ and $\operatorname{Im}(B)$ represent the real and imaginary parts of a complex matrix A. For $A, B \in \mathbb{C}^{m\times n}$, the real inner product for the real vector space $\mathbb{C}^{m\times n}$ then equals

$$\langle A, B\rangle_{\mathbb{C}^{m\times n}} := \operatorname{Re}(\operatorname{tr}(A^* B)), \tag{2.17}$$

where * is the conjugate transpose. We emphasize that (2.17) is a real inner product, rather than the complex Hilbert-Schmidt inner product. It is straightforward to verify that (2.17) can be written as

$$\langle A, B\rangle_{\mathbb{C}^{m\times n}} = \operatorname{tr}(\operatorname{Re}(A)^T \operatorname{Re}(B)) + \operatorname{tr}(\operatorname{Im}(A)^T \operatorname{Im}(B)) = \langle \operatorname{Re}(A), \operatorname{Re}(B)\rangle_{\mathbb{R}^{m\times n}} + \langle \operatorname{Im}(A), \operatorname{Im}(B)\rangle_{\mathbb{R}^{m\times n}} .$$

With the real inner product (2.17) for the real vector space $\mathbb{C}^{m\times n}$, a Fréchet derivative for any real valued function $f(X)$ can be defined as

$$\nabla f(X) = \frac{\partial f(X)}{\partial \operatorname{Re}(X)} + \mathbb{i}\frac{\partial f(X)}{\partial \operatorname{Im}(X)} \in \mathbb{C}^{m\times n}. \tag{2.18}$$

In particular, for $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_F^2$ with a linear operator $\mathcal{A}$, the Fréchet derivative (2.18) becomes

$$\nabla f(X) = \mathcal{A}^*(\mathcal{A}(X) - b)$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$.

2.3.2 Riemannian Gradient of the Embedded Manifold

The *Riemannian gradient* of f at $X \in \mathcal{H}_+^{n,p}$, denoted by $\operatorname{grad} f(X)$, is the projection of $\nabla f(X)$ onto $T_X\mathcal{H}_+^{n,p}$ ([30, Sect. 3.6.1]):

$$\operatorname{grad} f(X) = P_X^t(\nabla f(X)),$$

where P_X^t denotes the orthogonal projection onto $T_X\mathcal{H}_+^{n,p}$.

2.3.3 Riemannian Gradient of the Riemannian Quotient Manifold

For any real-valued function f defined on $\mathcal{H}_+^{n,p}$, there is a real-valued function F defined on $\mathbb{C}_*^{n\times p}$ that induces f: for any $X = YY^* \in \mathcal{H}_+^{n,p}$, $F(Y) := f \circ \beta(Y) = f(YY^*)$. Recall that $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is diffeomorphic to $\mathcal{H}_+^{n,p}$ under $\tilde{\beta}$. Given a smooth real-valued function f on $\mathcal{H}_+^{n,p}$, the corresponding cost function h on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is defined as

$$\begin{aligned} h : \mathbb{C}_*^{n\times p}/\mathcal{O}_p &\to \mathbb{R} \\ \pi(Y) &\mapsto f(\tilde{\beta}(\pi(Y))) = f(\beta(Y)) = f(YY^*). \end{aligned} \tag{2.19}$$

The Riemannian gradient of h at $\pi(Y)$ is a tangent vector in $T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. The next theorem shows that the horizontal lift of $\operatorname{grad} h(\pi(Y))$ can be obtained from the Riemannian gradient of F defined on $\mathbb{C}_*^{n\times p}$.

Theorem 2.3.1. The horizontal lift of the gradient of h at $\pi(Y)$ is the Riemannian gradient of F at Y. That is,

$$\overline{\operatorname{grad} h(\pi(Y))}_Y = \operatorname{grad} F(Y).$$

Therefore, $\operatorname{grad} F(Y)$ is automatically in $\mathcal{H}_Y$.

Proof. See [30, Section 3.6.2].

The next proposition summarizes the expression of $\operatorname{grad} F(Y)$ under different metrics.

Proposition 2.3.2. Let f be a smooth real-valued function defined on $\mathcal{H}_+^{n,p}$ and let $F : \mathbb{C}_*^{n\times p} \to \mathbb{R} : Y \mapsto f(YY^*)$. Assume $YY^* = X$. Then the Riemannian gradient of F is given by

$$\operatorname{grad} F(Y) = \begin{cases} 2\nabla f(YY^*)Y, & \text{if using metric } g^1 \\ 2\nabla f(YY^*)Y(Y^*Y)^{-1}, & \text{if using metric } g^2 \\ \left(I - \frac{1}{2}P_Y\right)\nabla f(YY^*)Y(Y^*Y)^{-1} & \text{if using metric } g^3 \end{cases}$$

where ∇f denotes Fréchet gradient (2.18) and $P_Y = Y(Y^*Y)^{-1}Y^*$.

Proof. Let $A \in T_Y\mathbb{C}_*^{n\times p} = \mathbb{C}^{n\times p}$. By the chain rule, we have

$$\mathrm{D}\, F(Y)[A] = \mathrm{D}\, f(YY^*)[YA^* + AY^*].$$

This yields to

$$g_Y^{\mathrm{i}}(\operatorname{grad} F(Y), A) = g_X\left(\operatorname{grad} f(YY^*), YA^* + AY^*\right),$$

where g_X is the metric (2.4). Since $YA^* + AY^* \in T_{YY^*}\mathcal{H}_+^{n,p}$, we have

$$\begin{aligned} g_X\left(\operatorname{grad} f(YY^*), YA^* + AY^*\right) &= \left\langle P_{YY^*}^t(\nabla f(YY^*)), YA^* + YA^*\right\rangle_{\mathbb{C}^{n\times n}} \\ &= \left\langle \nabla f(YY^*), YA^* + AY^*\right\rangle_{\mathbb{C}^{n\times n}}. \end{aligned}$$

It is straightforward to verify that

$$\begin{aligned} \left\langle \nabla f(YY^*), YA^* + AY^*\right\rangle_{\mathbb{C}^{n\times n}} &= g_Y^1\left(2\nabla f(YY^*)Y, A\right) \\ &= g_Y^2\left(2\nabla f(YY^*)Y(Y^*Y)^{-1}, A\right), \end{aligned}$$

which yields the expression of $\operatorname{grad} F(Y)$ under g^1 and g^2.

The Riemannian gradient for g^3 is due to

$$\begin{aligned}\left\langle P^t_{YY^*}(\nabla f(YY^*)), YA^* + YA^* \right\rangle_{\mathbb{C}^{n\times n}} &= g^3_Y\left(\left(I - \frac{1}{2}P_Y\right) P^t_X(\nabla f(YY^*))Y(Y^*Y)^{-1}, A\right)\\ &= g^3_Y\left(\left(I - \frac{1}{2}P_Y\right) \nabla f(YY^*)Y(Y^*Y)^{-1}, A\right).\end{aligned}$$

2.4 Riemannian Hessians

Definition 2.4.1 ([30, definition 5.5.1]). Given a real-valued function f on a Riemannian manifold $\mathcal{M}$, the *Riemannian Hessian* of f at a point x in $\mathcal{M}$ is the linear mapping $\operatorname{Hess} f(x)$ of $T_x\mathcal{M}$ into itself defined by

$$\operatorname{Hess} f(x)[\xi_x] = \nabla_{\xi_x} \operatorname{grad} f$$

for all ξ_x in $T_x\mathcal{M}$, where ∇ is the Riemannian connection on $\mathcal{M}$.

2.4.1 Riemannian Hessian of the Embedded Manifold

For a real-valued function $f(X)$ defined on the Euclidean space $\mathbb{C}^{n\times n}$, the Fréchet Hessian $\nabla^2 f(X)$ is defined in the sense of the Fréchet derivative; see Appendix A.0.2 for the definition of the Fréchet Hessian. The *Riemannian Hessian* of f at X, denoted by $\operatorname{Hess} f(X)$ is related to, but different from its Fréchet Hessian.

The following proposition gives the Riemannian Hessian of f. The proof follows similar ideas as in [7, Prop. 5.10] and [38, Prop. 2.3] where a second-order retraction based on a simple power expansion is constructed. We will leave the outline of the proof in Appendix B.0.1.

Proposition 2.4.1. Let $f(X)$ be a real-valued function defined on $\mathcal{H}^{n,p}_+$. Let $X \in \mathcal{H}^{n,p}_+$ and $\xi_X \in T_X\mathcal{H}^{n,p}_+$. Then the Riemannian Hessian operator of f at X is given by

$$\operatorname{Hess} f(X)[\xi_X] = P^t_X(\nabla^2 f(X)[\xi_X]) + P^p_X\left(\nabla f(X)(X^\dagger \xi^p_X)^* + (\xi^p_X X^\dagger)^* \nabla f(X)\right)$$

where $\xi^s_X = P^s_X(\xi_X)$ and $\xi^p_X = P^p_X(\xi_X)$ and P^t_X and P^p_X are defined in (2.7).

2.4.2 Riemannian Hessian of the Quotient Manifold

Recall that the cost function h on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is defined in (2.19). In this section, we summarize the Riemannian Hessian of h under the three different metrics g^{i}. The proofs are tedious calculations and given in Appendix C.0.1.

Proposition 2.4.2. 1. Using g^1, the Riemannian Hessian of h is given by

$$\overline{\left(\operatorname{Hess} h(\pi(Y))[\xi_{\pi(Y)}]\right)}_Y = P_Y^{\mathcal{H}^1}\left(2\nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*]Y + 2\nabla f(YY^*)\overline{\xi}_Y\right).$$

2. Using g^2, the Riemannian Hessian of h is given by

$$\begin{aligned}\overline{\left(\operatorname{Hess} h(\pi(Y))[\xi_{\pi(Y)}]\right)}_Y \quad = \quad & P_Y^{\mathcal{H}^2}\Big\{2\nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*]Y(Y^*Y)^{-1} \\ & +\nabla f(YY^*)P_Y^{\perp}\overline{\xi}_Y(Y^*Y)^{-1} + P_Y^{\perp}\nabla f(YY^*)\overline{\xi}_Y(Y^*Y)^{-1} \\ & +2skew(\overline{\xi}_Y Y^*)\nabla f(YY^*)Y(Y^*Y)^{-2} \\ & + 2skew\{\overline{\xi}_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\}Y(Y^*Y)^{-1}\Big\}.\end{aligned}$$

3. Using g^3, the Riemannian Hessian of h is given by

$$\begin{aligned}\overline{\left(\operatorname{Hess} h(\pi(Y))[\xi_{\pi(Y)}]\right)}_Y \quad = \quad & \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*]Y(Y^*Y)^{-1} \\ & +(I - P_Y)\nabla f(YY^*)(I - P_Y)\overline{\xi}_Y(Y^*Y)^{-1}.\end{aligned}$$

2.5 Computational Tools

In this section, we introduce some computational tools that will be used later.

2.5.1 Retraction

A retraction is essentially a first-order approximation to the exponential map; see [30, Def. 4.1.1].

Definition 2.5.1 ([30, Def. 4.1.1]). A retraction on a manifold $\mathcal{M}$ is a smooth mapping R from the tangent bundle $T\mathcal{M}$ onto $\mathcal{M}$ with the following properties. Let R_x denote the restriction of R to $T_x\mathcal{M}$.

1. $R_x(0_x) = x$, where 0_x denotes the zero element of $T_x\mathcal{M}$.

2. With the canonical identification $T_{0_x}T_x\mathcal{M} \equiv T_x\mathcal{M}$, R_x satisfies

$$\mathrm{D}\, R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}},$$

where $\mathrm{id}_{T_x\mathcal{M}}$ denotes the identity mapping on $T_x\mathcal{M}$.

Suppose $\mathcal{M}$ is an embedded submanifold of a Euclidean space $\mathcal{E}$, then by [39, Props. 3.2 and 3.3], the mapping R from the tangent bundle $T\mathcal{M}$ to the manifold $\mathcal{M}$ defined by

$$R : \begin{cases} T\mathcal{M} \to \mathcal{M} \\ (x, u) \mapsto P_{\mathcal{M}}(x + u) \end{cases} \tag{2.20}$$

is a retraction, where $P_{\mathcal{M}}$ is the orthogonal projection onto the manifold $\mathcal{M}$ with respect to the Euclidean distance, that is, the closest point. In our case $\mathcal{M} = \mathcal{H}_+^{n,p}$ and $\mathcal{E} = \mathbb{C}^{n\times n}$. Hence, a retraction on $\mathcal{H}_+^{n,p}$ is defined by the truncated SVD:

$$R_X(Z) := P_{\mathcal{H}_+^{n,p}}(X + Z) = \sum_{\mathrm{i}=1}^{p} \sigma_{\mathrm{i}}(X + Z) v_{\mathrm{i}} v_{\mathrm{i}}^*,$$

where v_{i} is the singular vector of $X + Z$ corresponding to the ith largest singular value $\sigma_{\mathrm{i}}(X + Z)$.

The retraction on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ can be defined using the retraction on the total space $\mathbb{C}_*^{n\times p}$. For any $A \in T_Y\mathbb{C}_*^{n\times p}$ and a step size $\tau > 0$,

$$\overline{R}_Y(\tau A) := Y + \tau A,$$

is a retraction on $\mathbb{C}_*^{n\times p}$ if $Y+\tau A$ remains full rank, which is ensured for small enough τ. Then Lemma 2.2.13 indicates that $\overline{R}$ satisfies the conditions of [30, Prop. 4.1.3], which implies that

$$R_{\pi(Y)}(\tau\eta_{\pi(Y)}) := \pi(\overline{R}_Y(\tau\overline{\eta}_Y)) = \pi(Y+\tau\overline{\eta}_Y) \tag{2.21}$$

defines a retraction on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ for a small enough step size $\tau > 0$.

2.5.2 Vector Transport

The vector transport is a mapping that transports a tangent vector from one tangent space to another tangent space.

Definition 2.5.2 ([30, definition 8.1.1]). A vector transport on a manifold $\mathcal{M}$ is a smooth mapping

$$T\mathcal{M} \oplus T\mathcal{M} \to T\mathcal{M} : (\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}(\xi_x) \in T\mathcal{M}$$

satisfying the following properties for all $x \in \mathcal{M}$:

1. (Associated retraction) There exists a retraction R, called the retraction associated with $\mathcal{T}$, such that the following diagram commutes

$$\begin{tikzcd} (\eta_x, \xi_x) \arrow[r, "\mathcal{T}"] \arrow[d] & \mathcal{T}_{\eta_x}(\xi_x) \arrow[d, "\Pi"] \\ \eta_x \arrow[r, "R"] & \Pi(\mathcal{T}_{\eta_x}(\xi_x)) \end{tikzcd}$$

   where $\Pi(\mathcal{T}_{\eta_x}(\xi_x))$ denotes the foot of the tangent vector $\mathcal{T}_{\eta_x}(\xi_x)$.

2. (Consistency) $\mathcal{T}_{0_x}\xi_x = \xi_x$ for all $\xi_x \in T_x\mathcal{M}$;

3. (Linearity) $\mathcal{T}_{\eta_x}(a\xi_x + b\zeta_x) = a\mathcal{T}_{\eta_x}(\xi_x) + b\mathcal{T}_{\eta_x}(\zeta_x)$.

Let $\xi_X, \eta_X \in T_X\mathcal{H}_+^{n,p}$ and let R be a retraction on $\mathcal{H}_+^{n,p}$. By [30, section 8.1.3], the projection of one tangent vector onto another tangent space is a vector transport,

$$\mathcal{T}_{\eta_X}\xi_X := P^t_{R_X(\eta_X)}\xi_X, \tag{2.22}$$

where P_Z^t is the projection operator onto $T_Z\mathcal{H}_+^{n,p}$. Namely, we first apply retraction to $X+\eta_X$ to arrive at a new point on the manifold, then we project the old tangent vector ξ_X onto the tangent space at that new point.

A vector transport on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ introduced in [30, Section 8.1.4] is projection to horizontal space.

$$\overline{\left(\mathcal{T}_{\eta_{\pi(Y)}}\xi_{\pi(Y)}\right)}_{Y+\overline{\eta}_Y} := P^{\mathcal{H}}_{Y+\overline{\eta}_Y}(\overline{\xi}_Y). \tag{2.23}$$

It can be shown that this vector transport is actually the differential of the retraction R defined in (2.21) (see [30, Section 8.1.2]) since

$$\begin{aligned}
\mathrm{D}\, R_{\pi(Y)}(\eta_{\pi(Y)})[\xi_{\pi(Y)}] &= \mathrm{D}\,\pi\left(\overline{R}_Y(\overline{\eta}_Y)\right)\left[\mathrm{D}\,\overline{R}_Y(\overline{\eta}_Y)[\overline{\xi}_Y]\right] \\
&= \mathrm{D}\,\pi(Y+\overline{\eta}_Y)\left[\overline{\xi}_Y\right] \\
&= \mathrm{D}\,\pi(Y+\overline{\eta}_Y)\left[P^{\mathcal{H}}_{Y+\overline{\eta}_Y}(\overline{\xi}_Y)\right].
\end{aligned}$$

Based on the projection formulae in Section 2.2.1, we can obtain formulae of vector transports using different Riemannian metrics. Denote $Y_2 = Y_1 + \overline{\eta}_{Y_1}$. If we use metric g^1, then

$$\overline{\left(\mathcal{T}_{\eta_{\pi(Y_1)}}\xi_{\pi(Y_1)}\right)}_{Y_1+\overline{\eta}_{Y_1}} = \overline{\xi}_{Y_1} - Y_2\Omega,$$

where Ω solves the Lyapunov equation

$$Y_2^*Y_2\Omega + \Omega Y_2^*Y_2 = Y_2^*\overline{\xi}_{Y_1} - \overline{\xi}_{Y_1}^* Y_2.$$

If we use metric g^2 or g^3, then

$$\begin{aligned}
\overline{\left(\mathcal{T}_{\eta_{\pi(Y_1)}}\xi_{\pi(Y_1)}\right)}_{Y_1+\overline{\eta}_{Y_1}} &= \overline{\xi}_{Y_1} - P^{\mathcal{V}}_{Y_2}(\overline{\xi}_{Y_1}) \\
&= \overline{\xi}_{Y_1} - Skew\left((Y_2^*Y_2)^{-1}Y_2^*\overline{\xi}_{Y_1}\right) \\
&= Y_2\left(\frac{(Y_2^*Y_2)^{-1}Y_2^*\overline{\xi}_{Y_1} + \overline{\xi}_{Y_1}^* Y_2(Y_2^*Y_2)^{-1}}{2}\right) + Y_{2\perp}Y_{2\perp}^*\overline{\xi}_{Y_1}.
\end{aligned}$$

3. A UNIFIED FRAMEWORK FOR RIEMANNIAN OPTIMIZATION ON FIXED-RANK HERMITIAN PSD MATRICES VIA QUOTIENT GEOMETRY

3.1 Introduction

In this chapter, we will consider three straightforward ideas and methodologies for solving (1.2).

3.2 Three Different Methodologies

3.2.1 The Burer–Monteiro Method

The Burer–Monteiro method [40], is to solve the unconstrained problem

$$\min_{Y \in \mathbb{C}^{n\times p}} F(Y) := f(YY^*). \tag{3.1}$$

As proven in Appendix A, the chain rule of Fréchet derivatives gives

$$\nabla F(Y) = 2\nabla f(YY^*)Y \in \mathbb{C}^{n\times p}.$$

The gradient descent method simply takes the form of

$$Y_{n+1} = Y_n - \tau\nabla F(Y_n) = Y_n - \tau 2\nabla f(Y_n Y_n^*)Y_n,$$

which is one of the simplest low-rank algorithms. The nonlinear conjugate gradient and quasi-Newton type methods, like L-BFGS, can also be easily used for (3.1). On the other hand, $F(Y) = F(YO)$ for any unitary matrix $O \in \mathbb{O}^{p\times p}$, where

$$\mathcal{O}_p = \{O \in \mathbb{C}^{p\times p} : O^*O = OO^* = I\}.$$

Even though this ambiguity of unitary matrices is never explicitly addressed in the Burer–Monteiro method, in this section, we will prove that the gradient descent and nonlinear conjugate gradient methods for solving (3.1) are exactly equivalent to the Riemannian gradient descent and Riemannian conjugate gradient methods on a quotient manifold with a Euclidean metric, which is also referred to as the Bures-Wasserstein metric [4, 32]. Thus the convergence of the Burer–Monteiro method can be understood within the context of Riemannian optimization on a quotient manifold.

3.2.2 Riemannian Optimization with the Embedded Geometry of $\mathcal{H}_+^{n,p}$

Another natural approach is to regard $\mathcal{H}_+^{n,p}$ as an embedded manifold in the Euclidean space $\mathbb{C}^{n\times n}$. For instance, Riemannian optimization algorithms on the embedded manifold of low-rank matrices and tensors are quite efficient and popular [41, 42]. Even though it is possible to study $\mathcal{H}_+^{n,p} \subset \mathbb{C}^{n\times n}$ as a complex manifold, we will regard $\mathbb{C}^{n\times n}$ as a $2n^2$-dimensional real vector space and $\mathcal{H}_+^{n,p} \subset \mathbb{C}^{n\times n}$ as a manifold over $\mathbb{R}$ since $f(X)$ is real-valued. In particular, the embedded geometry of $\mathcal{S}_+^{n,p}$, representing the set of real symmetric PSD low-rank matrices, was studied in [43].

3.2.3 Riemannian Optimization by Using Quotient Geometry

The third approach is to consider the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. Since there is a one-to-one correspondence between $X = YY^* \in \mathcal{H}_+^{n,p}$ and $\pi(Y) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p$, the optimization problem (1.2) is equivalent to

$$\begin{array}{ll} \underset{\pi(Y)}{\text{minimize}} & h(\pi(Y)) \\ \text{subject to} & \pi(Y) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p \end{array}, \tag{3.2}$$

where the cost function h is defined as $h(\pi(Y)) = F(Y) = f(YY^*)$.

For the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$, one can first choose a metric for its total space $\mathbb{C}_*^{n\times p}$, which induces a Riemannian metric on the quotient manifold under suitable conditions. In particular, a special metric was used in [36] to construct efficient Riemannian optimization algorithms for the problem (3.1). The horizontal lift of the Riemannian gradient for $h(\pi(Y))$ under this particular metric satisfies

$$(\operatorname{grad} h(\pi(Y)))_Y = \nabla F(Y)(Y^*Y)^{-1} = 2\nabla f(YY^*)Y(Y^*Y)^{-1}. \tag{3.3}$$

From the representation of the Riemannian gradient (3.3), we see that this approach generates different algorithms from the simpler Burer–Monteiro approach.

3.3 The Riemannian Conjugate Gradient Method

In this section, we introduce the Riemannian conjugate gradient (RCG) method described as Algorithm 1 in [41] with the geometric variant of PolakRibiére (PR+) for computing the conjugate direction. It is possible to explore other methods such as the limited-memory version of the Riemannian BFGS method (LRBFGS) as in [44]. However, RCG performs very well on a wide variety of problems and is easier to implement for our numerical examples.

We first summarize two Riemannian CG algorithms in Algorithm 1 and Algorithm 2 below. Algorithm 1 is the RCG on the embedded manifold for solving (1.2) and Algorithm 2 is the RCG on the quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^{\mathrm{i}})$ for solving (3.2). We remark that the explicit constants 0.0001 and 0.5 in the Armijo backtracking are chosen for convenience.

3.4 Equivalence Between Burer–Monteiro CG and RCG on the Riemannian Quotient Manifold with the Bures-Wasserstein Metric $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^1)$

In this section, we focus on establishing two equivalences in algorithms.First, we show that the Burer–Monteiro CG method, which is simply applying the CG method for the unconstrained problem (3.1), is equivalent to RCG on the Riemannian quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^1)$ with our retraction and vector transport defined in the previous sections.

Theorem 3.4.1. Using retraction (2.21), vector transport (2.23) and metric g^1, Algorithm 2 is equivalent to the conjugate gradient method solving (3.1) in the sense that they produce exactly the same iterates if started from the same initial point.

Algorithm 1 Riemannian Conjugate Gradient on the embedded manifold $\mathcal{H}_+^{n,p}$

```
Require: initial iterate $X_0 \in \mathcal{H}_+^{n,p}$, initial gradient $\xi_0 = \operatorname{grad} f(X_0)$, initial conjugate direction $\eta_0 = -\operatorname{grad} f(X_0)$, tolerance $\varepsilon > 0$
for $k = 1, 2, \dots$ do
    Compute an initial step $t_k$. For special cost functions, it is possible to compute:
        $t_k = \arg\min_t f(X_{k-1} + t\eta_{k-1})$
    Perform Armijo backtracking to find the smallest integer $m \geq 0$ such that
        $f(X_{k-1}) - f(R_{X_{k-1}}(0.5^m t_k \eta_{k-1})) \geq -0.0001 \times 0.5^m t_k g_{X_{k-1}}(\xi_{k-1}, \eta_{k-1})$
        $\zeta_k := 0.5^m t_k \eta_{k-1}$
    Obtain the new iterate by retraction
        $X_k = R_{X_{k-1}}(\zeta_k)$                                ▷ See Algorithm 6
    Compute gradient
        $\xi_k := \operatorname{grad} f(X_k)$                              ▷ See Algorithm 3
    Check convergence
        if $\|\xi_k\| := \sqrt{g_{X_k}(\xi_k, \xi_k)} < \varepsilon$ or $f(X_k) < \varepsilon$, then break
    Compute a conjugate direction by PR$_+$ and vector transport
        $\eta_k = -\xi_k + \beta_k \mathcal{T}_{\zeta_k}(\eta_{k-1})$,                    ▷ See Algorithm 4, 5
        with $\beta_k := \dfrac{g_{X_k}\left(\xi_k, \xi_k - \mathcal{T}_{\zeta_k}(\xi_{k-1})\right)}{g_{X_{k-1}}\left(\xi_{k-1}, \xi_{k-1}\right)}$.
end for
```

Algorithm 2 Riemannian Conjugate Gradient on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with metric g^{i}

```
Require: initial iterate $Y_0 \in \pi^{-1}(\pi(Y_0))$, initial horizontal lift of gradient $\overline{\xi}_0 = \operatorname{grad} F(Y_0)$,
    initial conjugate direction $\overline{\eta}_0 = -\overline{\xi}_0$, tolerance $\varepsilon > 0$
for $k = 1, 2, \ldots$ do
    Compute an initial step $t_k$. For special cost functions, it is possible to compute:
        $t_k = \arg\min_t F(Y_{k-1} + t\overline{\eta}_{k-1})$
    Perform Armijo backtracking to find the smallest integer $m \geq 0$ such that
        $$F(Y_{k-1}) - F(\overline{R}_{Y_{k-1}}(0.5^m t_k \overline{\eta}_{k-1})) \geq -0.0001 \times 0.5^m t_k g^{\mathrm{i}}_{Y_{k-1}}(\overline{\xi}_{k-1}, \overline{\eta}_{k-1})$$
        $$\overline{\zeta}_k := 0.5^m t_k \overline{\eta}_k$$
    Obtain the new iterate by retraction
        $Y_k = \overline{R}_{Y_{k-1}}(\overline{\zeta}_k)$
    Compute the horizontal lift of gradient
        $\overline{\xi}_k := \overline{(\operatorname{grad} h(\pi(Y_k)))}_{Y_k} = \operatorname{grad} F(Y_k)$        ▷ See Algorithm 7
    Check convergence
        if $\left\|\overline{\xi}_k\right\| := \sqrt{g^{\mathrm{i}}_{Y_k}(\overline{\xi}_k, \overline{\xi}_k)} < \varepsilon$ or $F(Y_k) < \varepsilon$, then break
    Compute a conjugate direction by PR$_+$ and vector transport
        $\overline{\eta}_k = -\overline{\xi}_k + \beta_k \overline{(\mathcal{T}_{\zeta_k}\eta_{k-1})}_{Y_k}$,        ▷ See Algorithm 8
        $$\text{with } \beta_k := \frac{g^{\mathrm{i}}_{Y_k}\left(\operatorname{grad} F(Y_k), \operatorname{grad} F(Y_k) - \overline{(\mathcal{T}_{\zeta_k}\xi_{k-1})}_{Y_k}\right)}{g^{\mathrm{i}}_{Y_{k-1}}\left(\operatorname{grad} F(Y_{k-1}), \operatorname{grad} F(Y_{k-1})\right)}.$$
end for
```

Proof. First of all, for g^1, the Riemannian gradient of F at Y is $\operatorname{grad} F(Y) = 2\nabla f(YY^*)Y$, which is equal to the Fréchet gradient of $F(Y) = f(YY^*)$ at Y. Since vector transport is the orthogonal projection to the horizontal space, the PR_+ β_k used in Riemannian CG becomes

$$\beta_k = \frac{g^1_{Y_k}\left(\operatorname{grad} F(Y_k), \operatorname{grad} F(Y_k) - P^{\mathcal{H}^1}_{Y_k}(\operatorname{grad} F(Y_{k-1}))\right)}{g^1_{Y_{k-1}}\left(\operatorname{grad} F(Y_{k-1}), \operatorname{grad} F(Y_{k-1})\right)}. \tag{3.4}$$

Now observe that

$$P^{\mathcal{H}^1}_{Y_k}(\operatorname{grad} F(Y_{k-1})) = \operatorname{grad} F(Y_{k-1}) - P^{\mathcal{V}}_{Y_k}(\operatorname{grad} F(Y_{k-1}))$$

and g^1 is equivalent to the classical inner product for $\mathbb{C}^{n\times p}$. Hence β_k computed by (3.4) is equal to PR_+ β_k in conjugate gradient for (3.1).

The first conjugate direction is $\eta_1 = -\operatorname{grad} F(Y_1) = -\nabla F(Y_1)$, so Burer–Monteiro CG coincides with Riemannian CG for the first iteration. It remains to show that η_k generated in Riemannian CG by

$$\eta_k = -\xi_k + \beta_k P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1})$$

is equal to η_k generated in Burer–Monteiro CG for each $k \geq 2$. It suffices to show that

$$P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1}) = \eta_{k-1}, \quad \forall k \geq 2.$$

Equivalently we need to show that for all $k \geq 2$, the Lyapunov equation

$$(Y_k^* Y_k)\Omega + \Omega(Y_k^* Y_k) = Y_k^* \eta_{k-1} - \eta_{k-1}^* Y_k \tag{3.5}$$

only has trivial solution $\Omega = 0$. By invertibility of the equation, this means that we only need to show the right hand side is zero. We prove it by induction.

For $k = 2$, $\eta_{k-1} = \eta_1 = -\xi_1 = -\operatorname{grad} F(Y_1)$. The following computations show that the RHS of (3.5) satisfies

$$Y_2^*\eta_1 - \eta_1^* Y_2 \quad = \quad -Y_2^*\xi_1 + \xi_1^* Y_2$$

$$
\begin{aligned}
&= -(Y_1 - c\xi_1)^*\xi_1 + \xi_1^*(Y_1 - c\xi_1) \\
&= \xi_1^* Y_1 - Y_1^*\xi_1 \\
&= Y_1^*(2\nabla f(Y_1Y_1^*))Y_1 - Y_1^*(2\nabla f(Y_1Y_1^*))Y_1 \\
&= 0.
\end{aligned}
$$

Hence $\Omega = 0$ and $P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1}) = \eta_{k-1}$ for $k = 2$.

Now suppose for $k \geq 2$, the RHS of (3.5) is 0 and hence $P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1}) = \eta_{k-1}$ holds. Then the RHS of the Lyapunov equation of step $k+1$ is

$$
\begin{aligned}
Y_{k+1}^*\eta_k - \eta_k^* Y_{k+1} &= (Y_k + c\eta_k)^*\eta_k - \eta_k^*(Y_k + c\eta_k) \\
&= Y_k^*\eta_k - \eta_k^* Y_k \\
&= Y_k^*\left(-\xi_k + \beta_k P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1})\right) - \left(-\xi_k + \beta_k P^{\mathcal{H}^1}_{Y_k}(\eta_{k-1})\right)^* Y_k \\
&= Y_k^*(-\xi_k + \beta_k\eta_{k-1}) - (-\xi_k + \beta_k\eta_{k-1})^* Y_k \\
&= -Y_k^*\xi_k + \xi_k^* Y_k \\
&= -Y_k^*(2\nabla f(Y_kY_k^*))Y_k + Y_k^*(2\nabla f(Y_kY_k^*))Y_k \\
&= 0.
\end{aligned}
$$

Hence $P^{\mathcal{H}^1}_{Y_{k+1}}(\eta_k) = \eta_k$ also holds. We have thus proven that Riemannian CG is equivalent to Burer–Monteiro CG.

Since the gradient descent corresponds to $\beta_k \equiv 0$, the same discussion also implies the following

Corollary 3.4.2. Using retraction (2.21) and metric g^1, the Riemannian gradient descent on the quotient manifold is equivalent to the Burer–Monteiro gradient descent method with a suitable step size (3.2.1) in the sense that they produce exactly the same iterates.

3.5 Equivalence Between RCG on Embedded Manifold and RCG on the Quotient Manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^3)$

In this subsection we show that Algorithm 1 is equivalent to Algorithm 2 with Riemannian metric g^3, a specific initial line-search in step 5, a specific retraction and a specific vector transport. The idea is to take advantage of the diffeomorphism $\tilde{\beta}$ between $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{H}_+^{n,p}$, as well as the fact that the metric g^3 of $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ is induced from the metric of $\mathcal{H}_+^{n,p}$.

The Lemma below shows that there is a one-to-one correspondence between $\operatorname{grad} f$ and $\operatorname{grad} h$.

Lemma 3.5.1. If we use g^3 as the Riemannian metric for $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$, then the Riemannian gradient of f and h is related by the diffeomorphism $\tilde{\beta}$ in the following way:

$$(\mathrm{D}\,\tilde{\beta})(\pi(Y))[\operatorname{grad} h(\pi(Y))] = \operatorname{grad} f(YY^*).$$

Proof. Recall that $\beta = \tilde{\beta} \circ \pi$ and we have Theorem 2.3.1. By chain rule and the definition of horizontal lift we have

$$\begin{aligned} LHS = (\mathrm{D}\,\tilde{\beta})(\pi(Y))[\operatorname{grad} h(\pi(Y))] &= (\mathrm{D}\,\tilde{\beta})(\pi(Y))\left[\mathrm{D}\,\pi(Y)\left[\overline{\operatorname{grad} h(\pi(Y))}_Y\right]\right] \\ &= \mathrm{D}\,\beta(Y)\left[\overline{\operatorname{grad} h(\pi(Y))}_Y\right] \\ &= \mathrm{D}\,\beta(Y)\left[\operatorname{grad} F(Y)\right], \end{aligned}$$

where the second equality follows from the inverse direction of chain rule.

Now recall that $F = f \circ \beta$. Let $A \in \mathbb{C}^{n\times p}$ then

$$\mathrm{D}\,F(Y)[A] = \mathrm{D}\,f(YY^*)[YA^* + YA^*].$$

Let $X = YY^*$. Then we have

$$g_Y^3(\operatorname{grad} F(Y), A) = g_X(\operatorname{grad} f(YY^*), YA^* + AY^*).$$

Applying the definition of g^3, we have

$$g_X\left(\mathrm{D}\,\beta(Y)[\operatorname{grad} F(Y)], YA^* + AY^*\right) = g_X\left(\operatorname{grad} f(YY^*), YA^* + AY^*\right),$$

or

$$g_X\left(LHS, YA^* + AY^*\right) = g_X\left(RHS, YA^* + AY^*\right).$$

Now notice that A is arbitrary and $YA^* + AY^*$ can be any tangent vector in $T_X\mathcal{H}_+^{n,p}$. Hence we must have $LHS = RHS$

Since $\tilde{\beta}$ is a diffeomorphism bewteen $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{H}_+^{n,p}$, $D\tilde{\beta}(\pi(Y))[\cdot]$ defines an isomorphism between the tangent space $T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ and $T_{YY^*}\mathcal{H}_+^{n,p}$. We denote this isomorphism by $L_{\pi(Y)}$. When the tangent space is clear from the context, $\pi(Y)$ is omitted and we only use the notation L for simplicity. The previous lemma then simply reads as

$$L_{\pi(Y)}(\operatorname{grad} h(\pi(Y))) = \operatorname{grad} f(\tilde{\beta}(\pi(Y))).$$

In Algorithm 1, we have a retraction R^E and a vector transport $\mathcal{T}^E$ on the embedded manifold $\mathcal{H}_+^{n,p}$, (with the superscript E for *Embedded*), such that R^E is the retraction associated with $\mathcal{T}^E$. Then we claim that there is a retraction R^Q and a vector transport $\mathcal{T}^Q$, (with the superscript Q denoting *Quotient*), on the Riemannian quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^3)$, such that Algorithm 2 is equivalent to Algorithm 1. The idea is again to use the diffeomorphism $\tilde{\beta}$ and the isomorphism $L_{\pi(Y)}$. We give the desired expression of R^Q and $\mathcal{T}^Q$ as follows.

$$R^Q_{\pi(Y)}(\xi_{\pi(Y)}) := \tilde{\beta}^{-1}\left(R^E_{\tilde{\beta}(\pi(Y))}\left(L(\xi_{\pi(Y)})\right)\right). \tag{3.6}$$

$$\mathcal{T}^Q_{\eta_{\pi(Y)}}(\xi_{\pi(Y)}) := L^{-1}_{\pi(Y_2)}\left(\mathcal{T}^E_{L(\eta_{\pi(Y)})}\left(L(\xi_{\pi(Y)})\right)\right), \tag{3.7}$$

where $\pi(Y_2)$ is in $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ such that $\tilde{\beta}(\pi(Y_2))$ denotes the foot of the tangent vector $\mathcal{T}^E_{L(\eta_{\pi(Y)})}\left(L(\xi_{\pi(Y)})\right)$.

Now it remains to show that R^Q defined in (3.6) is indeed a retraction and $\mathcal{T}^Q$ defined in (3.7) is indeed a vector transport.

Lemma 3.5.2. R^Q defined in (3.6) is a retraction.

Proof. First it is easy to see that $R^Q_{\pi(Y)}(0_{\pi(Y)}) = \pi(Y)$. Then we also have for all $v_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$

$$\begin{aligned}
\mathrm{D}\, R^Q_{\pi(Y)}(0_{\pi(Y)})[v_{\pi(Y)}] &= (\mathrm{D}\,\tilde{\beta}^{-1})(\tilde{\beta}(\pi(Y))\left[\mathrm{D}\, R^E_{\tilde{\beta}(\pi(Y))}(0)\left[\mathrm{D}\, L(0)\left[v_{\pi(Y)}\right]\right]\right] \\
&= (\mathrm{D}\,\tilde{\beta}^{-1})(\tilde{\beta}(\pi(Y))\left[\mathrm{D}\, R^E_{\tilde{\beta}(\pi(Y))}(0)\left[L(v_{\pi(Y)})\right]\right] \\
&= (\mathrm{D}\,\tilde{\beta}^{-1})(\tilde{\beta}(\pi(Y))\left[L(v_{\pi(Y)})\right] \\
&= \left(\mathrm{D}\,\tilde{\beta}(\pi(Y))\right)^{-1}[L(v_{\pi(Y)})] \\
&= L^{-1}(L(v_{\pi(Y)})) \\
&= v_{\pi(Y)}
\end{aligned}$$

Hence $\mathrm{D}\, R^Q_{\pi(Y)}(0_{\pi(Y)})[\cdot]$ is an identity map.

Lemma 3.5.3. $\mathcal{T}^E$ defined in (3.7) is a vector transport and R^Q is the retraction associated with $\mathcal{T}^E$.

Proof. Consistency and linearity are straightforward. It thus suffices to verify that the foot of $\mathcal{T}^Q_{\eta_{\pi(Y)}}(\xi_{\pi(Y)})$ is equal to $R^Q_{\pi(Y)}(\eta_{\pi(Y)})$. Since R^E is the associated retraction with $\mathcal{T}^E$, the foot of $\mathcal{T}^E_{L(\eta_{\pi(Y)})}(L(\xi_{\pi(Y)}))$ is equal to $R^E_{\tilde{\beta}(\pi(Y))}\left(L(\eta_{\pi(Y)})\right)$, which we denote by $\tilde{\beta}(\pi(Y_2))$ for some $\pi(Y_2)$. Hence $R^Q_{\pi(Y)}(\eta_{\pi(Y)}) = \tilde{\beta}^{-1}\left(R^E_{\tilde{\beta}(\pi(Y))}\left(L(\eta_{\pi(Y)})\right)\right) = \pi(Y_2)$.

Furthermore, we have that $\mathcal{T}^Q_{\eta_{\pi(Y)}}(\xi_{\pi(Y)}) = L^{-1}_{\pi(Y_2)}\left(\mathcal{T}^E_{L(\eta_{\pi(Y)})}\left(L(\xi_{\pi(Y)})\right)\right)$ is a tangent vector in $T_{\pi(Y_2)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. Hence, the foot of $\mathcal{T}^Q_{\eta_{\pi(Y)}}(\xi_{\pi(Y)})$ is also $\pi(Y_2)$.

Finally, in order to reach an equivalence, we also need the initial step size to match the one in step 5 of Algorithm 2. We simply replace the original initial step size t_k by

$$t_k = \arg\min_t f(Y_k Y_k^* + t(Y_k \eta_k^* + \eta_k Y_k^*)).$$

This value of t_k now is equivalent to the initial step size in step 5 of Algorithm 1. This gives us the following result:

Theorem 3.5.4. With the newly constructed initial step size, retraction and vector transport in this subsection, Algorithm 2 for solving (3.2) is equivalent to Algorithm 1 solving (1.2) in the sense that they produce exactly the same iterates.

3.6 Implementation details

The algorithms in this section can be applied for minimizing any smooth function $f(X)$ in (1.2). For problems with large n, however, it is advisable to avoid constructing and storing the Fréchet derivative $\nabla f(X) \in \mathbb{C}^{n\times n}$ explicitly. Instead, one directly computes the matrix-vector multiplications $\nabla f(X)U$. In the PhaseLift problem [8], for example, these matrix-vector multiplications can be implemented via the FFT at a cost of $\mathcal{O}(pn\log n)$ when $U \in \mathbb{C}^{n\times p}$; see [36].

Below, we detail the calculations needed in Algorithms 1 and 2. When giving flop counts, we assume that $\nabla f(X)U \in \mathbb{C}^{n\times p}$ can be computed in $spn\log n$ flops with s small. For g^2 and g^3 in Algorithms 7 and 8, we use forward slash "/" and backslash "\" in Matlab command to compute the inverse of Y^*Y.

3.6.1 Embedded manifold

```
**Algorithm 3** Calculate the Riemannian gradient grad $f(X)$
**Require:** $X = U\Sigma U^* \in \mathcal{H}_+^{n,p}$
**Ensure:** grad $f(X) = UHU^* + U_pU^* + UU_p^* \in T_X\mathcal{H}_+^{n,p}$
    $T \leftarrow \nabla f(X)U$                ▷ # $spn\log n$ flops
    $H \leftarrow U^*T$                        ▷ # $p^2(2n-1)$ flops
    $U_p \leftarrow T - UH$                    ▷ # $np + np(2p-1)$ flops
```

In implementation, we observe a vector transport that has better numerical performance if we only keep the first term in the above sum of H_2 and the second term of U_{2p} in Algorithm 4, which is outlined in Algorithm 5.

Algorithm 4 Calculate the vector transport by projection to tangent space $P^t_{X_2}(\nu)$

```
Require: $X_1 = U_1\Sigma_1 U_1^*$, $X_2 = U_2\Sigma_2 U_2^*$ and tangent vector $\nu = U_1 H_1 U_1^* + U_{p1}U_1^* + U_1 U_{p1}^* \in$
    $T_{X_1}\mathcal{H}_+^{n,p}$.
Ensure: $P^t_{X_2}(\nu) = U_2 H_2 U_2^* + U_{p2}U_2^* + U_2 U_{p2}^*$
        $A \leftarrow U_1^* U_2$                                                ▷ # $p^2(2n-1)$ flops
        $H_2^{(1)} \leftarrow A^* H_1 A$,   $U_p^{(1)} \leftarrow U_1(H_1 A)$        ▷ # $3p^2(2p-1) + np(2p-1)$ flops
        $H_2^{(2)} \leftarrow U_2^* U_{p1} A$,   $U_p^{(2)} \leftarrow U_{p1} A$         ▷ # $p^2(2n-1) + 2np(2p-1)$ flops
        $H_2^{(3)} \leftarrow H_2^{(2)^*}$,   $U_p^{(3)} \leftarrow U_1(U_{1_p}^* U_2)$     ▷ # $np(2p-1) + p^2(2n-1)$ flops
        $H_2 \leftarrow H_2^{(1)} + H_2^{(2)} + H_2^{(3)}$                              ▷ # $2p^2$ flops
        $U_{p2} \leftarrow U_p^{(1)} + U_p^{(2)} + U_p^{(3)}$,   $U_{p2} \leftarrow U_{p2} - U_2(U_2^* U_{p2})$   ▷ #
    $3np + np(2p-1) + p^2(2n-1)$ flops
```

Algorithm 5 Calculate the simpler form of vector transport used in implementation that has a better performance $P^t_{X_2}(\nu)$

```
Require: $X_1 = U_1\Sigma_1 U_1^*$, $X_2 = U_2\Sigma_2 U_2^*$ and tangent vector $\nu = U_1 H_1 U_1^* + U_{p1}U_1^* + U_1 U_{p1}^* \in$
    $T_{X_1}\mathcal{H}_+^{n,p}$.
Ensure: $P^t_{X_2}(\nu) = U_2 H_2 U_2^* + U_{p2}U_2^* + U_2 U_{p2}^*$
        $A \leftarrow U_1^* U_2$                                      ▷ # $p^2(2n-1)$ flops
        $H_2 \leftarrow A^* H_1 A$                                    ▷ # $2p^2(2p-1)$ flops
        $U_p \leftarrow U_{p1} A$                                     ▷ # $np(2p-1)$ flops
        $U_{p2} \leftarrow U_p - U_2(U_2^* U_p)$                      ▷ # $np + p^2(2n-1) + np(2p-1)$ flops
```

Algorithm 6 Calculate the retraction $R_X(Z) = P_{\mathcal{H}_+^{n,p}}(X + Z)$

```
Require: $X = U\Sigma U^* \in \mathcal{H}_+^{n,p}$, tangent vector $Z = UHU^* + U_p U^* + UU_p^*$.
Ensure: $P_{\mathcal{H}_+^{n,p}}(X + Z) = U_+\Sigma_+ U_+^*$.
        $(Q, R) \leftarrow \mathrm{qr}(U_p, 0)$   $M \leftarrow \begin{bmatrix} \Sigma + H & R^* \\ R & 0 \end{bmatrix}$      ▷ # $20np^2$ flops
        $[V, S] \leftarrow \mathrm{eig}(M)$                                        ▷ $O(p^3)$ flops
        $\Sigma+ \leftarrow S(1:p, 1:p)$,   $U_+ \leftarrow \begin{bmatrix} U & Q \end{bmatrix} V(:, 1:p)$      ▷ # $np(4p-1)$ flops
```

3.6.2 Quotient manifold

Algorithm 7 Calculate the Riemannian gradient grad $F(Y)$

```
Require: $Y \in \mathbb{C}_*^{n\times p}$
Ensure: $T = \operatorname{grad} F(Y)$
if metric is $g^1$ then
    $T \leftarrow 2\nabla f(YY^*)Y.$                    ▷ # $2spn\log n$ flops
else if metric is $g^2$ then
    $Z \leftarrow Y(Y^*Y)^{-1}$                         ▷ # $np(2p-1)+p^2(2n-1)+O(p^3)$ flops
    $T \leftarrow 2\nabla f(YY^*)Z$                     ▷ # $2spn\log n$ flops
else if metric is $g^3$ then
    $Z \leftarrow Y(Y^*Y)^{-1}$                         ▷ # $np(2p-1)+p^2(2n-1)+O(p^3)$ flops
    $T \leftarrow 2\nabla f(YY^*)Z$                     ▷ # $2spn\log n$ flops
    $M \leftarrow Y^*T, \quad T \leftarrow T-\frac{1}{2}ZM$   ▷ # $p^2(2n-1)+np+2np^2$ flops
end if
```

Algorithm 8 Calculate the quotient vector transport $P^{\mathcal{H}}_{Y_2}(h_1)$

```
Require: $Y_1 \in \mathbb{C}_*^{n\times p}$, $Y_2 \in \mathbb{C}_*^{n\times p}$ and horizontal vector $h_1 \in \mathcal{H}_{Y_1}$.
Ensure: $h_2 = P^{\mathcal{H}}_{Y_2}(h_1) \in \mathcal{H}_{Y_2}$.
if metric is $g^1$ then
        $E \leftarrow Y_2^* Y_2$                                   ▷ # $p^2(2n-1)$ flops
        $(Q, S) \leftarrow \mathrm{eig}(E), \quad d \leftarrow \mathrm{diag}(S)$                 ▷ # $O(p^3)$ flops
        $\lambda \leftarrow d\Big[1, 1, \cdots, 1\Big] + \Big[1, 1, \cdots, 1\Big]^T d^T$     ▷ # $2p^2$ flops
        $A \leftarrow Q^*(Y_2^* h_1 - h_1^* Y_2)Q$                  ▷ # $p^2(2n-1) + np + 2p^2(2p-1)$ flops
        $\Omega \leftarrow Q(A./\lambda)Q^*$                         ▷ # $p^2 + 2p^2(2p-1)$ flops
        $h_2 \leftarrow h_1 - Y_2\Omega$                              ▷ # $np + np(2p-1)$ flops
else if metric is $g^2$ or $g^3$ then
        $\tilde{\Omega} \leftarrow (Y^*Y)^{-1}(Y_2^* h_1)$             ▷ # $2p^2(2p-1) + p^2(2n-1) + O(p^3)$ flops
        $\Omega \leftarrow \frac{1}{2}(\tilde{\Omega} - \tilde{\Omega}^*)$                   ▷ # $2p^2$ flops
        $h_2 \leftarrow h_1 - Y_2\Omega$                              ▷ # $np + np(2p-1)$ flops
end if
```

3.6.3 Initial guess for the line search

The initial guess for the line search generally depends on the expression of the cost function $f(X)$. For the important case of $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_F^2$ where $\mathcal{A}$ is a linear operator and b is a matrix, the initial guess for embedded CG requires solving a linear equation and for quotient CG it requires solving a cubic equation. Below this calculation is detailed for b of size mn for some m and assuming that $\mathcal{A}(X), \mathcal{A}(T)$ and $\mathcal{A}(Y\eta^*)$ can be evaluated in $sp^\alpha n \log n$ flops for $X \in \mathcal{H}_+^{n,p}$, $T \in T_X\mathcal{H}_+^{n,p}$ and $Y, \eta \in \mathbb{C}_*^{n\times p}$.

Algorithm 9 Calculate the initial guess $t_* = \arg\min_t f(X + tT)$

```
Require: $X \in \mathcal{H}_+^{n,p}$ and a descend direction $T \in T_X\mathcal{H}_+^{n,p}$
Ensure: $t_* = \arg\min_t f(X + tT) = \arg\min_t \frac{1}{2}\|\mathcal{A}(X + tT) - b\|_F^2$
        $R \leftarrow \mathcal{A}(X) - b$                    ▷ # $sp^\alpha n \log n + mn$ flops
        $S \leftarrow \mathcal{A}(T)$                        ▷ # $sp^\alpha n \log n$ flops
        $t_* \leftarrow -\frac{\langle R,S\rangle}{\langle S,S\rangle}$                  ▷ # $4mn - 1$ flops
```

```
Algorithm 10 Calculate the initial guess $t_* = \arg\min_t F(Y + t\eta)$
Require: $Y \in \mathbb{C}_*^{n\times p}$, a descend direction $\eta \in \mathcal{H}_Y$,
Ensure: $t_* = \arg\min_t F(Y + t\eta) = \arg\min_t \frac{1}{2}\|\mathcal{A}((Y + t\eta)(Y + t\eta)^*) - b\|_F^2$
    $c_0 \leftarrow \mathcal{A}(YY^*) - b$                                                    ▷ # $sp^\alpha n \log n + mn$ flops
    $c_1^{(1)} \leftarrow \mathcal{A}(Y\eta^*)$,   $c_1^{(2)} \leftarrow \mathcal{A}(\eta Y^*)$,   $c_1 \leftarrow c_1^{(1)} + c_1^{(2)}$     ▷ # $2sp^\alpha n \log n + mn$ flops
    $c_2 \leftarrow \mathcal{A}(\eta\eta^*)$                                                      ▷ # $sp^\alpha n \log n$ flops
    $d_4 \leftarrow \langle c_2, c_2\rangle$,   $d_3 \leftarrow 2\langle c_2, c_1\rangle$                                     ▷ # $4mn - 1$ flops
    $d_2 \leftarrow 2\langle c_2, c_0\rangle + \langle c_1, c_1\rangle$,   $d_1 \leftarrow 2\langle c_1, c_0\rangle$                      ▷ # $6mn - 1$ flops
    $C \leftarrow \begin{bmatrix} 4d_4 & 3d_3 & 2d_2 & d_1 \end{bmatrix}$
    $S \leftarrow roots(C)$,   $t_* \leftarrow$ the smallest real positive root in $S$
```

3.7 Concluding Remarks

In this chapter, we have shown that the nonlinear conjugate gradient method on the unconstrained Burer–Monteiro formulation for Hermitian PSD fixed-rank constraints is equivalent to a Riemannian conjugate gradient method on a quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with the Bures-Wasserstein metric g^1, retraction, and vector transport. We have also shown that the Riemannian conjugate gradient method on the embedded geometry of $\mathcal{H}_+^{n,p}$ is equivalent to a Riemannian conjugate gradient method on a quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with a metric g^3, a special retraction, and a special vector transport. With these equivalences, we are able to unify three different methodologies within the same framework via optimizations on the quotient Riemannian manifold and conduct a fair comparison of the performance of algorithms.

4. CONDITION NUMBER ANALYSIS OF RIEMANNIAN HESSIANS AND RANK-DEFICIENCY EFFECTS

4.1 Introduction

In many applications, (1.2) or (3.2) is often used for solving (1.1). In [45], it was proven that first-order and second-order optimality conditions for the nonconvex Burer–Monteiro approach are sufficient to find the global minimizer of certain convex semi-definite programs under certain assumptions. In practice, even if the global minimizer of (1.1) has a known rank r, one might consider solving (1.2) or (3.2) for Hermitian PSD matrices with fixed rank $p > r$. For instance, in PhaseLift [8] and interferometry recovery [11], the minimizer to (1.1) is rank one, but in practice optimization over the set of PSD Hermitian matrices of rank p with $p \geq 2$ is often used because of a larger basin of attraction [11, 36]. If $p > r$, then an algorithm that solves (1.2) or (3.2) can generate a sequence that goes to the boundary of the manifold. Numerically, the smallest $p - r$ singular values of the iterates X_k will become very small as $k \to \infty$.

In this chapter, we analyze the eigenvalues of the Riemannian Hessian near the global minimizer. More specifically, we will obtain upper and lower bounds of the Rayleigh quotient at the point $X = YY^*$ (or $\pi(Y)$) that is close to the global minimizer $\hat{X} = \hat{Y}\hat{Y}^*$ (or $\pi(\hat{Y})$).

We first define the Rayleigh quotient and the condition number of the Riemannian Hessian.

Definition 4.1.1 (Rayleigh quotient of Riemannian Hessian). The Rayleigh quotient of the Riemannian Hessian of $(\mathcal{H}_+^{n,p}, g)$ is defined by

$$\rho^E(X, \zeta_X) = \frac{g_X(\operatorname{Hess} f(X)[\zeta_X], \zeta_X)}{g_X(\zeta_X, \zeta_X)}$$

for $\zeta_X \in T_X \mathcal{H}_+^{n,p}$.

The Rayleigh quotient of the Riemannian Hessian of $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^{\mathrm{i}})$ is defined by

$$\rho^{\mathrm{i}}(\pi(Y), \xi_{\pi(Y)}) = \frac{g^{\mathrm{i}}_{\pi(Y)}(\operatorname{Hess} h(\pi(Y))[\xi_{\pi(Y)}], \xi_{\pi(Y)})}{g^{\mathrm{i}}_{\pi(Y)}(\xi_{\pi(Y)}, \xi_{\pi(Y)})}$$

for $\xi_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$. If the Rayleigh quotient has lower bound a and upper bound b, then we define $\frac{b}{a}$ as the upper bound on the condition number of the Riemannian Hessian.

4.2 The Rayleigh Quotient Estimates

We assume that the Fréchet Hessian $\nabla^2 f$ is well conditioned when restricted to the tangent space. Formally, our bounds will be stated in terms of the constants A, B defined in the following assumption:

Assumption 4.2.1. For a fixed $\epsilon > 0$, there exist constants $A > 0$ and $B > 0$ such that for all X with $\left\|X - \hat{X}\right\|_F < \epsilon$, the following inequality holds for all $\zeta_X \in T_X\mathcal{H}_+^{n,p}$.

$$A\|\zeta_X\|_F^2 \leq \left\langle \nabla^2 f(X)[\zeta_X], \zeta_X \right\rangle_{\mathbb{C}^{n\times n}} \leq B\|\zeta_X\|_F^2.$$

Observe that this assumption is always satisfied for sufficiently small ϵ when f is smooth. However, the condition number B/A might be large in general. An important case for which this assumption holds everywhere is $f(X) = \frac{1}{2}\|X - H\|_F^2$ with H a given Hermitian PSD matrix. In this case, $\nabla^2 f(X)$ is the identity operator thus $A = B = 1$.

The main result in this chapter is given in the following theorem.

Theorem 4.2.1. Let $\hat{X} = \hat{Y}\hat{Y}^*$ be the global minimizer of (1.1) with rank $r \leq p$. For $X = YY^*$ near $\hat{X}$ where $Y \in \mathbb{C}_*^{n\times p}$, let $\zeta_X \in T_X\mathcal{H}_+^{n,p}$ be any tangent vector at X, $\xi_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ be any tangent vector at $\pi(Y)$, and $\overline{\xi}_Y \in \mathcal{H}_Y^{\mathrm{i}}$ be its horizontal lift at Y w.r.t. the metric g^{i}. Let $X = U\Sigma U^*$ denote the compact SVD of X and denote the ith diagonal entry of Σ to be σ_{i} with $\sigma_1 \geq \cdots \geq \sigma_p > 0$. Under the Assumption 4.2.1, for any arbitrary tangent vectors ζ_X and $\xi_{\pi(Y)}$, the following bounds hold:

1. For the embedded manifold,

$$A - \frac{2}{\sigma_p}\|\nabla f(X)\| \leq \rho^E(X, \zeta_X) \leq B + \frac{2}{\sigma_p}\|\nabla f(X)\|.$$

2. For the quotient manifold metric g^{i},

$$2A\sigma_p - 2\|\nabla f(YY^*)\| \leq \rho^1(\pi(Y), \xi_{\pi(Y)}) \leq B \cdot D^1_{\pi(Y)} + 2\|\nabla f(YY^*)\|,$$

$$2A - \frac{4(\sqrt{p}+1)}{\sigma_p}\|\nabla f(YY^*)\| \leq \rho^2(\pi(Y), \xi_{\pi(Y)}) \leq 4B + \frac{4(\sqrt{p}+1)}{\sigma_p}\|\nabla f(YY^*)\|,$$

$$A - \frac{1}{\sigma_p}\|\nabla f(YY^*)\| \leq \rho^3(\pi(Y), \xi_{\pi(Y)}) \leq B + \frac{1}{\sigma_p}\|\nabla f(YY^*)\|,$$

where $D^1_{\pi(Y)}$ satisfies $2\sigma_1 \leq D^1_{\pi(Y)} \leq 2\left(\frac{\sigma_1^2}{\sigma_p} + \sigma_1\right)$.

In particular, if $\hat{X} = \hat{Y}\hat{Y}^*$ has rank p, e.g., $\hat{X}$ has singular values $\hat{\sigma_1} \geq \cdots \geq \hat{\sigma_p} > 0$, then under the Assumption 4.2.1, we have the following limits, where the limits $X \to \hat{X}$ and $\pi(Y) \to \pi(\hat{Y})$ are taken in the sense of $\left\|X - \hat{X}\right\|_F \to 0$ and $\left\|YY^* - \hat{Y}\hat{Y}^*\right\|_F \to 0$:

1. For the embedded manifold

$$A - \frac{2}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\| \leq \lim_{X \to \hat{X}} \rho^E(X, \xi_X) \leq B + \frac{2}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\|.$$

2. For the quotient manifold metric g^{i},

$$2A\hat{\sigma}_p - 2\left\|\nabla f(\hat{X})\right\| \leq \lim_{\pi(Y)\to\pi(\hat{Y})} \rho^1(\pi(Y), \xi_{\pi(Y)}) \leq B \cdot D^1_{\pi(\hat{Y})} + 2\left\|\nabla f(\hat{X})\right\|,$$

$$2A - \frac{4(\sqrt{p}+1)}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\| \leq \lim_{\pi(Y)\to\pi(\hat{Y})} \rho^2(\pi(Y), \xi_{\pi(Y)}) \leq 4B + \frac{4(\sqrt{p}+1)}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\|,$$

$$A - \frac{1}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\| \leq \lim_{\pi(Y)\to\pi(\hat{Y})} \rho^3(\pi(Y), \xi_{\pi(Y)}) \leq B + \frac{1}{\hat{\sigma}_p}\left\|\nabla f(\hat{X})\right\|,$$

where $D^1_{\pi(\hat{Y})}$ satisfies $2\hat{\sigma}_1 \leq D^1_{\pi(\hat{Y})} \leq 2\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_p} + \hat{\sigma}_1\right)$.

Remark 4.2.2. If we further assume that $\nabla f(\hat{X}) = 0$, then the limits above can be further simplified. Such an assumption $\nabla f(\hat{X}) = 0$ may not be true in general, but it holds, e.g., for all cost functions that take the form $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_F^2$ for some matrix-valued linear operator $\mathcal{A}$, and the minimizer $\hat{X}$ for constrained minimization (1.2) or (1.1) satisfies

$f(\hat{X}) = 0$. Thus $\hat{X}$ is also the global minimizer for minimizing $f(X)$ over all $X \in \mathbb{C}$, which implies $\nabla f(\hat{X}) = 0$.

Remark 4.2.3. We can define the ratio of the upper and lower bounds of the Rayleigh quotient as the upper bound on the condition number of the Riemannian Hessian. Then under the assumption $\nabla f(\hat{X}) = 0$, the limit of the condition number of the Riemannian Hessian for the Bures-Wasserstein metric g^1 depends on the condition number of the minimizer $\hat{X}$. This reflects a significant difference between g^1 and the other two metrics.

Remark 4.2.4. For the case $\nabla f(\hat{X}) \neq 0$, if $\|\nabla f(\hat{X})\|$ is sufficiently small in the sense that

$$\|\nabla f(\hat{X})\| < a, \tag{4.1}$$

where a is equal to $\hat{\sigma}_p A/4$, $\hat{\sigma}_p A/8/(\sqrt{p}+1)$, and $\hat{\sigma}_p A/2$ for the embedded metric, the condition numbers of the embedded metric, quotient metric g^2 and g^3 are on the order of B/A. The quotient manifold with g^1 is still different from the other metrics since the condition number of its Riemannian Hessian additionally depends on the ratio $\hat{\sigma}_1/\hat{\sigma}_p$.

The rest of this subsection is the proof of Theorem 4.2.1. By the expressions of Riemannian Hessian, we have

$$\rho^E(X, \zeta_X) = \frac{\langle \nabla^2 f(X)[\zeta_X], \zeta_X \rangle_{\mathbb{C}^{n\times n}}}{g_X(\zeta_X, \zeta_X)} + \frac{g_X\left(P_X^p\left(\nabla f(X)(X^\dagger \zeta_X^p)^* + (\zeta_X^p X^\dagger)^* \nabla f(X)\right), \zeta_X\right)}{g_X(\zeta_X, \zeta_X)}.$$

$$\rho^1(\pi(Y), \xi_{\pi(Y)}) = \frac{\left\langle \nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*], Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^* \right\rangle_{\mathbb{C}^{n\times n}}}{g_Y^1(\overline{\xi}_Y, \overline{\xi}_Y)} + \frac{g_Y^1(2\nabla f(YY^*)\overline{\xi}_Y, \overline{\xi}_Y)}{g_Y^1(\overline{\xi}_Y, \overline{\xi}_Y)}.$$

$$\begin{aligned}\rho^2(\pi(Y), \xi_{\pi(Y)}) \quad = \quad & \frac{\left\langle \nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*], Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^* \right\rangle_{\mathbb{C}^{n\times n}}}{g_Y^2(\overline{\xi}_Y, \overline{\xi}_Y)} \\ & + \frac{\left\langle \nabla f(YY^*) P_Y^\perp \overline{\xi}_Y, \overline{\xi}_Y \right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\overline{\xi}_Y, \overline{\xi}_Y)} + \frac{\left\langle P_Y^\perp \nabla f(YY^*) \overline{\xi}_Y, \overline{\xi}_Y \right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\overline{\xi}_Y, \overline{\xi}_Y)} \\ & + \frac{\left\langle Y\overline{\xi}_Y^* \overline{\xi}_Y, 2\nabla f(YY^*) Y (Y^*Y)^{-1} \right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\overline{\xi}_Y, \overline{\xi}_Y)}\end{aligned}$$

$$-\frac{\left\langle \overline{\xi}_Y Y^* \overline{\xi}_Y, 2\nabla f(YY^*)Y(Y^*Y)^{-1} \right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\overline{\xi}_Y, \overline{\xi}_Y)}.$$

$$\begin{aligned}\rho^3(\pi(Y), \xi_{\pi(Y)}) &= \frac{\left\langle \nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*], Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^* \right\rangle_{\mathbb{C}^{n\times n}}}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)} \\ &\quad + \frac{g_Y^3((I-P_Y)\nabla f(YY^*)(I-P_Y)\overline{\xi}_Y(Y^*Y)^{-1}, \overline{\xi}_Y)}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)}.\end{aligned}$$

Observe that the leading terms in the above Rayleigh quotients take similar form: the numerator involves the Fréchet Hessian $\nabla^2 f$, and the denominator is the induced norm of tangent vector from the respective Riemannian metric. We call the leading term *second order term* (SOT) as it involves Fréchet Hessian of f as the second order information of f and we call the other terms that follow the leading term *first order terms* (FOTs) as they only contain the first order Fréchet gradient.

Under the Assumption 4.2.1, we get bounds of the SOT in $\rho^E(X, \zeta_X)$ as:

$$A = A\frac{\|\zeta_X\|_F^2}{g_X(\zeta_X, \zeta_X)} \leq \frac{\langle \nabla^2 f(X)[\zeta_X], \zeta_X \rangle_{\mathbb{C}^{n\times n}}}{g_X(\zeta_X, \zeta_X)} \leq B\frac{\|\zeta_X\|_F^2}{g_X(\zeta_X, \zeta_X)} = B.$$

For the quotient manifold, observe that $Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^* \in T_{YY^*}\mathcal{H}_+^{n,p}$. Hence Assumption 4.2.1 also applies and we get

$$A\frac{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2}{g_Y^{\mathrm{i}}\left(\overline{\xi}_Y, \overline{\xi}_Y\right)} \leq \frac{\left\langle \nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*], Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^* \right\rangle_{\mathbb{C}^{n\times n}}}{g_Y^{\mathrm{i}}(\overline{\xi}_Y, \overline{\xi}_Y)} \leq B\frac{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2}{g_Y^{\mathrm{i}}\left(\overline{\xi}_Y, \overline{\xi}_Y\right)}.$$

Hence the analysis of SOT for the quotient manifold now turns to analyzing $\frac{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2}{g_Y^{\mathrm{i}}\left(\overline{\xi}_Y, \overline{\xi}_Y\right)}$. We denote its infimum and supremum by

$$\begin{aligned}C_{\pi(Y)}^{\mathrm{i}} &:= \inf_{\xi_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p} \frac{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2}{g_Y^{\mathrm{i}}(\overline{\xi}_Y, \overline{\xi}_Y)}, \\ D_{\pi(Y)}^{\mathrm{i}} &:= \sup_{\xi_{\pi(Y)} \in T_{\pi(Y)}\mathbb{C}_*^{n\times p}/\mathcal{O}_p} \frac{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2}{g_Y^{\mathrm{i}}(\overline{\xi}_Y, \overline{\xi}_Y)}.\end{aligned}$$

The subscript is used to emphasize that the infimum and supremum are dependent on $\pi(Y)$. The next lemma characterizes these infimum and supremum.

Lemma 4.2.5. For any $Y \in \pi^{-1}(\pi(Y))$, let $YY^* = U\Sigma U^*$ denote the compact SVD of YY^* and denote the ith diagonal entry of Σ by σ_{i} with $\sigma_1 \geq \cdots \geq \sigma_p > 0$. Then the following estimates for the infimum $C^{\mathrm{i}}_{\pi(Y)}$ and the supremum $D^{\mathrm{i}}_{\pi(Y)}$ of $\frac{\left\|Y\bar{\xi}_Y^* + \bar{\xi}_Y Y^*\right\|_F^2}{g^{\mathrm{i}}_Y\left(\bar{\xi}_Y, \bar{\xi}_Y\right)}$ hold:

$$\begin{aligned}
C^1_{\pi(Y)} &= 2\sigma_p, \quad 2\sigma_1 \leq D^1_{\pi(Y)} \leq 2\left(\frac{\sigma_1^2}{\sigma_p} + \sigma_1\right). \\
C^2_{\pi(Y)} &= 2, \quad D^2_{\pi(Y)} = 4. \\
C^3_{\pi(Y)} &= D^3_{\pi(Y)} = 1.
\end{aligned}$$

Next we estimate the FOTs in the Rayleigh quotient. The result is given in the next lemma.

Lemma 4.2.6. Let $X = YY^*$ for any $Y \in \pi^{-1}(\pi(Y))$ with $X \in \mathcal{H}_+^{n,p}$ and $\pi(Y) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p$. Let $U\Sigma U^*$ be the compact SVD of X and denote the ith diagonal entry of Σ with $\sigma_1 \geq \cdots \geq \sigma_p > 0$. Then we have the following bounds for the FOTs in the Rayleigh quotient of the Riemannian Hessian.

1. For the embedded manifold we have

$$|\mathrm{FOT}| \leq \frac{2}{\sigma_p}\|\nabla f(X)\|.$$

2. For the quotient manifold with metric g^1 we have

$$|\mathrm{FOT}| \leq 2\|\nabla f(YY^*)\|.$$

3. For the quotient manifold with g^2 we have

$$|\mathrm{FOTs}| \leq \frac{4(\sqrt{p}+1)}{\sigma_p}\|\nabla f(YY^*)\|.$$

4. For the quotient manifold with g^3 we have

$$|\mathrm{FOTs}| \leq \frac{1}{\sigma_p}\|\nabla f(YY^*)\|.$$

The proofs for Lemma 4.2.6 and Lemma 4.2.5 are given in Section 4.4. With Lemma 4.2.6 and Lemma 4.2.5, the proof of Theorem 4.2.1 is concluded.

4.3 The Limit of the Rayleigh Quotient for a Rank-deficient Minimizer $\hat{X}$

Next, we consider the rank deficient case $p > r$ where r is the rank of the minimizer $\hat{X}$, i.e., the minimizer $\hat{X}$ lies on the boundary of the constraint manifold. Under the Assumption $\nabla f(\hat{X}) = 0$, any convergent algorithm that solves (1.2) or (3.2) will generate a sequence such that both $\sigma_{r+1}, \cdots, \sigma_p$ and $\nabla f(X)$ will vanish as $X \to \hat{X}$. We make one more assumption for a simpler quantification of the lower and upper bounds of the Rayleigh quotient near the minimizer.

Assumption 4.3.1. For a sequence $\{X_k\}$ with $X_k \in \mathcal{H}_+^{n,p}$ (or $\pi(Y_k) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p$) that converges to the minimizer $\hat{X}$ (or $\pi(\hat{Y})$), let $(\sigma_p)_k$ be the smallest nonzero singular value of $X_k = Y_k Y_k^*$, assume the following limits hold.

1. For the embedded manifold,

$$\lim_{k\to\infty} \frac{2}{(\sigma_p)_k}\|\nabla f(X_k)\| \leq \frac{A}{2}.$$

2. For the quotient manifold with metric g^1,

$$\lim_{k\to\infty} \frac{1}{(\sigma_p)_k}\|\nabla f(Y_k Y_k^*)\| \leq \frac{A}{2}.$$

3. For the quotient manifold with metric g^2,

$$\lim_{k\to\infty} \frac{4(\sqrt{p}+1)}{(\sigma_p)_k}\|\nabla f(Y_k Y_k^*)\| \leq A.$$

4. For the quotient manifold with metric g^3,

$$\lim_{k\to\infty} \frac{1}{(\sigma_p)_k} \|\nabla f(Y_k Y_k^*)\| \le \frac{A}{2}.$$

We remark that Assumption 4.3.1 may not always hold. In the next section, we will give some numerical evaluation of this assumption for four examples listed in Figure 4.3 (eigenvalue problem), Figure 4.5 (matrix completion), Figure 4.7 (phase retrieval), and Figure 4.9 (interferometry recovery). Assumption 4.3.1 holds numerically in most of these tests.

Remark 4.3.1. In general, there exists a sequence such that the FOT in $\rho^3(\pi(Y), \xi_{\pi(Y)})$ may blow up. Consider the following simple example of eigenvalue problem.

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) = \frac{1}{2}\left\|X - \hat{X}\right\|_F^2 \\ \text{subject to} & X \in \mathcal{H}_+^{3,2} \end{array},$$

where $\hat{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is a rank-1 minimizer. Suppose X takes the simple diagonal form $X = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Then we have

$$\nabla f(X) = \begin{bmatrix} \sigma_1 - 1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since $\nabla f(X) \to 0$ as $X \to \hat{X}$, we have $\sigma_1 \to 1$ and $\sigma_2 \to 0$.

Recall that the FOT in $\rho^3(\pi(Y), \xi_{\pi(Y)})$ is

$$\frac{g_Y^3((I - P_Y)\nabla f(YY^*)(I - P_Y)\overline{\xi}_Y (Y^*Y)^{-1}, \overline{\xi}_Y)}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)} = \frac{\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}}{2\|YSY^*\|_F^2 + \|Y_\perp K Y^*\|_F^2}.$$

Hence if we choose $S = 0$ and $Y_\perp K = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, we have

$$\frac{\langle \nabla f(YY^*) Y_\perp K, Y_\perp K \rangle_{\mathbb{C}^{n\times p}}}{2\|YSY^*\|_F^2 + \|Y_\perp K Y^*\|_F^2} = \frac{\sigma_1 - 1}{\sigma_2},$$

whose limit is dependent on the path that the tuple (σ_1, σ_2) goes to $(1, 0)$ and hence may blow up.

If $\hat{X}$ has rank $r < p$ and $\{X_k\}$ is a sequence that satisfies Assumption 4.3.1, then Theorem 4.2.1 implies

1. For the embedded manifold we have

$$\frac{A}{2} \leq \lim_{k\to\infty} \rho^E(X_k, \xi_{X_k}) \leq B + \frac{A}{2}.$$

2. For the quotient manifold with metric g^{i} we have

$$A \leq \lim_{k\to\infty} \frac{\rho^1(\pi(Y_k), \xi_{\pi(Y_k)})}{(\sigma_p)_k} \leq B \lim_{k\to\infty} \frac{D^1_{\pi(Y_k)}}{(\sigma_p)_k} + 2A,$$

$$A \leq \lim_{k\to\infty} \rho^2(\pi(Y_k), \xi_{\pi(Y_k)}) \leq 4B + A,$$

$$\frac{A}{2} \leq \lim_{k\to\infty} \rho^3(\pi(Y_k), \xi_{\pi(Y_k)}) \leq B + \frac{A}{2},$$

where $\lim_{k\to\infty} \frac{D^1_{\pi(Y_k)}}{(\sigma_p)_k} \geq \lim_{k\to\infty} \frac{2(\sigma_1)_k}{(\sigma_p)_k} = +\infty$ since $\sigma_p \to \hat{\sigma}_p = 0$.

Notice that the condition number in the Bures–Wassertein metric g^1 is fundamentally different from the other ones since it is the only metric that blows up.

4.4 Proof of Lemmas in This Chapter

4.4.1 Proof of Lemma 4.2.5

Proof. It is straightforward to see $C^3_{\pi(Y)} = D^3_{\pi(Y)} = 1$ by the definition of g^3.

For metric 2, write $\bar{\xi}_Y = YS + Y_\perp K$ for some $S = S^* \in \mathbb{C}^{p\times p}$ and $K \in \mathbb{C}^{n\times p}$. We have

$$\frac{\left\|Y\bar{\xi}_Y^* + \bar{\xi}_Y Y^*\right\|_F^2}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)} = 2 + \frac{2\|YSY^*\|_F^2}{\|YSY^*\|_F^2 + \|KY^*\|_F^2}.$$

Hence it is easy to see $C^2_{\pi(Y)} = 2$ when S is zero matrix and $D^2_{\pi(Y)} = 4$ when YSY^* is nonzero and K is zero matrix.

For metric 1, by its horizontal space, we can write $\bar{\xi}_Y = Y(Y^*Y)^{-1}S + Y_\perp K$ for some $S = S^* \in \mathbb{C}^{p\times p}$ and $K \in \mathbb{C}^{n\times p}$. Notice that the SVD of Y can be given as $Y = U\Sigma^{\frac{1}{2}}V^*$ where V is unitary. Let $\bar{S} = V^*SV$ and $\bar{K} = KV$, and $\bar{K}_\mathrm{i}$ be the ith column of $\bar{K}$, then

$$\begin{aligned}
\frac{\left\|Y\bar{\xi}_Y^* + \bar{\xi}_Y Y^*\right\|_F^2}{g_Y^1(\bar{\xi}_Y, \bar{\xi}_Y)} &= \frac{\|Y((Y^*Y)^{-1}S + S(Y^*Y)^{-1})Y^*\|_F^2 + 2\|KY^*\|_F^2}{\|Y(Y^*Y)^{-1}S\|_F^2 + \|K\|_F^2} \\
&= \frac{\left\|\Sigma^{-\frac{1}{2}}\bar{S}\Sigma^{\frac{1}{2}} + \Sigma^{\frac{1}{2}}\bar{S}\Sigma^{-\frac{1}{2}}\right\|_F^2 + 2\left\|\bar{K}\Sigma^{\frac{1}{2}}\right\|_F^2}{\left\|\Sigma^{-\frac{1}{2}}\bar{S}\right\|_F^2 + \left\|\bar{K}\right\|_F^2} \\
&= \frac{\sum_{\mathrm{i,j}=1}^{p}\left(\frac{\sigma_\mathrm{i}}{\sigma_\mathrm{j}} + \frac{\sigma_\mathrm{j}}{\sigma_\mathrm{i}} + 2\right)\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sum_{\mathrm{i}=1}^{p}\sigma_\mathrm{i}\left\|\bar{K}_\mathrm{i}\right\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\frac{\left|\bar{S}_\mathrm{ij}\right|^2}{\sigma_\mathrm{i}} + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2} \\
&= \frac{2\sum_{\mathrm{i,j}=1}^{p}\frac{\sigma_\mathrm{j}}{\sigma_\mathrm{i}}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sum_{\mathrm{i}=1}^{p}\sigma_\mathrm{i}\|K_\mathrm{i}\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\frac{\left|\bar{S}_\mathrm{ij}\right|^2}{\sigma_\mathrm{i}} + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2},
\end{aligned} \tag{4.2}$$

where symmetry $\bar{S}^* = \bar{S}$ is used in the last step. The lower bound is given by

$$\begin{aligned}
\frac{2\sum_{\mathrm{i,j}=1}^{p}\frac{\sigma_\mathrm{j}}{\sigma_\mathrm{i}}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sum_{\mathrm{i}=1}^{p}\sigma_\mathrm{i}\left\|\bar{K}_\mathrm{i}\right\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\frac{\left|\bar{S}_\mathrm{ij}\right|^2}{\sigma_\mathrm{i}} + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2} &\geq \frac{2\left(\frac{\sigma_p}{\sigma_1} + 1\right)\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sigma_p\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2}{\frac{1}{\sigma_p}\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2} \\
&= \frac{2\left(\frac{\sigma_p^2}{\sigma_1} + \sigma_p\right)\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + 2\sigma_p^2\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_\mathrm{ij}\right|^2 + \sigma_p\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_\mathrm{i}\right\|_F^2} \\
&\geq 2\sigma_p.
\end{aligned}$$

This lower bound is sharp as one can choose $S = 0$ and K with $\left\|\bar{K}_p\right\|_F = 1$ and $\left\|\bar{K}_{\mathrm{i}}\right\|_F = 0$ for $\mathrm{i} < p$.

We have the upper bound as follows.

$$\begin{aligned}
\frac{2\sum_{\mathrm{i,j}=1}^{p}\frac{\sigma_{\mathrm{j}}}{\sigma_{\mathrm{i}}}\left|\bar{S}_{\mathrm{ij}}\right|^2 + 2\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_{\mathrm{ij}}\right|^2 + 2\sum_{\mathrm{i}=1}^{p}\sigma_{\mathrm{i}}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\frac{\left|\bar{S}_{\mathrm{ij}}\right|^2}{\sigma_{\mathrm{i}}} + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2} &\leq \frac{2\left(\frac{\sigma_1}{\sigma_p}+1\right)\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_{\mathrm{ij}}\right|^2 + 2\sigma_1\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2}{\frac{1}{\sigma_1}\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_{\mathrm{ij}}\right|^2 + \sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2} \\
&= \frac{2\left(\frac{\sigma_1^2}{\sigma_p}+\sigma_1\right)\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_{\mathrm{ij}}\right|^2 + 2\sigma_1^2\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2}{\sum_{\mathrm{i,j}=1}^{p}\left|\bar{S}_{\mathrm{ij}}\right|^2 + \sigma_1\sum_{\mathrm{i}=1}^{p}\left\|\bar{K}_{\mathrm{i}}\right\|_F^2} \\
&\leq 2\left(\frac{\sigma_1^2}{\sigma_p}+\sigma_1\right),
\end{aligned}$$

where the last inequality is obtained by investigating the range of the rational function $f(x,y) = \frac{ax+by}{x+dy}$ with $a = 2\left(\frac{\sigma_1^2}{\sigma_p}+\sigma_1\right), b = 2\sigma_1^2$ and $d = \sigma_1$ on $\{(x,y)|x \geq 0, y \geq 0, xy \neq 0\}$.

This upper bound $2\left(\frac{\sigma_1^2}{\sigma_p}+\sigma_1\right)$ may not be the supremum as the inequalities are not sharp. However, we can show that $D^1_{\pi(Y)} \geq 2\sigma_1$. To see this, choose $\bar{S} = 0$ and K with $\left\|\bar{K}_1\right\|_F = 1$ and $\left\|\bar{K}_{\mathrm{i}}\right\|_F = 0$ for $\mathrm{i} > 1$. Then (4.2) reaches the value $2\sigma_1$. Hence the supremum must be at least $2\sigma_1$. So we have

$$2\sigma_1 \leq D^1_{\pi(Y)} \leq 2\left(\frac{\sigma_1^2}{\sigma_p}+\sigma_1\right). \tag{4.3}$$

4.4.2 Proof of Lemma 4.2.6

Proof. We will use the inequality $\|B^*A^*\|_F = \|AB\|_F \leq \|A\|\|B\|_F \leq \|A\|_F\|B\|_F$ for two matrices where $\|A\|$ is the spectral norm. In particular, if X is Hermitian, then we also have $\|AX\|_F = \|XA^*\|_F \leq \|X\|\|A^*\|_F = \|X\|\|A\|_F$.

For the embedded manifold, recall that $\xi_X^s = P_X^s(\xi_X)$ and $\xi_X^p = P_X^p(\xi_X)$ and P_X^t and P_X^p are defined in (2.7), and the bound for the FOT is given by

$$\begin{aligned}
&\frac{\left|g_X\left(P_X^p\left(\nabla f(X)(X^\dagger\zeta_X^p)^* + (\zeta_X^p X^\dagger)^*\nabla f(X)\right), \zeta_X\right)\right|}{g_X(\zeta_X,\zeta_X)} \\
= \;&\frac{\left|\left\langle P_X^p\left(\nabla f(X)\zeta_X^p X^\dagger + X^\dagger\zeta_X^p\nabla f(X)\right), \zeta_X\right\rangle_{\mathbb{C}^{n\times n}}\right|}{\langle\zeta_X,\zeta_X\rangle_{\mathbb{C}^{n\times n}}}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{\left|\left\langle P_X^p\left(\nabla f(X)\zeta_X^p X^\dagger\right), \zeta_X\right\rangle_{\mathbb{C}^{n\times n}}\right|}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} + \frac{\left|\left\langle P_X^p\left(X^\dagger\zeta_X^p \nabla f(X)\right), \zeta_X\right\rangle_{\mathbb{C}^{n\times n}}\right|}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} \\
&\leq 2\frac{\|\nabla f(X)\zeta_X^p X^\dagger\|_F\|\zeta_X\|_F}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} \leq 2\frac{\|\nabla f(X)\|\|\zeta_X^p X^\dagger\|_F\|\zeta_X\|_F}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} \leq 2\frac{\|\nabla f(X)\|\|X^\dagger\|\|\zeta_X^p\|_F\|\zeta_X\|_F}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} \\
&\leq \frac{2\|\nabla f(X)\|\left\|X^\dagger\right\|\|\zeta_X\|_F^2}{\langle \zeta_X, \zeta_X\rangle_{\mathbb{C}^{n\times n}}} = 2\|\nabla f(X)\|\left\|X^\dagger\right\| = \frac{2}{\sigma_p}\|\nabla f(X)\|.
\end{aligned}$$

For the quotient manifold with g^1, the FOT is bounded by

$$\begin{aligned}
\frac{\left|g_Y^1(2\nabla f(YY^*)\bar{\xi}_Y, \bar{\xi}_Y)\right|}{g_Y^1(\bar{\xi}_Y, \bar{\xi}_Y)} &= \frac{\left|\left\langle 2\nabla f(YY^*)\bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}\right|}{\left\langle \bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}} \leq \frac{2\left\|\nabla f(YY^*)\bar{\xi}_Y\right\|_F\left\|\bar{\xi}_Y\right\|_F}{\left\langle \bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}} \\
&\leq \frac{2\|\nabla f(YY^*)\|\left\|\bar{\xi}_Y\right\|_F^2}{\left\langle \bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}} = 2\|\nabla f(YY^*)\|.
\end{aligned}$$

For the quotient manifold with g^2, the FOTs are

$$\begin{aligned}
\text{FOTs} &= \frac{\left\langle \nabla f(YY^*)P_Y^\perp\bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)} + \frac{\left\langle P_Y^\perp\nabla f(YY^*)\bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)} &(4.4)\\
&+\frac{\left\langle Y\bar{\xi}_Y^*\bar{\xi}_Y, 2\nabla f(YY^*)Y(Y^*Y)^{-1}\right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)} &(4.5)\\
&-\frac{\left\langle \bar{\xi}_Y Y^*\bar{\xi}_Y, 2\nabla f(YY^*)Y(Y^*Y)^{-1}\right\rangle_{\mathbb{C}^{n\times p}}}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)}. &(4.6)
\end{aligned}$$

We can estimate each term separately. Notice that the SVD of Y can be given as $Y = U\Sigma^{\frac{1}{2}}V^*$ where V is unitary. Let $\bar{S} = V^*SV$ and $\bar{K} = KV$, and $\bar{K}_{\mathrm{i}}$ be the ith column of $\bar{K}$. For the first summand we have

$$\begin{aligned}
\frac{\left|\left\langle \nabla f(YY^*)P_Y^\perp\bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}\right|}{g_Y^2(\bar{\xi}_Y, \bar{\xi}_Y)} &= \frac{\left|\left\langle \nabla f(YY^*)P_Y^\perp\bar{\xi}_Y, \bar{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}\right|}{\left\langle \bar{\xi}_Y Y^*, \bar{\xi}_Y Y^*\right\rangle_{\mathbb{C}^{n\times n}}} \\
&\leq \frac{\|\nabla f(YY^*)\|\left\|\bar{\xi}_Y\right\|_F^2}{\left\langle \bar{\xi}_Y Y^*, \bar{\xi}_Y Y^*\right\rangle_{\mathbb{C}^{n\times n}}}. \\
&= \frac{\|YS\|_F^2 + \|K\|_F^2}{\|YSY^*\|_F^2 + \|KY^*\|_F^2}\|\nabla f(YY^*)\|
\end{aligned}$$

$$
\begin{aligned}
&\le \left(\frac{\|YS\|_F^2}{\|YSY^*\|_F^2}+\frac{\|K\|_F^2}{\|KY^*\|_F^2}\right)\|\nabla f(YY^*)\| \\
&= \left(\frac{\left\|\sqrt{\Sigma}\bar{S}\right\|_F^2}{\left\|\sqrt{\Sigma}\bar{S}\sqrt{\Sigma}\right\|_F^2}+\frac{\left\|\bar{K}\right\|_F^2}{\left\|\bar{K}\sqrt{\Sigma}\right\|_F^2}\right)\|\nabla f(YY^*)\| \\
&\le \frac{2}{\sigma_p}\|\nabla f(YY^*)\|.
\end{aligned}
$$

Similarly we have the bounds for the second term:

$$
\frac{\left|\left\langle P_Y^{\perp}\nabla f(YY^*)\overline{\xi}_Y,\overline{\xi}_Y\right\rangle_{\mathbb{C}^{n\times p}}\right|}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)}\le\frac{2}{\sigma_p}\|\nabla f(YY^*)\|.
$$

For the third term, with the fact $\|A^*A\|_F=\|A\|_F^2$, we have

$$
\begin{aligned}
\frac{\left|\left\langle Y\overline{\xi}_Y^*\overline{\xi}_Y,2\nabla f(YY^*)Y(Y^*Y)^{-1}\right\rangle_{\mathbb{C}^{n\times p}}\right|}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)} &= \frac{\left|\left\langle Y\overline{\xi}_Y^*\overline{\xi}_YY^*,2\nabla f(YY^*)Y(Y^*Y)^{-2}Y^*\right\rangle_{\mathbb{C}^{n\times n}}\right|}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)} \\
&\le \frac{\left\|Y\overline{\xi}_Y^*\overline{\xi}_YY^*\right\|_F\|2\nabla f(YY^*)Y(Y^*Y)^{-2}Y^*\|_F}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)} \\
&\le \frac{\left\|\overline{\xi}_YY^*\right\|_F^2\|2\nabla f(YY^*)\|\|Y(Y^*Y)^{-2}Y^*\|_F}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)} \\
&= 2\left\|Y(Y^*Y)^{-2}Y^*\right\|_F\|\nabla f(YY^*)\| \\
&\le \frac{2\sqrt{p}}{\sigma_p}\|\nabla f(YY^*)\|.
\end{aligned}
$$

Similarly we can bound the fourth term:

$$
\frac{\left|\left\langle\overline{\xi}_YY^*\overline{\xi}_Y,2\nabla f(YY^*)Y(Y^*Y)^{-1}\right\rangle\right|_{\mathbb{C}^{n\times p}}}{g_Y^2(\overline{\xi}_Y,\overline{\xi}_Y)}\le\frac{2\sqrt{p}}{\sigma_p}\|\nabla f(YY^*)\|.
$$

Thus, for the quotient manifold with g^2 we have

$$
|\mathrm{FOTs}|\le\frac{4(\sqrt{p}+1)}{\sigma_p}\|\nabla f(YY^*)\|.
$$

For the quotient manifold with g^3, recall that $P_Y^\perp = I - P_Y = I - Y(Y^*Y)^{-1}Y^*$, with the property (2.14) and the fact $(I - P_Y)^*Y = 0$, the FOT can be bounded as follows:

$$\begin{aligned}
|\mathrm{FOT}| &= \frac{\left|g_Y^3((I-P_Y)\nabla f(YY^*)(I-P_Y)\overline{\xi}_Y(Y^*Y)^{-1}, \overline{\xi}_Y)\right|}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)} \\
&= \frac{2\left|\left\langle P_Y^\perp \nabla f(YY^*) P_Y^\perp \overline{\xi}_Y, \overline{\xi}_Y \right\rangle_{\mathbb{C}^{n\times p}}\right|}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)} \\
&= \frac{2|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{g_Y^3(\overline{\xi}_Y, \overline{\xi}_Y)} \\
&= \frac{2|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{\left\|Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*\right\|_F^2} \\
&= \frac{2|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{\|2YSY^* + Y_\perp KY^* + YK^*Y_\perp^*\|_F^2} \\
&= \frac{2|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{\|2YSY^*\|_F^2 + \|Y_\perp KY^*\|_F^2 + \|YK^*Y_\perp^*\|_F^2} \\
&= \frac{|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{2\|YSY^*\|_F^2 + \|Y_\perp KY^*\|_F^2} \\
&\leq \frac{|\langle \nabla f(YY^*) Y_\perp K, Y_\perp K\rangle_{\mathbb{C}^{n\times p}}|}{\|Y_\perp KY^*\|_F^2} \\
&\leq \frac{\|Y_\perp K\|_F^2}{\|Y_\perp KY^*\|_F^2} \|\nabla f(YY^*)\| \\
&\leq \frac{1}{\sigma_p} \|\nabla f(YY^*)\|.
\end{aligned}$$

4.5 Numerical Experiments and Interpretations

In this section, we report on the numerical performance of the conjugate gradient methods on three kinds of cost functions of $f(X)$: eigenvalue problem, matrix completion, phase-retrieval, and interferometry. In particular, we implement and compare the following four algorithms:

1. Riemannian CG on the quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^1)$, i.e., Algorithm 2 with metric g^1. This algorithm is equivalent to Burer–Monteiro CG, that is, CG applied directly to (3.1).

2. Riemannian CG on the quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^2)$, i.e., Algorithm 2 with metric g^2. The same metric g^2 was used in [36].

3. Riemannian CG on the quotient manifold $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^3)$, i.e., Algorithm 2 with metric g^3, and also a specific retraction, vector transport and initial step as described in Section 3.5. This special implementation is equivalent to Riemannian CG on embedded manifold, i.e., Algorithm 1.

4. Burer–Monteiro L-BFGS method, that is, using the L-BFGS method directly applied to (3.1). This method was used in [11].

4.5.1 Eigenvalue Problem

For any n-by-n Hermitian PSD matrix A, its top p eigenvalues and associated eigenvectors can be found by solving the following minimization problem:

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) := \frac{1}{2}\|X - A\|_F^2 \\ \text{subject to} & X \in \mathcal{H}_+^{n,p} \end{array},$$

or equivalently

$$\begin{array}{ll} \underset{\pi(Y)}{\text{minimize}} & h(\pi(Y)) := \frac{1}{2}\|YY^* - A\|_F^2 \\ \text{subject to} & \pi(Y) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p \end{array}.$$

It is easy to verify that

$$\nabla f(X) = X - A, \quad \nabla^2 f(X)[\zeta_X] = \zeta_X, \quad \zeta_X \in \mathbb{C}^{n\times n}.$$

In practice we only need A as an operator $A : v \mapsto Av$. We consider a numerical test for a random Hermitian PSD matrix A of size 50 000-by-50 000 with rank 10. We solve the minimization problem above with $p = 15$. Obviously, the minimizer is rank-10 thus rank deficient for $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with $p = 15$. This corresponds to a scenario of finding the eigenvalue decomposition of a low rank Hermitian PSD matrix A with estimated rank at most 15. The results are shown in Figure 4.1. The initial guess is the same random initial matrix for all four algorithms. We see that the simpler Burer–Monteiro approach, including the L-BFGS method and the CG method with metric g^1, is significantly slower.

In the second test of Figure 4.2, the minimizer has rank $r = 15$, and the fixed rank for the manifold is also set to $p = 15$; i.e., there is no rank deficiency. But the condition number of the minimizer A causes a difference in the asymptotic convergence rate for the CG method with metric g^1. In 4.2a, the condition number of A is large and we observe a slower asymptotic convergence rate for the CG method with metric g^1; while in 4.2b, the condition number of A is smaller and the asymptotic convergence rate becomes much faster. This is consistent with Theorem 4.2.1. In the third test of Figure 4.3, we show the ratio term $\frac{\left\|\nabla f(Y_k Y_k^*)\right\|}{(\sigma_p)_k}$ in Assumption 4.3.1 versus the iteration number k. This ratio does not blow up as $\pi(Y_k)$ converges to $\pi(\hat{Y})$.

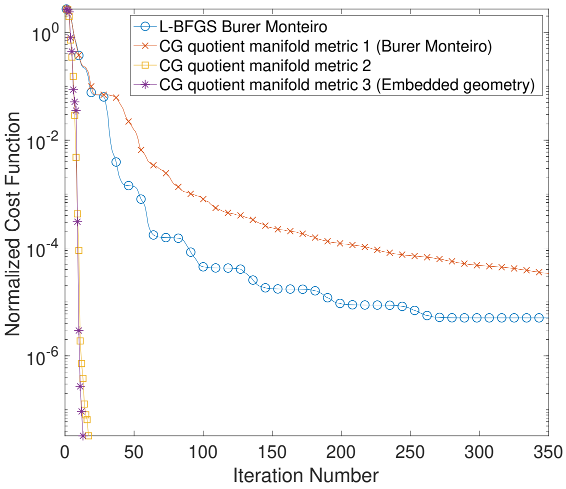
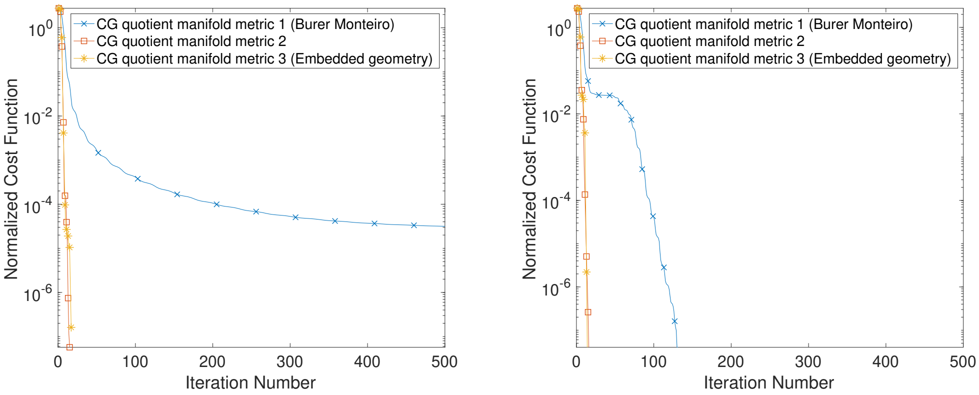


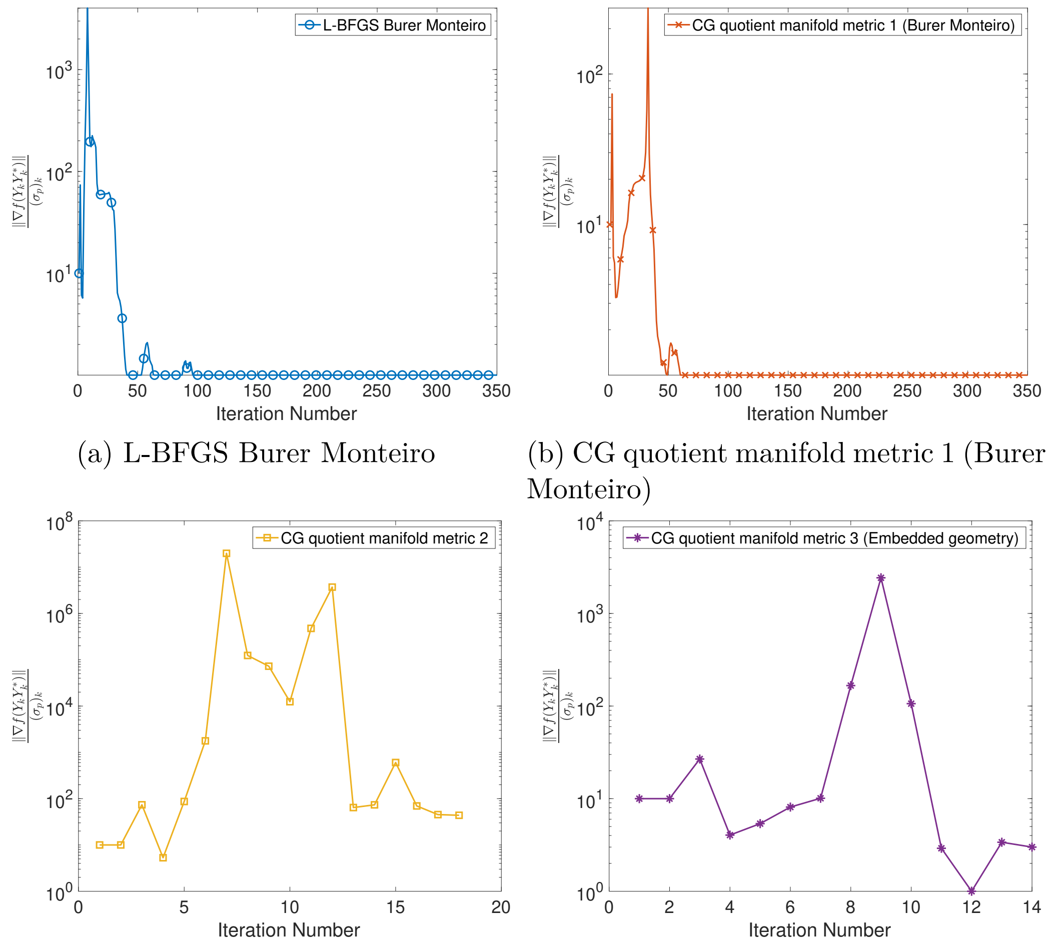
Figure 4.1. Eigenvalue problem of a random 50 000-by-50 000 PSD matrix of rank 10 solved on the rank 15 manifold: a comparison of normalized cost function value $\frac{\left\|Y_k Y_k^* - A\right\|_F}{\|A\|_F}$ decrease versus iteration number k when using L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$.



(a) $\frac{\hat{\sigma}_1}{\hat{\sigma}_p} = 10^6$

(b) $\frac{\hat{\sigma}_1}{\hat{\sigma}_p} = 10^3$

Figure 4.2. Numerical justification of Theorem 4.2.1 for the eigenvalue problem of a random 50 000-by-50 000 PSD matrix of rank 15 on the rank 15 manifold. Effect of condition number of A on the convergence speed of normalized cost function value $\frac{\|Y_k Y_k^* - A\|_F}{\|A\|_F}$ versus iteration number k. (a): when the condition number of A is large, CG with metric g^1 is slower; (b): when its smaller, CG with metric g^1 becomes faster.



(a) L-BFGS Burer Monteiro

(b) CG quotient manifold metric 1 (Burer Monteiro)

(c) CG quotient manifold metric 2

(d) CG quotient manifold metric 3 (Embedded geometry)

Figure 4.3. Numerical justification of Assumption 4.3.1 for the eigenvalue problem of a random 50 000-by-50 000 PSD matrix of rank 10 on the rank 15 manifold, same setup as the numerical test shown in Fig 4.1. Plots show the ratio term $\frac{\left\|\nabla f(Y_k Y_k^*)\right\|}{(\sigma_p)_k}$ in Assumption 4.3.1 versus the iteration number k for L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$.

4.5.2 Matrix Completion Problem

Let Ω be a subset of the complete set $\{1, \cdots, n\} \times \{1, \cdots, n\}$. Then the projection operator onto Ω is a sampling operator defined as

$$P_\Omega : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n} : X_{\mathrm{i,j}} \mapsto \begin{cases} X_{\mathrm{i,j}} & \text{if } (\mathrm{i,j}) \in \Omega, \\ 0 & \text{if } (\mathrm{i,j}) \notin \Omega. \end{cases}$$

The original matrix completion problem has no symmetry or Hermitian constraint. Here, we just consider an artificial Hermitian matrix completion problem for a given $A \in \mathcal{H}_+^{n,p}$:

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) := \frac{1}{2}\|P_\Omega(X - A)\|_F^2 \\ \text{subject to} & X \in \mathcal{H}_+^{n,p} \end{array},$$

or equivalently

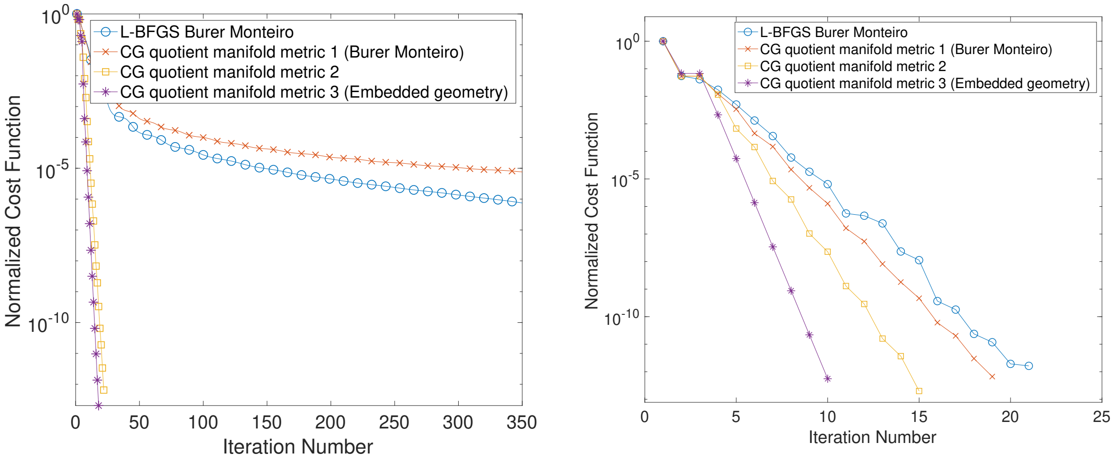
$$\begin{array}{ll} \underset{\pi(Y)}{\text{minimize}} & h(\pi(Y)) := \frac{1}{2}\|P_\Omega(YY^* - A)\|_F^2 \\ \text{subject to} & \pi(Y) \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p \end{array}.$$

Straightforward calculation shows

$$\nabla f(X) = P_\Omega(X - A), \quad \nabla^2 f(X)[\zeta_X] = P_\Omega(\zeta_X), \quad \zeta_X \in \mathbb{C}^{n\times n}.$$

We consider a Hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$ with $n = 10\,000$ and P_Ω a random 90% sampling operator. In the first test of Figure 4.4a, the minimizer has rank $r = 25$, and the fixed rank for the manifold is set to $p = 30$. In the second test of Figure 4.4b, the minimizer has rank $r = 25$, and the fixed rank for the manifold is set to $p = 25$. The initial guess is the same random matrix for all four algorithms. For both cases, we see that the simpler Burer–Monteiro approach, including the L-BFGS method and the CG method with metric g^1, is significantly slower.

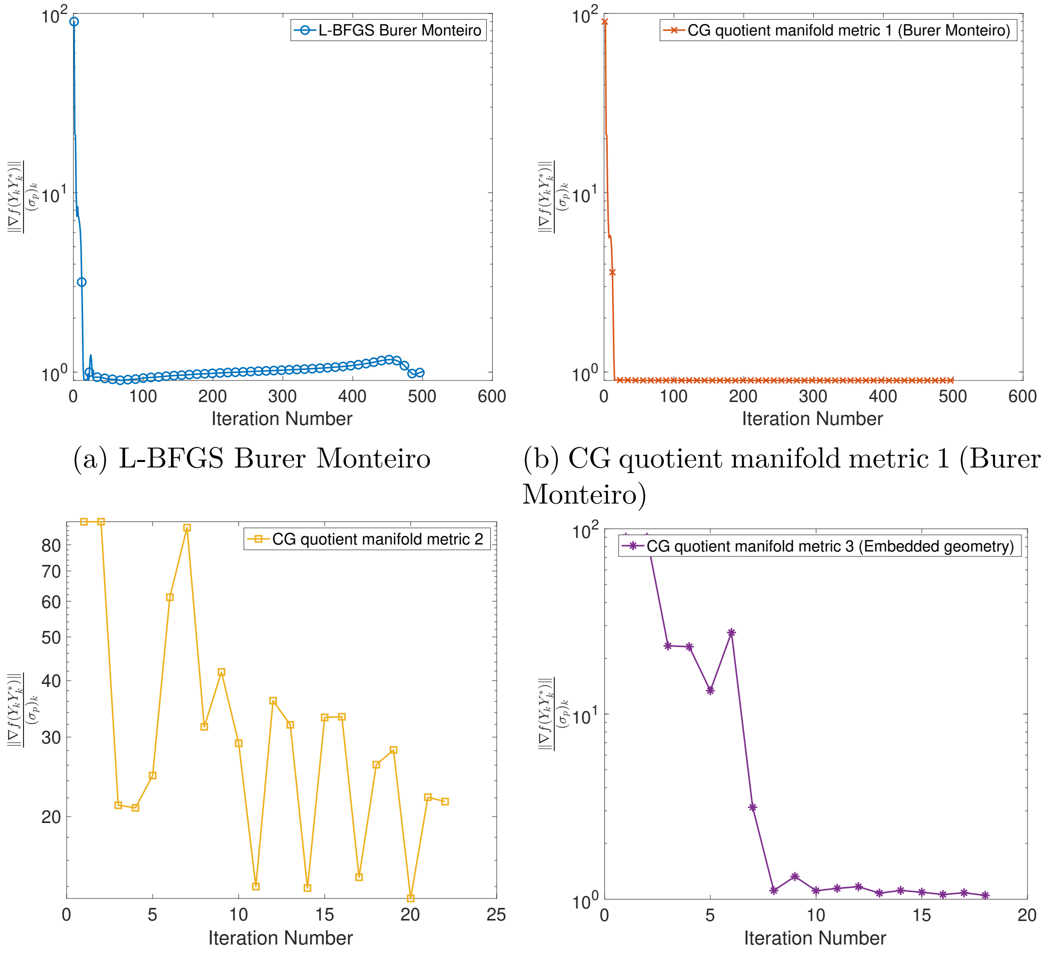
In the third test of Figure 4.5, we show that the ratio term $\frac{\|\nabla f(Y_k Y_k^*)\|}{(\sigma_p)_k}$ in Assumption 4.3.1 versus the iteration number k does not blow up as $\pi(Y_k)$ converges to $\pi(\hat{Y})$.



(a) The algorithms are solved on the rank 30 manifold

(b) The algorithms are solved on the rank 25 manifold

Figure 4.4. Matrix completion of a random 10 000-by-10 000 PSD matrix of rank 25 observed at random 90% entries. A comparison of decrease in normalized cost function value $\frac{\left\|P_\Omega(Y_kY_k^*-A)\right\|_F}{\|P_\Omega(A)\|_F}$ versus iteration number k when using L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$. When the minimizer is rank deficient (the case in (a)), L-BFGS approach and CG method with metric g^1 is significantly slower.



(a) L-BFGS Burer Monteiro

(b) CG quotient manifold metric 1 (Burer Monteiro)

(c) CG quotient manifold metric 2

(d) CG quotient manifold metric 3 (Embedded geometry)

Figure 4.5. Numerical justification of Assumption 4.3.1 for the matrix completion problem of a random 10 000-by-10 000 PSD matrix of rank 25 observed at random 90% entries solved on the rank 30 manifold (same setup as the numerical test shown in Fig 4.4a). Plots show the ratio term $\frac{\left\|\nabla f(Y_k Y_k^*)\right\|}{(\sigma_p)_k}$ in the Assumption 4.3.1 versus the iteration number k for L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$.

4.5.3 The PhaseLift Problem

We now solve the phase retrieval problem as described in [8]: Take an image $x \in \mathbb{C}^{N^2 \times 1}$ and a collection of masks denoted by $\{M_{\rm i}\}_{\rm i=1}^m$ where $N^2 = n$ is the size of the flattened image. Each $M_{\rm i}$ is of the same size as x and the elements in each $M_{\rm i}$ are real or complex numbers with both real and imaginary parts between 0 and 1. We can choose $M_{\rm i}$ to be random numbers or i.i.d. Gaussian. We have m number of observations for each mask $\rm i = 1, \cdots, m$:

$$d^{\rm i} = \mathcal{N}(x) := |(\text{DFT}(\text{Diag}(M_{\rm i}) * x)|^2, \tag{4.7}$$

where $\mathcal{N}$ denotes the nonlinear operator. The squared power is taken element-wisely. $\text{Diag}(M_{\rm i})$ denotes the diagonal matrix whose diagonal is $M_{\rm i}$. DFT denotes the $n \times n$ discrete Fourier transform matrix. Therefore, $d^{\rm i}$ is a vector of size $n \times 1$.

Now we lift x so that $\mathcal{N}$ can be treated as a linear operator. Let $d^{\rm i}_{\rm j}$ denote the jth component of $d^{\rm i}$. Let $z^{{\rm i}*}$ denote $\text{DFT} \cdot \text{Diag}(M_{\rm i})$ and $z_{\rm j}^{{\rm i}*}$ denote the jth row of $\text{DFT} \cdot \text{Diag}(M_{\rm i})$. Then equation (4.7) can be written as

$$\left|\langle z^{\rm i}_{\rm j}, x\rangle\right|^2 = z_{\rm j}^{{\rm i}*} x x^* z^{\rm i}_{\rm j} = d^{\rm i}_{\rm j}, \quad \rm j = 1, \dots n, \quad i = 1, \dots, \mathit{m}.$$

Denoting $X := xx^*$, the nonlinear operator $\mathcal{N}$ can be rewritten as the linear operator

$$\mathcal{A} : \mathbb{C}^{n \times n} \to \mathbb{R}^{mn \times 1}, \quad X \mapsto [tr(z_1^1 z_1^{1*} X), \cdots, tr(z_n^1 z_n^{1*} X), \cdots, tr(z_1^m z_1^{m*} X), \cdots, tr(z_n^m z_n^{m*} X)]^T.$$

Let $Z^{\rm i} := \text{DFT} \cdot \text{Diag}(M_{\rm i}) = \begin{bmatrix} -{z_1^{\rm i}}^* - \\ \cdots \\ -{z_n^{\rm i}}^* - \end{bmatrix}$, then we have alternatively

$$\mathcal{A} : \mathbb{C}^{n \times n} \to \mathbb{R}^{mn \times 1}, \quad X \mapsto [\text{diag}(Z^1 X Z^{1^*}), \cdots, \text{diag}(Z^m X Z^{m*})]^T.$$

Denote $b = [d^1, \cdots, d^m]^T$. Then the cost function can be written as

$$f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|^2$$

The conjugate of operator $\mathcal{A}$, denoted by $\mathcal{A}^*$ is given by

$$A^*(b) = \begin{cases} \displaystyle\sum_{\mathrm{i}=1}^{m}\sum_{\mathrm{j}=1}^{n} b^{\mathrm{i}}_{\mathrm{j}} z^{\mathrm{i}}_{\mathrm{j}} z^{\mathrm{i}*}_{\mathrm{j}} = \sum_{\mathrm{i}=1}^{m} Z^{\mathrm{i}^*} \operatorname{Diag}(b^{\mathrm{i}}) Z^{\mathrm{i}}, & \text{if domain of } \mathcal{A} \text{ is } \mathbb{C}^{n\times n} \\ \operatorname{Re}\left(\displaystyle\sum_{\mathrm{i}=1}^{m}\sum_{\mathrm{j}=1}^{n} b^{\mathrm{i}}_{\mathrm{j}} z^{\mathrm{i}}_{\mathrm{j}} z^{\mathrm{i}*}_{\mathrm{j}}\right) = \operatorname{Re}\left(\displaystyle\sum_{\mathrm{i}=1}^{m} Z^{\mathrm{i}^*} \operatorname{Diag}(b^{\mathrm{i}}) Z^{\mathrm{i}}\right), & \text{if domain of } \mathcal{A} \text{ is } \mathbb{R}^{n\times n}. \end{cases}$$

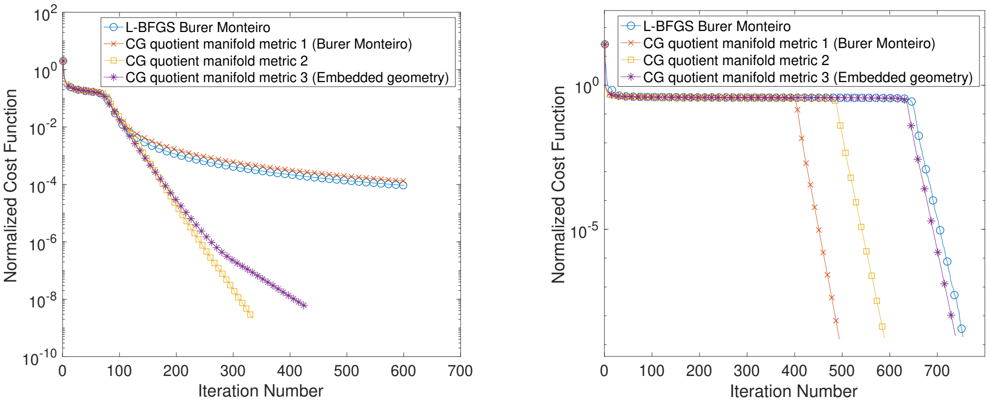
Straightforward calculation shows

$$\nabla f(X) = \mathcal{A}^*(\mathcal{A}(X) - b), \quad \nabla^2 f(X)[\zeta_X] = \mathcal{A}^*(\mathcal{A}(\zeta_X)) \quad \text{for all } \zeta_X \in \mathbb{C}^{n\times n}.$$

For the numerical experiments, we take the phase retrieval problem for a complex gold ball image of size 256 × 256 as in [36]. Thus $n = 256^2 = 65,536$ in (1.1) or (1.2). We consider the operator $\mathcal{A}$ that corresponds to 6 Gaussian random masks. Hence, the size of b is $6n = 393,216$. Remark that problem is easier to solve with more masks.

We first test the algorithms on the rank 3 manifold, and then on the rank 1 manifolds. The results are visible in Figure 4.6. The initial guess is randomly generated. First, we observe that solving the PhaseLift problem on the rank p manifold with $p > 1$ can accelerate the convergence, compared to solving it on the rank 1 manifold. Second, when $p = r = 1$, the asymptotic convergence rates of all algorithms are essentially the same, though the algorithms differ in the length of their convergence "plateaus". When $p = 3 > r = 1$, we can see that the Burer–Monteiro approach has slower asymptotic convergence rates.

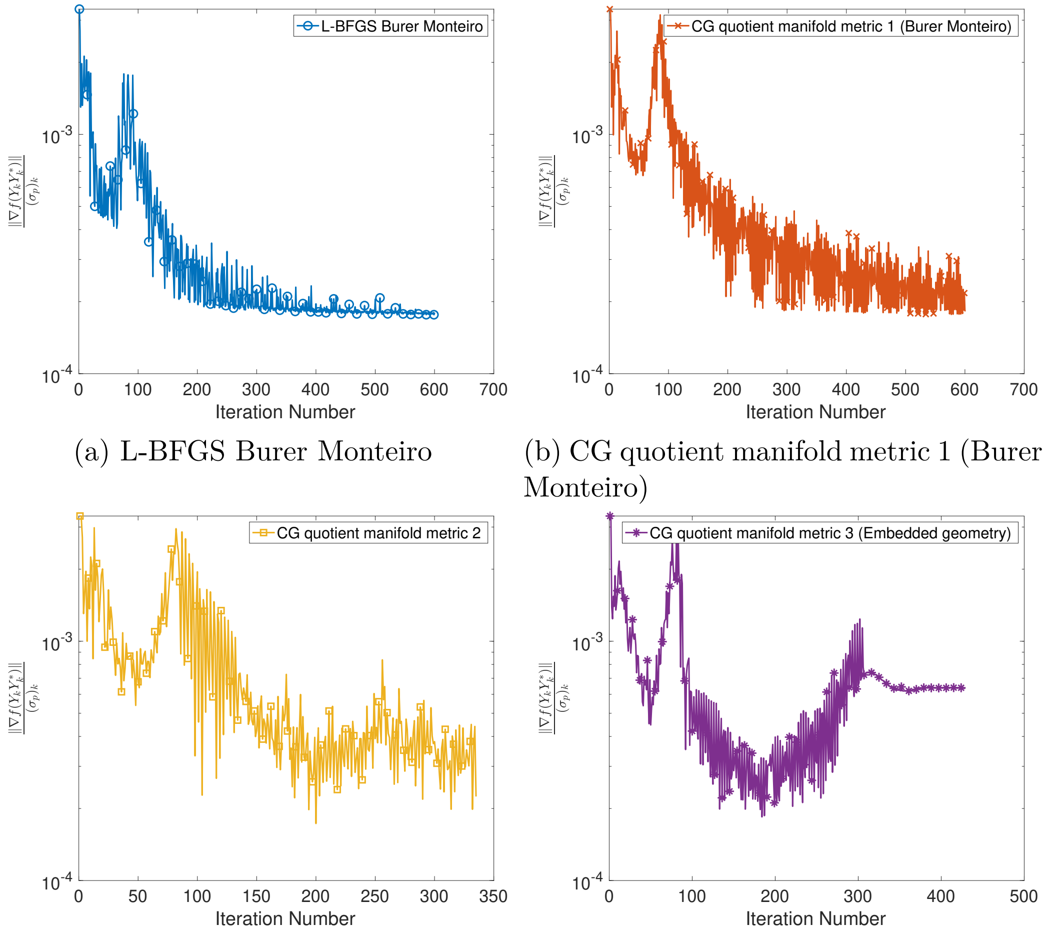
In the second test of Figure 4.7, we show that the ratio term $\frac{\|\nabla f(Y_k Y_k^*)\|}{(\sigma_p)_k}$ in Assumption 4.3.1 versus the iteration number k does not blow up as $\pi(Y_k)$ converges to $\pi(\hat{Y})$.



(a) The algorithms are solved on the rank 3 manifold

(b) The algorithms are solved on the rank 1 manifold

Figure 4.6. Phase retrieval of a 256-by-256 image with 6 Gaussian masks. A comparison of normalized cost function value $\frac{\left\|\mathcal{A}(Y_k Y_k^*)-b\right\|}{\|b\|}$ versus iteration number k when using L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$. When the minimizer is rank deficient (the case in 4.6a), L-BFGS approach and CG method with metric g^1 is significantly slower.



(a) L-BFGS Burer Monteiro

(b) CG quotient manifold metric 1 (Burer Monteiro)

(c) CG quotient manifold metric 2

(d) CG quotient manifold metric 3 (Embedded geometry)

Figure 4.7. Numerical justification of Assumption 4.3.1 for the phase retrieval problem of a 256-by-256 image with 6 Gaussian masks solved on the rank 3 manifold (same setup as the numerical test shown in Fig 4.6a). Plots show the ratio term $\frac{\|\nabla f(Y_k Y_k^*)\|}{(\sigma_p)_k}$ in the Assumption 4.3.1 versus the iteration number k for L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$.

4.5.4 Interferometry Recovery Problem

As last example, we consider solving the interferometry recovery problem described in [11]. Consider solving the linear system $Fx = d$ where $F \in \mathbb{C}_*^{m\times n}$ with $m > n$ and $x \in \mathbb{C}^{n\times 1}$. For the sake of robustness, the interferometry recovery [11] requires solving the lifted problem

$$\begin{array}{ll} \underset{X}{\text{minimize}} & f(X) = \frac{1}{2}\|P_\Omega(FXF^* - dd^*)\|_F^2 \\ \text{subject to} & X \in \mathcal{H}_+^{n,p} \end{array},$$

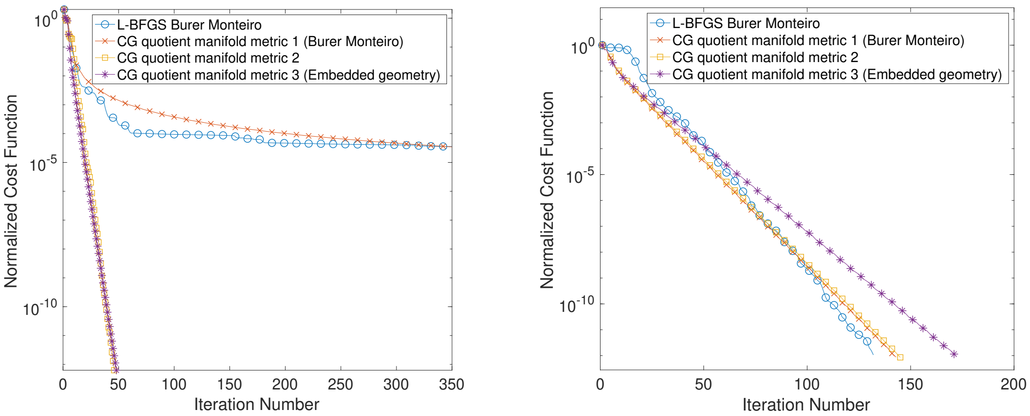
where Ω is a sparse and symmetric sampling index that includes all of the diagonal.

Straightforward calculation again shows

$$\nabla f(X) = F^* P_\Omega(FXF^* - dd^*)F, \quad \nabla^2 f(X)[\zeta_X] = F^* P_\Omega(F\zeta_X F^*)F \quad \text{for all } \zeta_X \in \mathbb{C}^{n\times n}.$$

We solve an interferometry problem with a randomly generated $F \in \mathbb{C}^{10\,000\times 1000}$. Hence $n = 1000$ in (1.1) or (1.2). The sampling operator Ω is also randomly generated, with 1% density. In Figure4.8a, $p = 3$ and $r = 1$ and we can see that the Burer–Monteiro approach has slower asymptotic convergence rates. In Figure4.8b, $p = r = 1$ and we can see now that all algorithms have more or less the same asymptotic convergence rates.

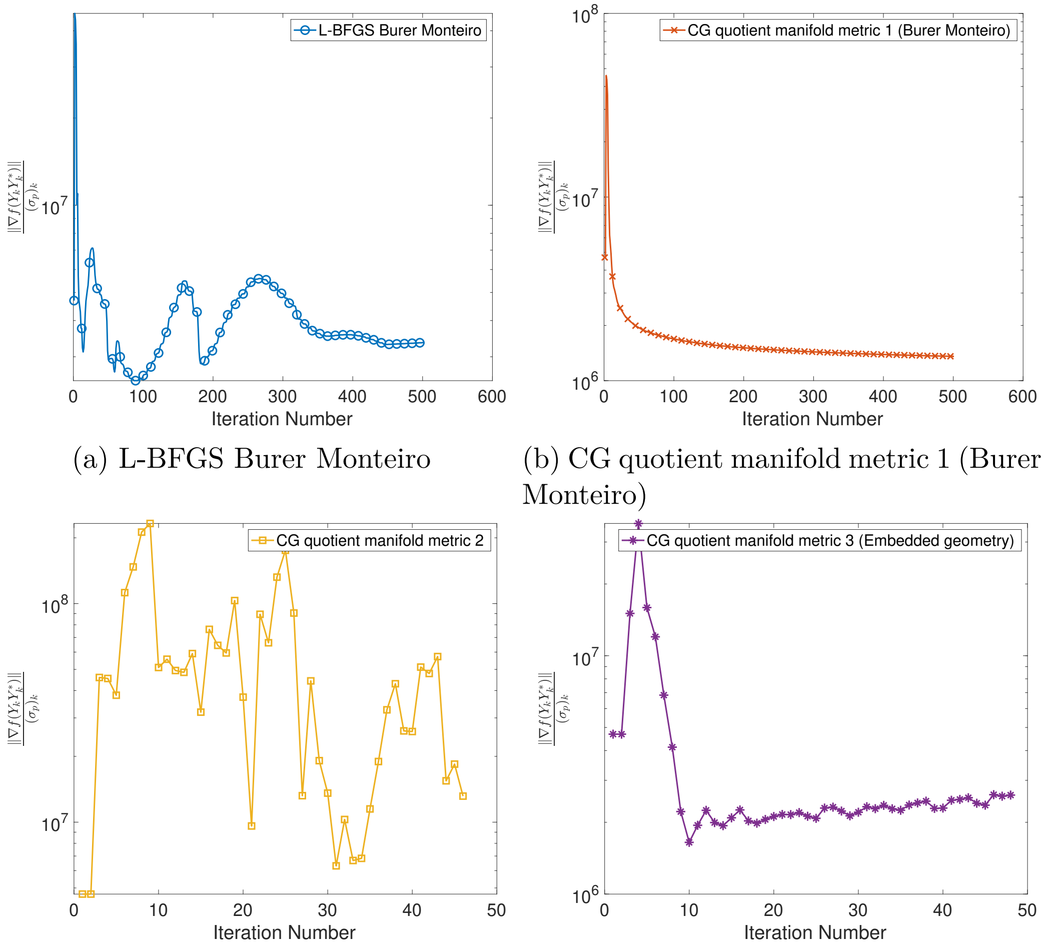
In the second test of Figure 4.9, we show that the ratio term $\frac{\|\nabla f(Y_k Y_k^*)\|}{(\sigma_p)_k}$ in Assumption 4.3.1 versus the iteration number k does not blow up as $\pi(Y_k)$ converges to $\pi(\hat{Y})$.



(a) The algorithms are solved on the rank 3 manifold

(b) The algorithms are solved on the rank 1 manifold

Figure 4.8. Interferometry recovery of a random 10 000-by-1000 F with 1% sampling. A comparison of normalized cost function value $\frac{\left\|P_\Omega(FY_kY_k^*F^*-dd^*)\right\|_F}{\|P_\Omega(dd^*)\|_F}$ versus iteration number k when using L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$. When the minimizer is rank deficient (the case in (a)), L-BFGS approach and CG method with metric g^1 is significantly slower.



(a) L-BFGS Burer Monteiro

(b) CG quotient manifold metric 1 (Burer Monteiro)

(c) CG quotient manifold metric 2

(d) CG quotient manifold metric 3 (Embedded geometry)

Figure 4.9. Numerical justification of Assumption 4.3.1 for the interferometry recovery problem of a random 10 000-by-1000 F with 1% sampling solved on a rank 3 manifold. (same setup as the numerical test shown in Fig 4.8a). Plots show the ratio term $\frac{\|\nabla f(Y_kY_k^*)\|}{(\sigma_p)_k}$ in the Assumption 4.3.1 versus the iteration number k for L-BFGS approach and CG method with metric $g^{\mathrm{i}}, \mathrm{i} = 1, 2, 3$.

4.6 Concluding Remarks

In this chapter, We have analyzed the condition numbers of the Riemannian Hessians on $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^{\mathrm{i}})$ for these metrics g^1, g^3 and another metric g^2 used in the literature. As a noteworthy result, we have shown that when the rank p of the optimization manifold is larger than the rank of the minimizer to the original PSD constrained minimization, the condition number of the Riemannian Hessian on $(\mathbb{C}_*^{n\times p}/\mathcal{O}_p, g^1)$ can be unbounded, which is consistent with the observation that the Burer–Monteiro approach often has a slower asymptotic convergence rate in numerical tests.

5. CONVERGENCE OF ORTHOGONALIZATION-FREE RCG VIA RIEMANNIAN INTERPRETATION

5.1 Introduction

In many applications, it is desired to obtain extreme eigenvalues and eigenvectors of large Hermitian matrices by efficient and compact algorithms. In particular, orthogonalization-free methods are preferred for large-scale problems for finding eigenspaces of extreme eigenvalues without explicitly computing orthogonal vectors in each iteration. For the top p eigenvalues, the simplest orthogonalization-free method is to find the best rank-p approximation to a positive semi-definite Hermitian matrix by algorithms solving the unconstrained Burer-Monteiro formulation. In this chapter, we show that the nonlinear conjugate gradient method for the unconstrained Burer-Monteiro formulation is equivalent to a Riemannian conjugate gradient method on a quotient manifold with the Bures-Wasserstein metric, thus its global convergence to a stationary point can be proven. Numerical tests suggest that it is efficient for computing the largest k eigenvalues for large-scale matrices if the largest k eigenvalues are nearly distributed uniformly.

Given a Hermitian matrix $B \in \mathbb{C}^{n\times n}$, the goal is to find its largest p eigenvalues and the corresponding eigenvectors.

For large enough $\mu > 0$, $A := B + \mu I \in \mathbb{C}^{n\times n}$ is a positive definite Hermitian matrix with the same extreme eigenspaces. Thus we focus only on Hermitian positive definite or semi-definite matrices.

Extreme eigenvalue problems for Hermitian matrices naturally arise in many applications [46–52]. For example, many problems can be cast as a graph, for which the adjacency matrix and the graph Laplacian are real symmetric thus Hermitian [53]. The extreme eigenvalues and eigenvectors of these matrices contain information about the graph and the point cloud data such as diffusion maps [54]. Notice that the discussion in this chapter also applies to the smallest k eigenvalues for a positive definite Hermitian matrix B by considering either $A = \mu I - B$ with large enough μ or $A = B^{-1}$ if an efficient implementation of the linear system solver for $Bx = b$ is available, i.e., the matrix-vector multiplication $B^{-1}b$ can be efficiently implemented.

In the literature, notable convergence results for orthogonalization-free methods include global convergence of perturbed gradient descent for (5.4) in [55] and global convergence of TriOFM in [56].

The same CG algorithm (5.5) was also considered in [57] for real symmetric matrices. Both our algorithm and convergence proof also apply to the Hermitian matrices. We also verify the numerical performance of the discussed algorithms on large matrices of the size millions by millions. In particular, our numerical tests for large matrices are consistent with the observation in [57] that the simple CG method (5.5) is superior for nearly uniformly distributed extreme eigenvalues.

This chapter mainly focuses on the convergence analysis of the simplest orthogonalization-free method (5.5) which is fully scalable in parallel computing. Developing distributed and parallel numerical implementation will be left as future work. In the literature, most numerical solvers for eigenvalue problems rely on orthogonalization to achieve high efficiency in sequential computing. Well-developed algorithms with orthogonalization include [58–61]. To achieve better parallel efficiency for a full eigendecomposition, spectrum slicing can be applied to estimate different eigenpairs in different spectrum regions simultaneously [62–67].

In the rest of this chapter, we will first review the equivalence of the conventional CG method to the Riemannian CG method in Section 5.3, as first shown in Section 3.4. The convergence proof of the Riemannian CG method is then shown in Section 5.4. In Section 5.5, we show that the simple coordinate descent method of minimizing (5.4) is also equivalent to a coordinate Riemannian gradient descent method. Section 5.6 includes numerical tests. Concluding remarks are given in Section 5.7. This chapter is based on [68]

5.2 Problem Formulation and the Riemannian Optimization Viewpoint

The extreme eigenvalue problem can be written as an optimization problem, with many different cost functions to consider. The most well-known one is to minimize the multicolumn Rayleigh quotient

$$\underset{Y\in\mathbb{C}^{n\times p}}{\text{minimize}} \quad F(Y) := \operatorname{tr}\left((Y^*Y)^{-1}Y^*AY\right). \tag{5.1}$$

If assuming the spectrum of Y^*Y is bounded by one and taking the inverse of Y^*Y as the first order approximation of the Neumann series expansion, then as an approximation to multicolumn Rayleigh quotient, a popular method known as the orbital minimization method (OMM) is to minimize the cost function [69]:

$$\underset{Y\in\mathbb{C}^{n\times p}}{\text{minimize}} \quad F(Y) := \operatorname{tr}\left((2I - Y^*Y)Y^*AY\right) \ . \tag{5.2}$$

Another simple formulation is to consider optimization over the noncompact Stiefel manifold $\mathbb{C}_*^{n\times p} = \{Y \in \mathbb{C}^{n\times p}\colon \text{rank(Y)=p}\}$:

$$\underset{Y\in\mathbb{C}_*^{n\times p}}{\text{minimize}} \quad F(Y) := \tfrac{1}{2}\|YY^* - A\|_F^2 \ , \tag{5.3}$$

where $\|\cdot\|_F$ is the matrix Frobenius norm. Various orthogonalization-free algorithms for solving both (5.2) and (5.3) were considered and compared numerically in [57].

Notice that $\mathbb{C}_*^{n\times p}$ is an open set in the Euclidean space $\mathbb{C}^{n\times p}$, thus any line search method $x_{k+1} = x_k + \alpha_k\eta_k$ starting with the iterate $x_k \in \mathbb{C}_*^{n\times p}$ and a small enough step size α_k will give $x_{k+1} \in \mathbb{C}_*^{n\times p}$. Therefore, any such line search algorithm can be regarded as the same algorithm solving an unconstrained problem with a non-degenerate $x_k \in \mathbb{C}_*^{n\times p}$:

$$\underset{x\in\mathbb{C}^{n\times p}}{\text{minimize}} \quad f(x) := \tfrac{1}{2}\|xx^* - A\|_F^2 \ . \tag{5.4}$$

In the literature, the formulation (5.4) is often called the Burer-Monteiro method for Hermitian positive semi-definite (PSD) fixed rank p constraint, i.e., for minimizing $\|X - A\|_F^2$ where X is a Hermitian PSD matrix of rank p.

The nonlinear conjugate gradient method for (5.4) can be written as

$$\begin{cases} x_{k+1} &= x_k + \alpha_k\eta_k, \\ \eta_{k+1} &= -\nabla f(x_k) + \beta_k\eta_k = -2(xx^* - A)x + \beta_k\eta_k, \end{cases} \tag{5.5}$$

where α_k is the step size, β_k is a nonlinear coefficient computed by various formulae, and η_k is the search direction in CG method. In this chapter, we only consider two variants for how to compute β_k: one is the PolakRibiére CG method, and the other one is the Fletcher-Reeves CG method for computing the conjugate direction [70].

A third choice is LOBPCG method first introduced in [58]. A critical step in the LOBPCG method is a Rayleigh-Ritz procedure in which an orthonormal basis is computed to simplify calculations and ensure numerical stability, and it is the only orthogonalization step. LOBPCG without orthogonalization also gives an orthogonalization-free method, which may still work well for many problems in practice, though it might suffer from some instability when the number of eigenpairs to be computed becomes large. Careful base selection strategies [71] [72] can improve its robustness.

The landscape of (5.4) has been well studied in [55, 57, 73, 74] and its local minimizers must also be global minimizers. Theorem 2.1 in [57] implies that, if $\hat{Y} \in \mathbb{C}_*^{n\times p}$ satisfies $\nabla F(\hat{Y}) = 0$ for $F(Y) = \frac{1}{2}\|YY^* - A\|_F^2$, then $\hat{Y} = UO$ where $O \in \mathbb{C}^{p\times p}$ is a unitary matrix, and $U \in \mathbb{C}^{n\times p}$ has orthogonal columns as some eigenvectors of A. Furthermore, any local minimum is a global minimum, i.e., any local minimizer of (5.4) in $\mathbb{C}_*^{n\times p}$ has the form $\hat{Y} = UO$ with columns of U being eigenvectors of a Hermitian PSD matrix A corresponding to its top p eigenvectors.

However, the convergence of CG method (5.5) for (5.4) has never been rigorously justified.

Notice that there is an ambiguity up to unitary matrices in both formulations (5.4) and (5.3), that is $F(YO) = F(Y)$ for any $O \in \mathcal{O}_p$, where $\mathcal{O}_p$ are all $p \times p$ unitary matrices. To this end, mathematically it is proper to consider an equivalence class for each $x \in \mathbb{C}_*^{n\times p}$:

$$[Y] = \{YO : \forall O \in \mathcal{O}_p\},$$

and a quotient set

$$\mathbb{C}_*^{n\times p}/\mathcal{O}_p := \{[Y] : \forall Y \in \mathbb{C}_*^{n\times p}\}.$$

The quotient set with a proper metric becomes a quotient manifold.

Now, we abuse notation by letting x denote the equivalent class $[x]$, and $\overline{x}$ denote one representation of this equivalent class. So we can instead consider the optimization over the quotient manifold:

$$\underset{x\in\mathbb{C}_*^{n\times p}/\mathcal{O}_p}{\text{minimize}} \quad h(x) := F(\overline{x}) = \tfrac{1}{2}\|\overline{x}\overline{x}^* - A\|_F^2 \ . \tag{5.6}$$

Following the recent progress in [75] for Riemannian optimization over Hermitian PSD fixed rank manifolds, we first show that the simple unconstrained Burer-Monteiro CG method (5.5) is equivalent to a Riemannian CG method solving (5.6) over the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with the Bures-Wasserstein metric [32] and proper retraction and vector transport operators. Then with existing Riemannian optimization convergence theory, we can establish the global convergence of the simple algorithm (5.5) to a stationary point of (5.3). We emphasize that the main result of this chapter is the global convergence proof for the classical simple algorithm (5.5), and we do not modify the algorithm (5.5) at all. The Riemannian optimization is used only for proving convergence of (5.5), and (5.5) should not be implemented via much more complicated Riemannian optimization over a quotient manifold.

5.3 The Conjugate Gradient Methods

We first recall the traditional conjugate gradient method for solving (5.4), which is summarized as Algorithm 11. We present the abstract Riemannian conjugate gradient method for solving (5.6) over the quotient manifold as Algorithm 12, with Wolfe conditions

$$h(R_{x_k}(\alpha_k\eta_k)) \leq h(x_k) + c_1\alpha_k g_{x_k}(\operatorname{grad} h(x_k), \eta_k), \tag{5.7}$$

$$\left|g_{R_{x_k}(\alpha_k\eta_k)}(\operatorname{grad} h(R_{x_k}(\alpha_k\eta_k)), \mathrm{D}\, R_{x_k}(\alpha_k\eta_k)[\eta_k])\right| \leq c_2|g_{x_k}(\operatorname{grad} h(x_k), \eta_k)|, \tag{5.8}$$

where $0 < c_1 < c_2 < 1$; and the Riemannian metric on $\mathbb{C}_*^{n\times p}$ is chosen as the Bures-Wasserstein metric g^1 introduced in Section 2.2.1, which is also the canonical Euclidean inner product on $\mathbb{C}^{n\times p}$,

$$g_{\overline{x}}(A, B) := \langle A, B\rangle_{\mathbb{C}^{n\times p}} = \operatorname{Re}(\operatorname{tr}(A^*B)), \quad \forall A, B \in T_{\overline{x}}\mathbb{C}_*^{n\times p} = \mathbb{C}^{n\times p}. \tag{5.9}$$

The abstract Algorithm 12 can be implemented as Algorithm 13, in which each tangent vector is treated as horizontal lift and each iterate is a representative of its equivalence class, and it is independent of the choice of the representative of the equivalent class.

Algorithm 11 (PolakRibiére or Fletcher-Reeves) Conjugate Gradient on $\mathbb{C}^{n\times p}$

```
Require: initial iterate $Y_0 \in \mathbb{C}^{n\times p}$, tolerance $\varepsilon > 0$, initial descent direction as negative
    gradient $\eta_0 = -\nabla F(Y_0) = -2(Y_0 Y_0^* - A)Y_0$
for $k = 0, 1, 2, \dots$ do
    Use backtracking to compute the step size $\alpha_k > 0$ satisfying the strong Wolfe conditions
    Obtain the new iterate by
        $Y_{k+1} = Y_k + \alpha_k \eta_k$
    Compute the gradient
        $\xi_{k+1} := \nabla F(Y_{k+1})$
    Check for convergence
        if $\|\xi_{k+1}\|_F < \varepsilon$, then break
    Compute a conjugate direction by the PolakRibiére method or the Fletcher-Reeves method
        $\eta_{k+1} = -\xi_{k+1} + \beta_{k+1}\eta_k$
        where $\beta_{k+1} = \begin{cases} \max\left(0, \dfrac{\langle \nabla F(Y_{k+1}), \nabla F(Y_{k+1}) - \nabla F(Y_k)\rangle}{\langle \nabla F(Y_k), \nabla F(Y_k)\rangle}\right) & \text{if using PolakRibiére} \\ \dfrac{\langle \nabla F(Y_{k+1}), \nabla F(Y_{k+1})\rangle}{\langle \nabla F(Y_k), \nabla F(Y_k)\rangle} & \text{if using Fletcher-Reeves.} \end{cases}$
end for
```

Algorithm 12 Formal form of the (PolakRibiére or Fletcher-Reeves) Riemannian Conjugate Gradient on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with metric g(c.f. Algorithm 2)

```
Require: initial iterate $x_0 \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p$, tolerance $\varepsilon > 0$, tangent vector $\eta_0 = -\operatorname{grad} h(x_0)$
for $k = 0, 1, 2, \dots$ do
    Compute the step size $\alpha_k > 0$ satisfying the strong Wolfe conditions (5.7) and (5.8)
    Obtain the new iterate by retraction
        $$x_{k+1} = R_{x_k}(\alpha_k \eta_k)$$
    Compute the gradient
        $\xi_{k+1} := \operatorname{grad} h(x_{k+1})$
    Check for convergence
        if $\|\xi_{k+1}\| := \sqrt{g_{x_{k+1}}(\xi_{k+1}, \xi_{k+1})} < \varepsilon$, then break
    Compute a conjugate direction by the PolakRibiére ($\mathrm{PR}_+$) method or the Fletcher-Reeves (FR) method, and vector transport
        $\eta_{k+1} = -\xi_{k+1} + \beta_{k+1} \mathcal{T}_{\alpha_k \eta_k}(\eta_k)$
        $$\text{where } \beta_{k+1} = \begin{cases} \max\left(0, \dfrac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1}), \operatorname{grad} h(x_{k+1}) - \mathcal{T}_{\alpha_k\eta_k}(\xi_k)\right)}{g_{x_k}\left(\operatorname{grad} h(x_k), \operatorname{grad} h(x_k)\right)}\right) & \mathrm{PR}_+ \\ \dfrac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1}), \operatorname{grad} h(x_{k+1})\right)}{g_{x_k}\left(\operatorname{grad} h(x_k), \operatorname{grad} h(x_k)\right)} & \mathrm{FR} \end{cases}$$
end for
```

Algorithm 13 Implementation for Riemannian Conjugate Gradient on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with metric g

```
Require: initial iterate $\overline{x}_0 \in \mathbb{C}_*^{n\times p}$, tolerance $\varepsilon > 0$, initial descent direction as $\overline{\eta}_0 = -\operatorname{grad} F(\overline{x}_0) = -2(\overline{x}_0\overline{x}_0^* - A)\overline{x}_0$
for $k = 0, 1, 2, \dots$ do
    Compute the step size $\alpha_k > 0$ satisfying the strong Wolfe conditions
    Obtain the new iterate by retraction
        $$\overline{x}_{k+1} = \overline{R}_{\overline{x}_k}(\alpha_k\overline{\eta}_k) = \overline{x}_k + \alpha_k\overline{\eta}_k$$
    Compute the horizontal lift of gradient
        $\overline{\xi}_{k+1} := \operatorname{grad} F(\overline{x}_{k+1}) = 2(\overline{x}_{k+1}\overline{x}_{k+1}^* - A)\overline{x}_{k+1}$
    Check for convergence
        if $\left\|\overline{\xi}_{k+1}\right\| := \sqrt{g_{\overline{x}_{k+1}}(\overline{\xi}_{k+1}, \overline{\xi}_{k+1})} < \varepsilon$, then break
    Compute a conjugate direction by PR$_+$ or by FR and vector transport
        $\overline{\eta}_{k+1} = -\overline{\xi}_{k+1} + \beta_{k+1}\overline{\mathcal{T}_{\alpha_k\eta_k}(\eta_k)}_{\overline{x}_{k+1}}$
        $$\text{where } \beta_{k+1} = \begin{cases} \max\left(0, \dfrac{g_{\overline{x}_{k+1}}\left(\operatorname{grad} F(\overline{x}_{k+1}), \operatorname{grad} F(\overline{x}_{k+1}) - \overline{\mathcal{T}_{\alpha_k\eta_k}(\xi_k)}_{\overline{x}_{k+1}}\right)}{g_{\overline{x}_k}\left(\operatorname{grad} F(\overline{x}_k), \operatorname{grad} F(\overline{x}_k)\right)}\right) & \text{PR}_+ \\ \dfrac{g_{\overline{x}_{k+1}}\left(\operatorname{grad} F(\overline{x}_{k+1}), \operatorname{grad} F(\overline{x}_{k+1})\right)}{g_{\overline{x}_k}\left(\operatorname{grad} F(\overline{x}_k), \operatorname{grad} F(\overline{x}_k)\right)} & \text{FR} \end{cases}$$
end for
```

The following results are already shown in Section 3.4 and we simply restate them here without proof for completeness of this chapter.

Lemma 5.3.1. Let η_k be the descent direction generated by Algorithm 12. Then we have

$$\overline{\mathcal{T}_{\alpha_k\eta_k}(\eta_k)}_{\overline{x}_{k+1}} = P^{\mathcal{H}}_{\overline{x}_k+\alpha_k\overline{\eta}_k}(\overline{\eta}_k) = \overline{\eta}_k. \tag{5.10}$$

Theorem 5.3.2. Algorithm 13 is equivalent to Algorithm 11, which is the conjugate gradient method solving (5.4), in the sense that they produce exactly the same iterates if started from the same initial point.

5.4 The Convergence of the Fletcher-Reeves Riemannian Conjugate Gradient Method

In this section, we will prove that the Fletcher-Reeves Riemannian Conjugate Gradient method converges to a stationary point, thus Algorithm 11 also converges by Theorem 5.3.2.

The discussion in this section follows the same arguments as in standard convergence theory, e.g., [76]. The cost function and vector transport considered in this chapter satisfy the conditions for convergence analysis in [76]. Many results in this section are standard convergence results for a line search method, see [70]. For completeness, we include the full proof.

Let $\eta_k \in T_{x_k}\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ be a descent direction. Define the angle θ_k between $-\operatorname{grad} h(x_k)$ and η_k by

$$\cos\theta_k = -\frac{g_{x_k}\left(\operatorname{grad} h(x_k), \eta_k\right)}{\|\operatorname{grad} h(x_k)\|_{x_k}\|\eta_k\|_{x_k}}. \tag{5.11}$$

Let $\mathcal{L} := \{x \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p : 0 \le h(x) \le h(x_0)\}$ and $\pi^{-1}(\mathcal{L}) = \{\overline{x} \in \mathbb{C}_*^{n\times p} : 0 \le F(\overline{x}) \le F(\overline{x}_0)\}$. We can show that $\pi^{-1}(\mathcal{L})$ is bounded.

Lemma 5.4.1. There is a constant C such that $\|\bar{x}\|_F \le C, \quad \forall \bar{x} \in \pi^{-1}(\mathcal{L})$.

Proof. Assume it is not true, then $\forall n \in \mathbb{N}, \exists \bar{x}_n \in \pi^{-1}(\mathcal{L})$ such that $\|\bar{x}_n\|_F \ge n$. Let $y_n = \frac{\bar{x}_n}{\|\bar{x}_n\|_F}$, then $\|y_n\|_F = 1$ and $\bar{x}_n = \|\bar{x}_n\|_F y_n = a_n y_n$ with $a_n \ge n$. Thus $F(\bar{x}_n) = \frac{1}{2}\|a_n^2 y_n y_n^* - A\|_F^2 \to \infty$ since $a_n \to \infty$ and $\|y_n\|_F = 1$. On the other hand, $\bar{x}_n \in \pi^{-1}(\mathcal{L})$ implies that $F(\bar{x}_n)$ should be bounded, which is a contradiction.

Lemma 5.4.2. The Riemannian gradient of F, i.e., $\operatorname{grad} F(\overline{x}) = 2(\overline{x}\overline{x}^* - A)\overline{x}$ is Lipschitz continuous on $\pi^{-1}(\mathcal{L})$. That is, there exists a constant $L > 0$ such that

$$\|\operatorname{grad} F(\overline{y}) - \operatorname{grad} F(\overline{x})\|_F \le L\|\overline{y} - \overline{x}\|_F, \quad \text{for all } \overline{x}, \overline{y} \in \pi^{-1}(\mathcal{L}). \tag{5.12}$$

Proof. It suffices to show that $q : \overline{x} \mapsto \overline{xx}^*\overline{x}$ is Lipschitz continuous on $\pi^{-1}(\mathcal{L})$. Let $\overline{x}, \overline{y} \in \pi^{-1}(\mathcal{L})$. Then $\|\overline{x}\|_F \leq C, \|\overline{y}\|_F \leq C$ by Lemma 5.4.2.

$$\begin{aligned}
\|q(\overline{x}) - q(\overline{y})\|_F &= \|\overline{xx}^*\overline{x} - \overline{yy}^*\overline{y}\|_F = \|\overline{xx}^*\overline{x} - \overline{xx}^*\overline{y} + \overline{xx}^*\overline{y} - \overline{yy}^*\overline{y}\|_F \\
&\leq \|\overline{xx}^*\overline{x} - \overline{xx}^*\overline{y}\|_F + \|\overline{xx}^*\overline{y} - \overline{yy}^*\overline{y}\|_F \\
&= \|\overline{xx}^*\overline{x} - \overline{xx}^*\overline{y}\|_F + \|\overline{xx}^*\overline{y} - \overline{yx}^*\overline{y} + \overline{yx}^*\overline{y} - \overline{yy}^*\overline{y}\|_F \\
&\leq \|\overline{xx}^*\overline{x} - \overline{xx}^*\overline{y}\|_F + \|\overline{xx}^*\overline{y} - \overline{yx}^*\overline{y}\|_F + \|\overline{yx}^*\overline{y} - \overline{yy}^*\overline{y}\|_F \\
&\leq \|\overline{xx}^*\|\|\overline{x} - \overline{y}\|_F + \|\overline{x} - \overline{y}\|_F\|\overline{x}^*\|_F\|\overline{y}\|_F + \|\overline{y}\|_F\|\overline{x}^* - \overline{y}^*\|_F\|\overline{y}\|_F \\
&\leq 3C^2\|\overline{x} - \overline{y}\|_F.
\end{aligned}$$

Theorem 5.4.3 (Zoutendijks theorem on manifold). Let η_k be a descent direction and let α_k satisfy the strong Wolfe conditions (5.7) and (5.8). Then for the cost function h defined in 2.19, the following series converges.

$$\sum_k^\infty \cos^2\theta_k \|\operatorname{grad} h(x_k)\|_{x_k}^2 < \infty.$$

Proof. From the strong Wolfe condition (5.8) we have

$$\begin{aligned}
(c_2 - 1)g_{x_k}(\operatorname{grad} h(x_k), \eta_k) &\leq g_{x_{k+1}}\left((\operatorname{grad} h(R_{x_k}(\alpha_k\eta_k)), \mathrm{D}\, R_{x_k}(\alpha_k\eta_k)[\eta_k]\right) - g_{x_k}\left(\operatorname{grad} h(x_k), \eta_k\right) \\
&= g_{\overline{x}_{k+1}}\left(\operatorname{grad} F(\overline{x}_k + \alpha_k\overline{\eta}_k), P^{\mathcal{H}}_{\overline{x}_k+\alpha_k\overline{\eta}_k}(\overline{\eta}_k)\right) - g_{\overline{x}_k}\left(\operatorname{grad} F(\overline{x}_k), \overline{\eta}_k\right) \\
&= g_{\overline{x}_{k+1}}\left(\operatorname{grad} F(\overline{x}_k + \alpha_k\overline{\eta}_k), \overline{\eta}_k\right) - g_{\overline{x}_k}\left(\operatorname{grad} F(\overline{x}_k), \overline{\eta}_k\right).
\end{aligned}$$

Notice that our Riemannian metric g is simply the inner product on the Euclidean space $\mathbb{C}^{n\times p}$, hence

$$g_{\overline{x}_{k+1}}\left(\operatorname{grad} F(\overline{x}_k + \alpha_k\overline{\eta}_k), \overline{\eta}_k\right) - g_{\overline{x}_k}\left(\operatorname{grad} F(\overline{x}_k), \overline{\eta}_k\right) = \left\langle \operatorname{grad} F(\overline{x}_k + \alpha_k\overline{\eta}_k) - \operatorname{grad} F(\overline{x}_k), \overline{\eta}_k \right\rangle. \tag{5.13}$$

From Lemma 5.4.2 we know

$$\langle \operatorname{grad} F(\overline{x}_k + \alpha_k \overline{\eta}_k) - \operatorname{grad} F(\overline{x}_k), \overline{\eta}_k \rangle \leq \alpha_k L \|\overline{\eta}_k\|_F^2.$$

Hence for any k we have

$$\alpha_k \geq \frac{(c_2 - 1) g_{x_k}(\operatorname{grad} h(x_k), \eta_k)}{L \|\overline{\eta}_k\|_F^2}. \tag{5.14}$$

Now it follows from (5.7) and (5.14) that

$$\begin{aligned} 0 \leq h(x_{k+1}) &\leq h(x_k) + c_1 \alpha_k g_{x_k}(\operatorname{grad} h(x_k), \eta_k) \\ &\leq h(x_k) - \frac{c_1(1-c_2)}{L} \cos^2 \theta_k \|\operatorname{grad} h(x_k)\|_{x_k}^2 \\ &\leq h(x_0) - \frac{c_1(1-c_2)}{L} \sum_{\mathrm{j}=0}^{k} \cos^2 \theta_{\mathrm{j}} \|\operatorname{grad} h(x_{\mathrm{j}})\|_{x_{\mathrm{j}}}^2. \end{aligned}$$

Hence

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\operatorname{grad} h(x_k)\|_{x_k}^2 \leq \frac{L}{c_1(1-c_2)} h(x_0) < \infty. \tag{5.15}$$

Lemma 5.4.4. If using Fletcher-Reeves method in Algorithm 12, then for $0 < c_1 < c_2 < 1/2$, the search direction η_k is a descent direction satisfying

$$-\frac{1}{1-c_2} \leq \frac{g_{x_k}(\operatorname{grad} h(x_k), \eta_k)}{\|\operatorname{grad} h(x_k)\|_{x_k}^2} \leq \frac{2c_2 - 1}{1 - c_2}. \tag{5.16}$$

Proof. We prove it by induction on k.

When $k = 0$, (5.16) holds since

$$\frac{g_{x_0}(\operatorname{grad} h(x_0), \eta_0)}{\|\operatorname{grad} h(x_0)\|_{x_0}^2} = \frac{g_{x_0}(\operatorname{grad} h(x_0), -\operatorname{grad} h(x_0))}{\|\operatorname{grad} h(x_0)\|_{x_0}^2} = -1.$$

Now suppose (5.16) holds for some $k \geq 0$.

Recall that we use differentiated retraction as our vector transport:

$$\mathcal{T}_{\alpha_k \eta_k}(\eta_k) = \mathrm{D}\, R_{x_k}(\alpha_k \eta_k)[\eta_k].$$

And the β_{k+1} in Fletcher-Reeves method is defined as

$$\beta_{k+1} = \frac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1}), \operatorname{grad} h(x_{k+1})\right)}{g_{x_k}\left(\operatorname{grad} h(x_k), \operatorname{grad} h(x_k)\right)}.$$

Hence the middle term in (5.16) for $k+1$ is

$$\begin{aligned}
\frac{g_{x_{k+1}}(\operatorname{grad} h(x_{k+1}), \eta_{k+1})}{\|\operatorname{grad} h(x_{k+1})\|^2_{x_{k+1}}} &= \frac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1}), -\operatorname{grad} h(x_{k+1}) + \beta_{k+1}\mathcal{T}_{\alpha_k\eta_k}(\eta_k)\right)}{\|\operatorname{grad} h(x_{k+1})\|^2_{x_{k+1}}} \\
&= \frac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1}), -\operatorname{grad} h(x_{k+1}) + \beta_{k+1}\mathrm{D}\, R_{x_k}(\alpha_k\eta_k)[\eta_k])\right)}{\|\operatorname{grad} h(x_{k+1})\|^2_{x_{k+1}}} \\
&= -1 + \frac{g_{x_{k+1}}\left(\operatorname{grad} h(x_{k+1})\right), \mathrm{D}\, R_{x_k}(\alpha_k\eta_k)[\eta_k])}{\|\operatorname{grad} h(x_k)\|^2_{x_k}}. \qquad (5.17)
\end{aligned}$$

From the strong Wolfe condition (5.8) we have

$$c_2 g_{x_k}(\operatorname{grad} h(x_k), \eta_k) \le g_{x_{k+1}}(\operatorname{grad} h(x_{k+1}), \mathrm{D}\, R_{x_k}(\alpha_k\eta_k)[\eta_k]) \le -c_2 g_{x_k}(\operatorname{grad} h(x_k), \eta_k). \qquad (5.18)$$

Hence from (5.17) and (5.18) we have

$$-1 + c_2\frac{g_{x_k}(\operatorname{grad} h(x_k), \eta_k)}{\|\operatorname{grad} h(x_k)\|^2_{x_k}} \le \frac{g_{x_{k+1}}(\operatorname{grad} h(x_{k+1}), \eta_{k+1})}{\|\operatorname{grad} h(x_{k+1})\|^2_{x_{k+1}}} \le -1 - c_2\frac{g_{x_k}(\operatorname{grad} h(x_k), \eta_k)}{\|\operatorname{grad} h(x_k)\|^2_{x_k}}.$$

And the result (5.16) follows from the induction hypothesis.

Theorem 5.4.5. For cost function h in (2.19), the Algorithm 12 with Fletcher-Reeves method generates iterates x_k such that

$$\liminf_{k\to\infty} \|\operatorname{grad} h(x_k)\|_{x_k} = 0. \qquad (5.19)$$

Proof. If $\operatorname{grad} h(x_k) = 0$ for some $k = k_0$. Then $\operatorname{grad} h(x_k) = 0$ for all $k \ge k_0$.

So we consider $\operatorname{grad} h(x_k) \ne 0$ for all k. We shall prove (5.19) by contradiction. Suppose (5.19) does not hold. Then there exists a constant $c > 0$ such that

$$\|\operatorname{grad} h(x_k)\|_{x_k} \ge c > 0, \quad \forall k \ge 0. \qquad (5.20)$$

From (5.11) and (5.16) we have

$$\cos\theta_k \geq \frac{1-2c_2}{1-c_2}\frac{\|\operatorname{grad} h(x_k)\|_{x_k}}{\|\eta_k\|_{x_k}}. \tag{5.21}$$

It follows by Theorem 5.4.3 that the following series converges.

$$\sum_{k=0}^{\infty} \frac{\|\operatorname{grad} h(x_k)\|_{x_k}^4}{\|\eta_k\|_{x_k}^2} < \infty. \tag{5.22}$$

For $k \geq 1$, the strong Wolfe condition (5.8) and (5.16) gives rise to

$$\left|g_{x_k}\left(\operatorname{grad} h(x_k), \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right)\right| \;\leq\; -c_2 g_{x_{k-1}}\left(\operatorname{grad} h(x_{k-1}), \eta_{k-1}\right) \leq \frac{c_2}{1-c_2}\|\operatorname{grad} h(x_{k-1})\|_{x_{k-1}}^2.$$

Hence we have the following recurrence equation for $\|\eta_k\|_{x_k}^2$.

$$\begin{aligned}
\|\eta_k\|_{x_k}^2 &= \left\|-\operatorname{grad} h(x_k) + \beta_k \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2 \\
&\leq \|\operatorname{grad} h(x_k)\|_{x_k}^2 + 2\beta_k\left|g_{x_k}\left(\operatorname{grad} h(x_k), \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right)\right| + \beta_k^2\left\|\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2 \\
&\leq \|\operatorname{grad} h(x_k)\|_{x_k}^2 + \frac{2c_2}{1-c_2}\beta_k\|\operatorname{grad} h(x_{k-1})\|_{x_{k-1}}^2 + \beta_k^2\left\|\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2 \\
&= \|\operatorname{grad} h(x_k)\|_{x_k}^2 + \frac{2c_2}{1-c_2}\|\operatorname{grad} h(x_k)\|_{x_k}^2 + \beta_k^2\left\|\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2 \\
&= \frac{1+c_2}{1-c_2}\|\operatorname{grad} h(x_k)\|_{x_k}^2 + \beta_k^2\left\|\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2.
\end{aligned} \tag{5.23}$$

Recall that we use differentiated retraction as our vector transport:

$$\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1}) = \mathrm{D}\, R_{x_{k-1}}(\alpha_{k-1}\eta_{k-1})[\eta_{k-1}] = \mathrm{D}\,\pi(\overline{x}_{k-1} + \alpha_{k-1}\overline{\eta}_{k-1})\left[P^{\mathcal{H}}_{\overline{x}_{k-1}+\alpha_{k-1}\overline{\eta}_{k-1}}(\overline{\eta}_{k-1})\right].$$

Hence

$$\begin{aligned}
\left\|\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right\|_{x_k}^2 &= g_{x_k}\left(\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1}), \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})\right) \\
&= g_{\overline{x}_k}\left(\overline{\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})}_{\overline{x}_k}, \overline{\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1})}_{\overline{x}_k}\right) \\
&= g_{\overline{x}_k}\left(P^{\mathcal{H}}_{\overline{x}_{k-1}+\alpha_{k-1}\overline{\eta}_{k-1}}(\overline{\eta}_{k-1}), P^{\mathcal{H}}_{\overline{x}_{k-1}+\alpha_{k-1}\overline{\eta}_{k-1}}(\overline{\eta}_{k-1})\right) \\
&= g_{\overline{x}_{k-1}}\left(\overline{\eta}_{k-1}, \overline{\eta}_{k-1}\right) = \|\eta_{k-1}\|_{x_{k-1}}^2.
\end{aligned}$$

Hence (5.23) becomes the following recurrence formula for $\|\eta_k\|^2_{x_k}$.

$$\|\eta_k\|^2_{x_k} \le \frac{1+c_2}{1-c_2}\|\mathrm{grad}\, h(x_k)\|^2_{x_k} + \beta_k^2\|\eta_{k-1}\|^2_{x_{k-1}}. \tag{5.24}$$

By recursively using (5.23) and recall the definition of β_k in Fletcher-Reeves method we obtain

$$\begin{aligned}
\|\eta_k\|^2_{x_k} \;&\le\; \frac{1+c_2}{1-c_2}\left(\|\mathrm{grad}\, h(x_k)\|^2_{x_k} + \beta_k^2\|\mathrm{grad}\, h(x_{k-1})\|^2_{x_{k-1}} + \cdots + \beta_k^2\beta_{k-1}^2\ldots\beta_2^2\|\mathrm{grad}\, h(x_1)\|^2_{x_1}\right) \\
&\quad +\beta_k^2\beta_{k-1}^2\ldots\beta_0^0\|\eta_0\|^2_{x_0} \\
&= \frac{1+c_2}{1-c_2}\|\mathrm{grad}\, h(x_k)\|^4_{x_k}\left(\|\mathrm{grad}\, h(x_k)\|^{-2}_{x_k} + \|\mathrm{grad}\, h(x_k)\|^{-2}_{x_{k-1}} + \cdots + \|\mathrm{grad}\, h(x_k)\|^{-2}_{x_1}\right) \\
&\quad +\|\mathrm{grad}\, h(x_k)\|^4_{x_k}\|\mathrm{grad}\, h(x_0)\|^{-2}_{x_0} \\
&< \frac{1+c_2}{1-c_2}\|\mathrm{grad}\, h(x_k)\|^4_{x_k}\sum_{\mathrm{j}=0}^{k}\|\mathrm{grad}\, h(x_{\mathrm{j}})\|^{-2}_{x_{\mathrm{j}}} \le \frac{1+c_2}{1-c_2}\|\mathrm{grad}\, h(x_k)\|^4_{x_k}\frac{k+1}{c^2},
\end{aligned}$$

where we have used the contradiction assumption (5.20) in the last inequality. (5.25) results in the divergence of the following series.

$$\sum_{k=0}^{\infty}\frac{\|\mathrm{grad}\, h(x_k)\|^4_{x_k}}{\|\eta_k\|^2_{x_k}} \ge c^2\frac{1-c_2}{1+c_2}\sum_{k=0}^{\infty}\frac{1}{k+1} = \infty. \tag{5.25}$$

This contradicts to (5.22) and hence we have completed the proof.

In general, it is more difficult to prove the convergence of the Riemannian PR$_+$ CG method. It is possible to extend the convergence proof of PR$_+$ CG method in [77] to Riemannian PR$_+$ CG method, but it is beyond the scope of this chapter.

5.5 Coordinate Riemannian Gradient Descent (CRGD)

The orthogonalization-free methods are preferred for large scale problems. For much larger problems, the coordinate descent method is favored, since the full gradient can be too large to even store. For instance, the coordinate gradient descent method for finding leading eigenvalue in [74] is the coordinate descent method for minimizing (5.4) with rank $p = 1$. In this section, following the same Riemannian manifold notation as in previous sections, we show that the a Riemmanian coordinate descent method is also equivalent to the coordinate descent method for minimizing (5.4) with any rank $p > 0$, which is the generalization of the algorithm in [74].

In [78], a method called the tangent subspace descent method was proposed: this method generalized the block coordinate descent method to manifold settings. Instead of updating the full gradient at each iteration, the tangent direction in each update is a projected vector of the full Riemannian gradient to a subspace of the tangent space by some subspace selection rule P_k. In the specific case of $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ considered in this chapter, this method is written as Algorithm 14 and we denote it as Coordinate Riemannian Gradient Descent (CRGD).

Since the horizontal lift of $\operatorname{grad} h(x_k)$ is a n-by-p matrix, we can simply choose the subspace selection rule by cyclically selecting the N-column block of the n-by-p matrix $\operatorname{grad} F(\overline{x}_k)$. Let M_k denote the mask that evaluates the k-th N-column block of a n-by-p matrix cyclically. That is, if Z is a n-by-p matrix, then

$$M_k(Z) = Z_{kN+1:(k+1)N,:} \tag{5.26}$$

where $Z_{kN+1:(k+1)N,:}$ denotes the N-by-p matrix that takes the $(kN+1)$-th to $(k+1)N$-th columns of Z. And the index that exceeds the matrix range is understood as modulo by the matrix size, namely, cyclically. Then our update to $\overline{x}_k$ is written through the following

$$\overline{x}_{k+1} = \overline{R}_{\overline{x}_k}(\alpha M_k(\operatorname{grad} F(\overline{x}_k))), \tag{5.27}$$

where α is a constant step size.

With the simple retraction as in Section 2.5.1, (5.27) simply reduces to

$$\overline{x}_{k+1} = \overline{x}_{k+1} - \alpha M_k(2(\overline{x}_k\overline{x}_k^* - A)\overline{x}_k). \tag{5.28}$$

Notice that (5.28) with $p = 1$ and $N = 1$ reduces to the coordinate descent method for the leading eigenvalue in [74]. In particular, if $p = 1$ and we set $N = 1$ and P_k in Algorithm 14 to be M_k, defined in (5.26), then Algorithm 14 is equivalent to Algorithm 2 in [74]. So the generalization of the method in [74] to top p eigenvalues can be equivalently written as (5.28) or (5.27), which is a Riemannian coordinate descent method.

To take the advantage of CRGD to solve large-scaled problems, one should implement it through compact implementation. That is, each update should only depend on the block size N and should be independent of the problem size n. In the case of eigenvalue problem, $F(\overline{x}) = \frac{1}{2}\|\overline{x}\overline{x}^* - A\|_F^2$. If we assume that A is a sparse matrix such that we can achieve $M_k(Av)$ in $O(N)$, then we can indeed achieve a compact implementation of CRGD as in Algorithm 15.

Algorithm 14 Coordinate Riemannian gradient descent (CRGD) on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with metric g

```
Require: initial iterate $x_0 \in \mathbb{C}_*^{n\times p}/\mathcal{O}_p$, tolerance $\varepsilon > 0$, tangent vector $\xi_0 = -\operatorname{grad} h(x_0)$,
    subspace selection rule $P_k$, $\delta_0 := P_0(\xi_0)$, stepsize $\alpha > 0$.
for  $k = 0, 1, 2, \dots$ do
    Obtain the new iterate by retraction
            $x_{k+1} = R_{x_k}(\alpha\delta_k)$
    Compute the projection of $\xi_{k+1} := -\operatorname{grad} h(x_{k+1})$ to a subspace of $T_{x_{k+1}}\mathbb{C}^{n\times p}/\mathcal{O}_p$
        $\delta_{k+1} := P_{k+1}(\xi_{k+1})$
    Check for convergence
        if $\|\delta_{k+1}\| := \sqrt{g_{x_{k+1}}(\delta_{k+1}, \delta_{k+1})} < \varepsilon$, then break
end for
```

5.6 Numerical Experiments

The numerical performance of the simple CG methods (5.5) has been well studied in the literature, e.g., see [57] for a comparison with other orthogonalization-free methods. In general, the performance of (5.5) for solving (5.4) depends on the spectrum of the matrix A. For completeness, in this section we verify the numerical performance of the simple CG methods (5.5) on large matrices A.

Algorithm 15 Compact implementation for cyclic coordinate Riemannian gradient descent on the quotient manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ with metric g

```
Require: initial iterate $\overline{x}_0 \in \mathbb{C}_*^{n\times p}$, $\overline{\eta}_0 = -\operatorname{grad} F(\overline{x}_0) \in \mathbb{C}^{n\times p}$, first $N$ columns of $\overline{\eta}_0$: $\overline{\delta}_0 = \mathcal{M}_0(\overline{\eta}_0)$, $a_0 = x_0^* x_0$, $b_0 = \delta_0^* x_0$, $c_0 = \delta_0^* \delta_0$, stepsize $\alpha > 0$, $s_0 = a_0 + \alpha b_0 + \alpha b_0^* + \alpha^2 c_0$, tolerance $\varepsilon > 0$.
for $k = 0, 1, 2, \dots$ do
    Obtain the new iterate by retraction
        $$\overline{x}_{k+1} = \overline{R}_{\overline{x}_k}(\alpha \overline{\delta}_k) = \overline{x}_k + \alpha \overline{\delta}_k$$
    Cyclically compute the next $N$ columns of $\overline{\eta}_{k+1} = -\operatorname{grad} F(\overline{x}_{k+1})$
        $\overline{\delta}_{k+1} := -2M_{k+1}(\overline{x}_k s_k) - 2\alpha M_{k+1}(\overline{\delta}_k s_k) + 2M_{k+1}(A\overline{x}_k) + 2\alpha M_{k+1}(A\overline{\delta}_k)$
    Check for convergence
        if $\left\|\overline{\delta}_{k+1}\right\| := \sqrt{g_{\overline{x}_{k+1}}(\overline{\delta}_{k+1}, \overline{\delta}_{k+1})} < \varepsilon$, then break
    Compute and update $a_{k+1}, b_{k+1}, c_{k+1}$
        $$a_{k+1} = a_k + \alpha \overline{x}_k^* \overline{\delta}_k + \alpha \overline{\delta}_k^* x_k + \alpha^2 \overline{\delta}_k^* \overline{\delta}_k$$
        $$b_{k+1} = \overline{\delta}_{k+1}^* \overline{x}_{k+1}$$
        $$c_{k+1} = \overline{\delta}_{k+1}^* \delta_{k+1}$$
    Compute temporary variable $s_{k+1} \in \mathbb{C}^{p\times p}$
        $s_{k+1} = a_{k+1} + \alpha b_{k+1} + \alpha b_{k+1}^* + \alpha^2 c_{k+1}$
end for
```

5.6.1 Real Symmetric PSD Matrices

We consider two types of matrices A. The first type is a 2D Laplacian matrix, which has a nearly uniform eigenvalue gap for a few top eigenvalues. Consider the discretization of a 2D Poisson equation with homogeneous Dirichlet boundary conditions on $[0, 1] \times [0, 1]$ using m-by-m interior grid points. Then the matrix representing the Laplacian operator is a 2D Laplacian matrix A of size m^2-by-m^2 given as

$$A = \frac{1}{\Delta x^2} K \otimes I_m + I_m \otimes \frac{1}{\Delta y^2} K, \tag{5.29}$$

where $\Delta x = \Delta y = \frac{1}{m+1}$, and K is a m-by-m tridiagonal matrix.

$$K = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \tag{5.30}$$

The second type is constructed by eigenvalue decomposition $A = V\Lambda V^{-1}$ where eigenvectors V are given by discrete cosine transform. We assign Λ so that the eigenvalues λ_{i} have four types of distribution of eigenvalues, similar to the numerical experiments considered in [57] but with a much larger matrix size:

1. (random) $\lambda_{\mathrm{i}} \sim |\mathcal{N}(0,1)|$, where $\mathcal{N}(0,1)$ is standard normal distribution.

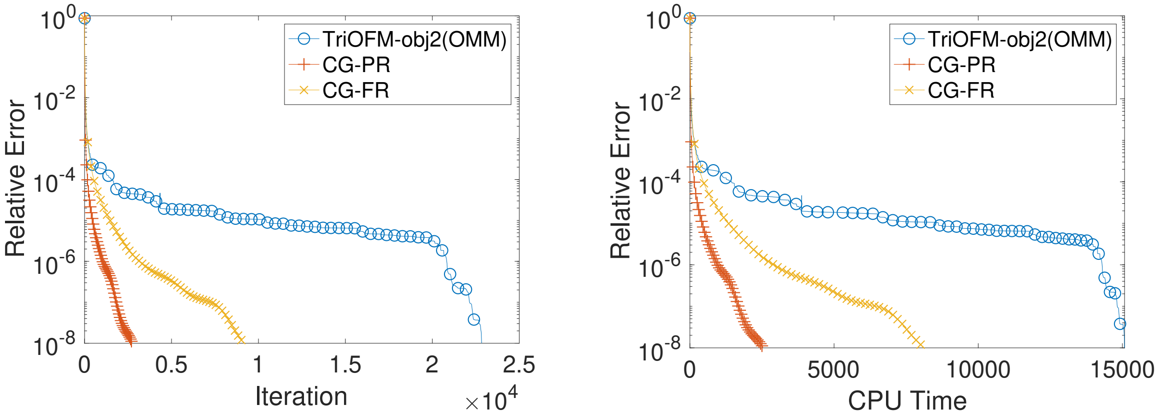
2. (uniform) $\lambda_{\mathrm{i}} = 1 - \frac{\mathrm{i}-1}{n}, \quad 1 \le \mathrm{i} \le r$.

3. (u-shape) $\lambda_1 = \frac{14}{16}, \lambda_2 = \frac{10}{16}, \lambda_3 = \frac{8}{16}, \lambda_4 = \frac{7}{16}, \lambda_5 = \frac{5}{16}, \lambda_{\mathrm{i}} = \frac{1}{16}$.

4. (logarithm) $\lambda_{\mathrm{i}} = \frac{2^{1+\lfloor \log_2 n \rfloor}}{n} \frac{1}{2^{\mathrm{i}}}, \quad 1 \le \mathrm{i} \le r$.

We first compare the simple CG methods (5.5) with the TriOFM method in [56] for a 2D discrete Laplacian matrix, shown in Figure 5.1.

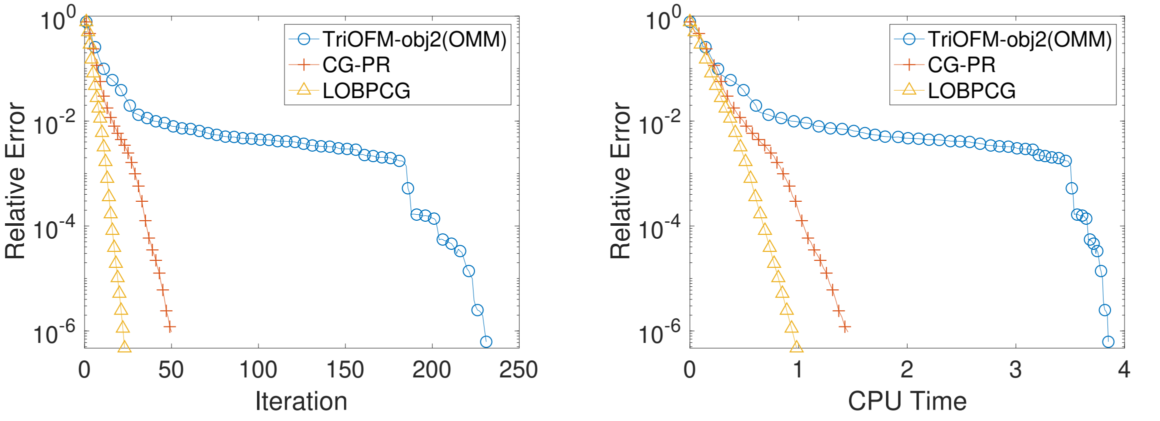
Next, we compare TriOFM, CG and LOBPCG for different distributed eigenvalues. We use Algorithm 1 in [72] as the orthogonalization-free LOBPCG method in numerical tests. The comparison is shown for randomly distributed eigenvalues in Figure 5.2, uniformly distributed eigenvalues in Figure 5.3, U-shape distribution of eigenvalues in Figure 5.4, and log distribution of eigenvalues in Figure 5.5. In all these comparisons, the orthogonalization-free LOBPCG method is the most efficient one. Notice that the simple CG-PR method is much less efficient than the TriOFM method for the log distribution of eigenvalues. However, this slowness is due to the eigenvalue gap between σ_p and σ_{p+1}. In Figure 5.6, the top p eigenvalues with $p = 5$ have a log distribution but the gap between σ_p and σ_{p+1} is enlarged by shifting the top p eigenvalues from the same matrix in Figure 5.5, and we observe that the simple CG-PR method is efficient in this scenario. In other words, the matrix in Figure 5.5 has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and the matrix in Figure 5.6 has eigenvalues $\lambda_1 + C \geq \lambda_2 + C \geq \cdots \geq \lambda_p + C \geq \lambda_{p+1} \geq \cdots \geq \lambda_n$.



(a) Relative error vs iteration

(b) Relative error vs CPU time

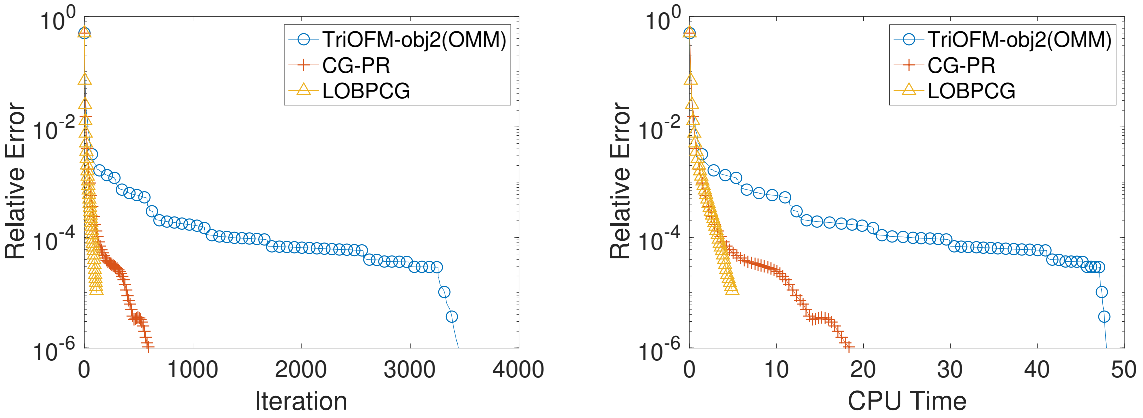
Figure 5.1. Comparison for computing the top-10 eigenvalues of a 2D Laplacian matrix of size $10^6 \times 10^6$.



(a) Relative error vs iteration

(b) Relative error vs CPU time

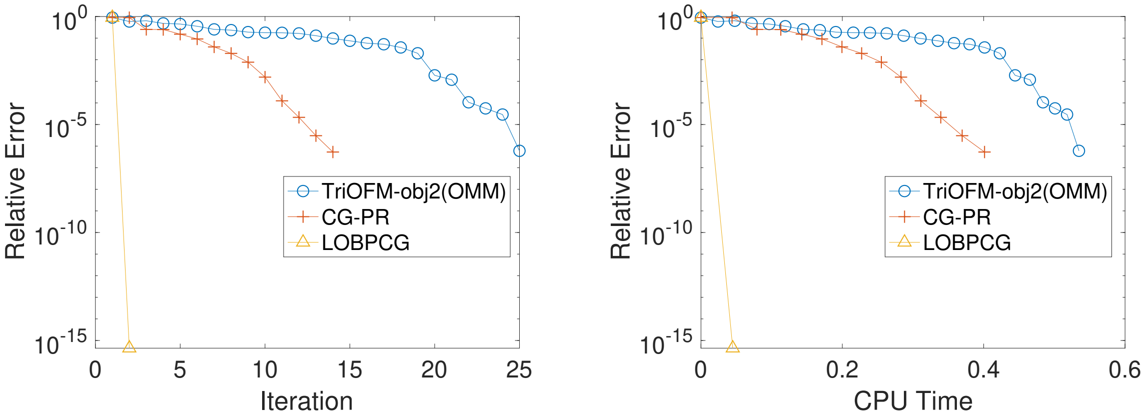
Figure 5.2. Comparison for computing the top-10-eigenvalue problem of a 10^4-by-10^4 matrix with randomly distributed eigenvalues.



(a) Relative error vs iteration

(b) Relative error vs CPU time

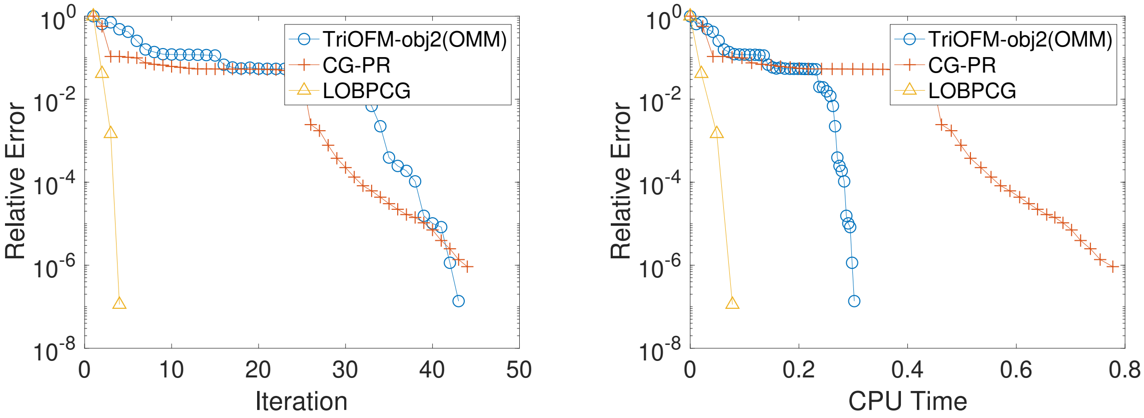
Figure 5.3. Comparison for computing the top-10-eigenvalue problem of a 10^4-by-10^4 matrix with uniformly distributed eigenvalues.



(a) Relative error vs iteration

(b) Relative error vs CPU time

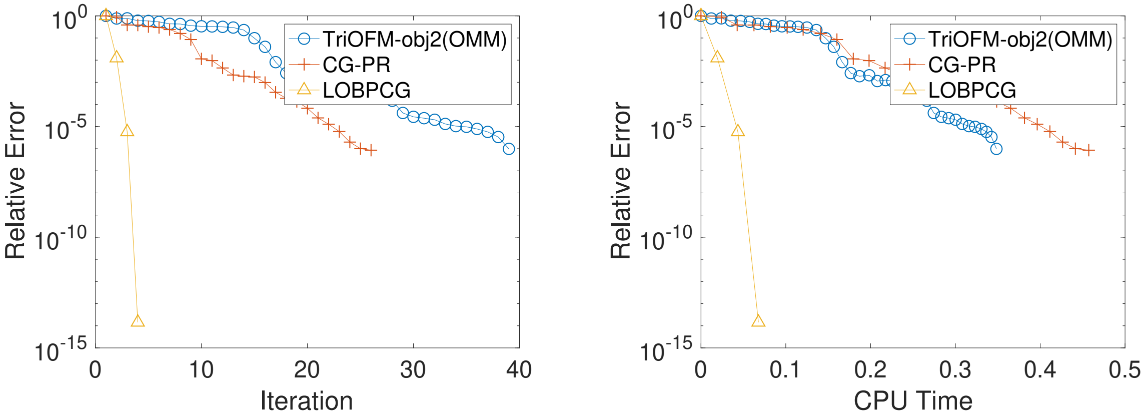
Figure 5.4. Comparison for computing the top-10-eigenvalue problem of a 10^4-by-10^4 matrix with U-shape distributed eigenvalues.



(a) Relative error vs iteration

(b) Relative error vs CPU time

Figure 5.5. Comparison for computing the top-5-eigenvalue problem of a 10^4-by-10^4 matrix with logarithm distributed eigenvalues.



(a) Relative error vs iteration

(b) Relative error vs CPU time

Figure 5.6. Comparison for computing the top-5-eigenvalue problem of a 10^4-by-10^4 matrix with eigenvalues $\lambda_1 + C \geq \lambda_2 + C \geq \cdots \geq \lambda_5 + C \geq \lambda_{5+1} \geq \cdots \geq \lambda_n$, where $C = \lambda_1$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ has a log distribution.

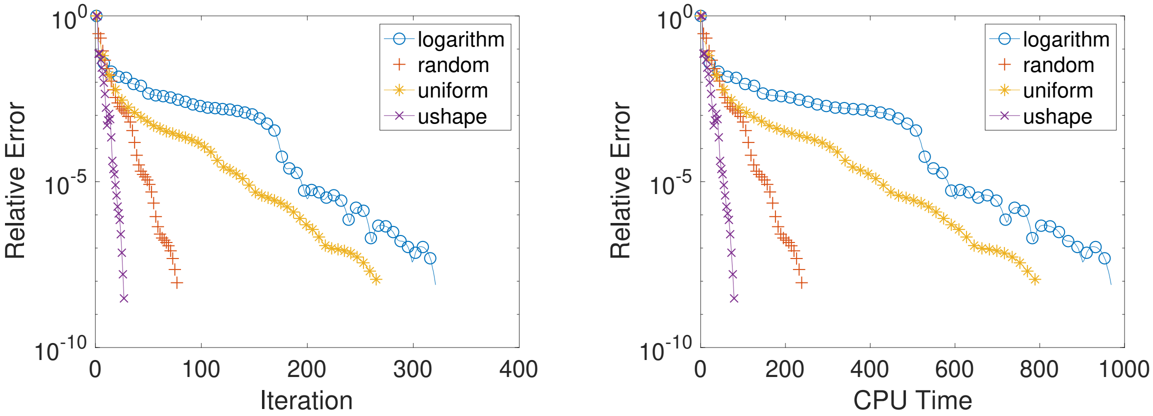
5.6.2 Hermitian PSD Matrices

It is shown in [75] that Algorithm 12 can be used for finding the top eigenvalues of a Hermitian PSD matrix. We test Algorithm 12 on 5.4 for a matrix A with eigenvectors defined by 2D Fast Fourier Transform. Namely, the linear operator of applying A to a 2D array u is defined by

$$Au = \mathrm{i}fft2(\Sigma. * fft2(u)),$$

where $.*$ denotes the entrywise product and Σ is a 2D array consisting of nonnegative eigenvalues of A.

The performance of the CG-PR method is shown in Figure 5.7 for four kinds of eigenvalue distributions in such a Hermitian PSD matrix.



(a) Relative error vs iteration

(b) Relative error vs CPU time

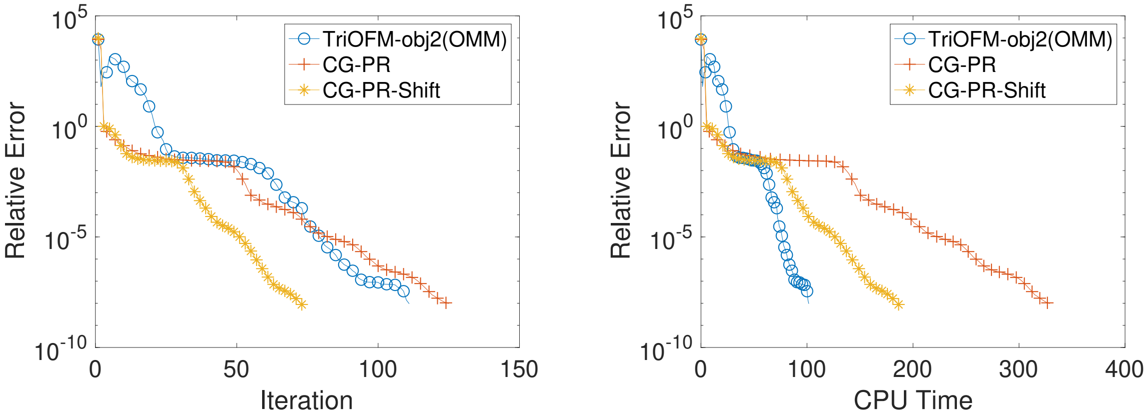
Figure 5.7. The CG-PR method for the top-10-eigenvalue problem with rank-1000 Hermitian matrices of 10^6-by-10^6 with different distributions of eigenvalues.

5.6.3 Smallest Eigenvalues

Inverse 2D Laplacian Matrix

One technique to find the smallest eigenvalues of a given invertible matrix A is through the shift-and-inverse method. That is, to find the largest eigenvalues of $(A+\mu I)^{-1}$, where $\mu > 0$ is a shift constant such that $A+\mu I$ becomes positive definite. We use this method to find the smallest eigenvalues of the 2D Laplacian matrix A as in (5.29).

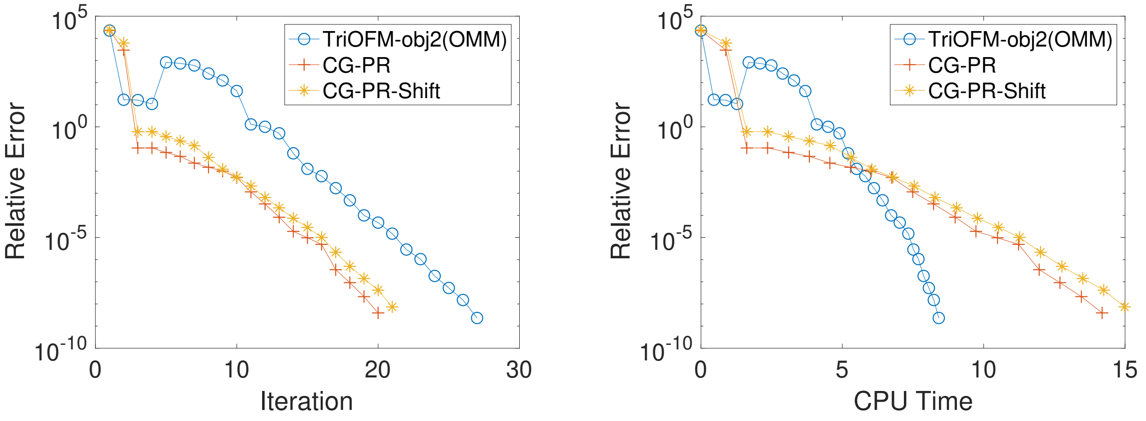
Notice that the top eigenvalues of A^{-1} almost follow a logarithm distribution. Based on our observation, we can choose μ appropriately to make the top eigenvalues of $(A+\mu I)^{-1}$ have a uniform distribution to accelerate the convergence of the CG method. Since we know the true eigenvalues of A, we shift it by choosing μ to be the smallest desired eigenvalue. That is, suppose the smallest r eigenvalues of A are $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_r$. Then we choose $\mu = \sigma_1$. As a result, the top eigenvalues of $(A+\mu I)^{-1}$ would be $\frac{1}{\sigma_1+\sigma_1} \geq \frac{1}{\sigma_2+\sigma_1} \geq \cdots \geq \frac{1}{\sigma_r+\sigma_1}$ that almost follow a uniform distribution. A fast matrix inversion is implemented by using the eigendecomposition of the matrix. The performance is shown in Figure 5.8 and Figure 5.9.



(a) Relative error vs iteration

(b) Relative error vs CPU time

Figure 5.8. The shift-and-inverse method on the smallest-10-eigenvalue problem of a 10^6-by-10^6 2D-Laplacian matrix.



(a) Relative error vs iteration

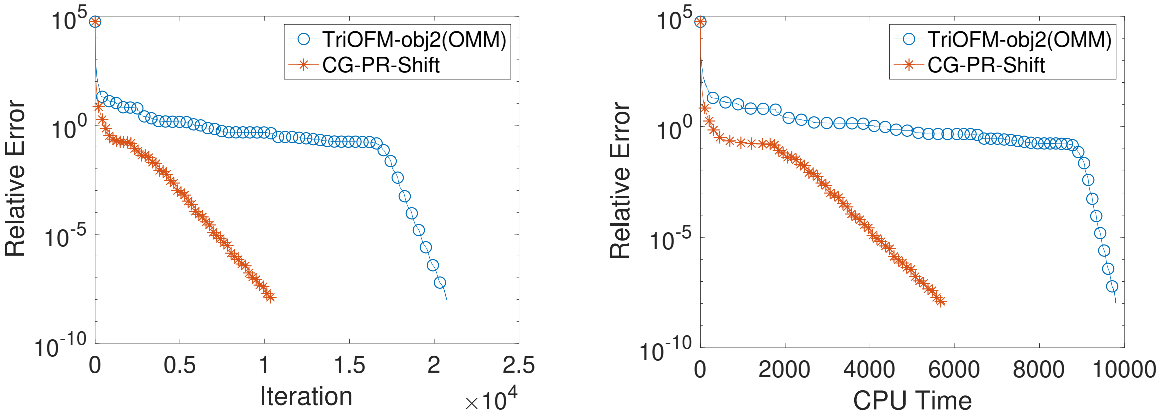
(b) Relative error vs CPU time

Figure 5.9. The shift-and-inverse method on the smallest-3-eigenvalue problem of a 10^6-by-10^6 2D-Laplacian matrix.

Negative 2D Laplacian Matrix

Another way to find the smallest eigenvalues of a given matrix A is through the negative-shift method. That is, to consider finding the largest eigenvalues of $\mu I - A$, where $\mu > 0$ is a shift constant such that $\mu I - A$ is positive semi-definite. We use this method to find the smallest eigenvalues of the 2D Laplacian matrix defined in (5.29).

Notice we need to shift at least the largest eigenvalue of A to ensure that $\mu I - A$ is PSD. And once we find the top eigenvalues of $\mu I - A$ we need to shift back and extract the smallest eigenvalues of A by computing $\mu - (\mu - \sigma)$, where σ's are the smallest eigenvalues of A. Hence when the condition number of A is bad, i.e., if $\mu >> \sigma$, then we might lose a significant number of digits of accuracy for computing $\mu - (\mu - \sigma)$. In our numerical tests, we did not encounter this numerical accuracy issue. The performance is shown in Figure 5.10. Notice that the negative-shift method is much slower than the shift-and-inverse method, because of the different distributions of the largest eigenvalues of $\mu I - A$ and $(A + \mu I)^{-1}$.



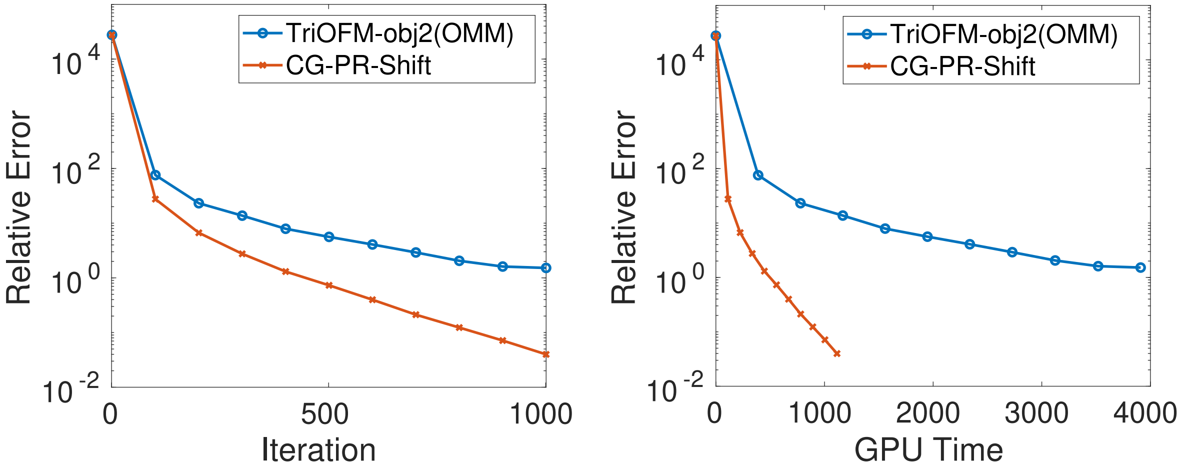
(a) Relative error vs iteration

(b) Relative error vs CPU time

Figure 5.10. The negative-shift method on the smallest-10-eigenvalue problem of a 10^6-by-10^6 2D-Laplacian matrix.

Negative 3D Laplacian matrix

We repeat the same test as in the previous subsection for a larger problem of finding the smallest eigenvalues of a 3D discrete Laplacian on a 500^3 grid, which corresponds to a matrix of size 1.25E8×1.25E8. We implement both the simple CG method (5.5) and TriOFM method on an Nvidia GPU A100 80G.



(a) Relative error vs iteration (b) Relative error vs GPU time

Figure 5.11. The shift-and-inverse method on the smallest-3-eigenvalue problem of a 3D-Laplacian matrix on a 500^3 grid. The matrix size is 1.25E8×1.25E8. Computation was done on Nvidia GPU A100 80G.

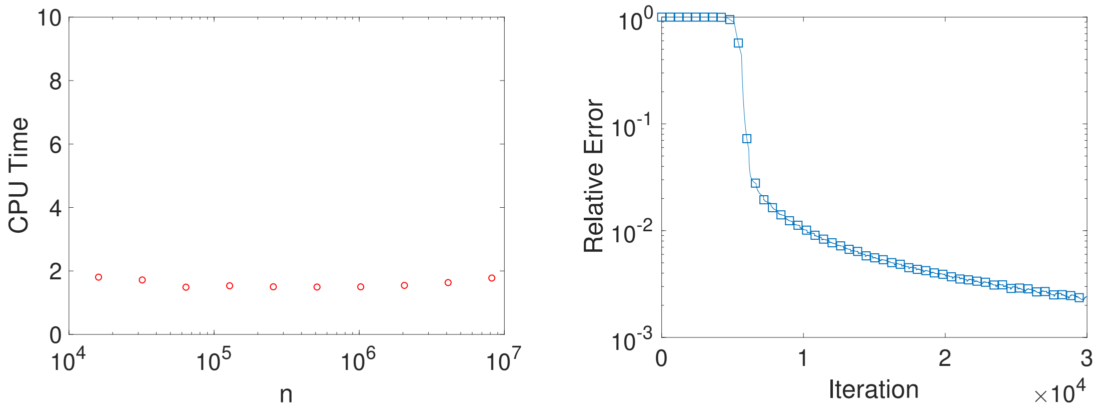
5.6.4 Coordinate Riemannian gradient descent

We consider applying the coordinate Riemannian gradient descent method described in Section 5.5 to a 1D Laplacian matrix of size n-by-n given by $A = \frac{1}{\Delta x^2} K$, where $\Delta x = \frac{1}{n+1}$ and K are the tridiagonal matrix defined in (5.30). This example is only for the demonstration purpose of the coordinate gradient descent method. Choosing this simple A makes it easy for the compact implementation of the matrix-vector multiplication of Au. One can also apply this method to any sparse matrix A as long as one has the compact implementation of $M_k(Au)$ in $O(N)$, where N is a constant independent of the problem size n.

As we can see from Figure 5.12, the CPU time for running the first 3000 iterations is independent of problem size. This demonstrated the $O(1)$ computational complexity of the coordinate Riemannian gradient descent method for leading eigenpairs.

5.7 Concluding Remarks

In this chapter, we have shown the orthogonalization-free method to find leading eigenpairs of a positive semi-definite Hermitian matrix via an unconstrained Burer-Monteiro formulation. For this optimization problem, we have shown the equivalence between the



(a) CPU time of the first 3000 iterations vs problem size $n = 100 * 2^k$ for k goes from 4 to 13. Each iteration cyclically updates $N = 1000$ columns.

(b) Relative error vs iteration. Problem size $n = 100 * 2^9$. Each iteration cyclically updates $N = 100$ columns with constant step size 10^{-10}.

Figure 5.12. Coordinate Riemannian gradient descent for solving the top-10 eigenvalues of a Laplacian matrix.

nonlinear conjugate gradient method and a Riemannian conjugate gradient method on a quotient manifold with the Bures-Wasserstein metric, leading to a new understanding of the global convergence of the nonlinear conjugate gradient method in Burer-Monteiro formulation to a stationary point. We have also shown that the simple coordinate descent method in Burer-Monteiro formulation is equivalent to a coordinate Riemannian gradient descent method. Numerical tests on large scale matrices have verified the numerical performance of the simple conjugate gradient method in Burer-Monteiro formulation for computing leading eigen-pairs, which is consistent with findings in the literature.

6. RIEMANNIAN LANGEVIN MONTE CARLO SCHEMES FOR SAMPLING PSD MATRICES WITH FIXED RANK

6.1 Introduction

In this chapter, we turn our attention from Riemannian optimization to Riemannian sampling.

We will introduce two explicit numerical schemes to sample matrices from the Gibbs distributions on $\mathcal{S}_+^{n,p}$, the manifold of real PSD matrices of size $n \times n$ and rank p. Gibbs distributions originate in statistical physics, while the sampling problem may also be seen as a stochastic variant of the optimization problem. Given an energy function $\mathcal{E} : \mathcal{S}_+^{n,p} \to \mathbb{R}$ and certain Riemannian metrics g on $\mathcal{S}_+^{n,p}$, these schemes rely on an Euler-Maruyama discretization of the Riemannian Langevin equation (RLE) with Brownian motion on the manifold. We present numerical schemes for RLE under two fundamental metrics on $\mathcal{S}_+^{n,p}$: (a) the metric obtained from the embedding of $\mathcal{S}_+^{n,p} \subset \mathbb{R}^{n\times n}$; and (b) the Bures-Wasserstein metric corresponding to quotient geometry. We also provide examples of energy functions with explicit Gibbs distributions that allow numerical validation of these schemes.

This chapter is based on [79]. The main contribution in this chapter is the efficient sampling schemes for ρ_β based on Langevin dynamics. Our approach builds on the geometric theory of optimization; in particular, we extend Riemannian optimization on $\mathcal{S}_+^{n,p}$ [75, 80] to Gibbs sampling as follows. In [80] it was recognized that two commonly used gradient descent schemes over $\mathcal{S}_+^{n,p}$ are time discretizations of *Riemannian* gradient flows, where $\mathcal{S}_+^{n,p}$ is equipped with the two natural Riemannian metrics listed below. We combine this observation with the theory of Brownian motion on Riemannian manifolds to obtain Riemannian Langevin equations and explicit sampling schemes. This sampling problem is related to the optimization problem $\min_{X\in\mathcal{S}_+^{n,p}} \mathcal{E}(X)$ since in the limit $\beta \to \infty$ the Gibbs distribution concentrates at the global minima of $\mathcal{E}(X)$.

The reader unfamiliar with these concepts should note that while the abstract theory serves to guide our work, the schemes presented in this chapter may be implemented without requiring a complete understanding of the underlying theory. Further, while this chapter is focused on the two numerical schemes below, the underlying framework can be used to extend other Riemannian gradient descent schemes to sampling schemes for the Gibbs measure. The new phenomenon that arises is the interplay between Brownian motion and curvature in the Riemannian Langevin equation. This interplay has been studied in depth by two of the authors (TY and GM) and their co-workers in recent papers for geometries used in optimization and physics [81–83].

6.2 Problem Statement

Consider the space of real, symmetric positive semi-definite matrices with size $n \times n$ and rank p, denoted by

$$\mathcal{S}_+^{n,p} = \{X \in \mathbb{R}^{n \times n} | X = X^T, X \succeq 0, \mathrm{rank}(X) = p\}. \tag{6.1}$$

Given an energy $\mathcal{E} : \mathcal{S}_+^{n,p} \to \mathbb{R}$ and a parameter $\beta > 0$ referred to as the inverse temperature, our goal is to sample efficiently from the Gibbs distribution.

$$\rho_\beta(X) = \frac{1}{Z_\beta} \mathrm{e}^{-\beta \mathcal{E}(X)} \rho_{\mathrm{ref}}(X), \quad Z_\beta = \int_{\mathcal{S}_+^{n,p}} \mathrm{e}^{-\beta \mathcal{E}(X')} \rho_{\mathrm{ref}}(X')\, dX'. \tag{6.2}$$

Gibbs measures must be defined with respect to a base measure. In this work, we equip the space $\mathcal{S}_+^{n,p}$ with a Riemannian metric g and choose $\rho_{\mathrm{ref}}(X)dX = \sqrt{\det g(X)} dX$ to be the canonical volume form associated to the metric g. This volume form is expressed in coordinates for the metrics studied in Section 6.5.

6.3 Riemannian Langevin Equations on $\mathcal{S}_+^{n,p}$

In this section , we will show how Langevin equations are defined intrinsically on the Riemannian manifold $(\mathcal{S}_+^{n,p}, g)$. We will state the Itô form of the Riemannian Langevin equation (6.6) for both Riemannian geometries. The main ideas are as follows: (a) the abstract theory of Brownian motion on Riemannian manifolds is used to define the Riemannian Langevin equation in Stratonovich form for the metrics g_E and g_{BW} on $\mathcal{S}_+^{n,p}$; (b) the Itô-Stratonovich conversion rule is used to compute the associated Itô form of these SDEs and it is observed that the Itô-Stratonovich correction term corresponds to mean curvature. This approach yields the SDEs below.

6.3.1 The Classical Euclidean Langevin Equation on $\mathbb{R}^n$

Let us first recall the Langevin equation on $\mathbb{R}^n$. Given a potential or energy function $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ and let W_t denote the standard Wiener process on $\mathbb{R}^n$. The Langevin equation for the potential $\mathcal{E}$ is the Itô differential equation

$$dx_t = -\nabla\mathcal{E}(x_t)\,dt + \sqrt{\frac{2}{\beta}}\,dW_t. \tag{6.3}$$

The Fokker-Planck equation describes the evolution of the probability density of x_t. With $\rho(x,t)\,dx = \mathbb{P}(x_t \in (x, x+dx))$, we have

$$\partial_t\rho = \frac{1}{\beta}\triangle\rho + \nabla\cdot(\rho\nabla\mathcal{E})\,. \tag{6.4}$$

The Gibbs density (with reference density being uniform with respect to Lebesgue measure) is the unique equilibrium of equation (6.4) under natural growth assumptions on the energy $\mathcal{E}$ as $|x| \to \infty$.

The Langevin equation immediately yields a numerical scheme for (approximate) sampling from the Gibbs distribution. Fix a step size $\Delta t > 0$, let $t_k = k\Delta t$, $k = 0, 1, \ldots$, and let x_k denote the numerical approximation to (6.3) at time t_k. The Euler-Maruyama scheme to approximate equation (6.3), also known as Langevin Monte Carlo in the statistics literature, is

$$x_{k+1} = x_k - \Delta t\nabla\mathcal{E}(x_k) + \sqrt{\frac{2\Delta t}{\beta}}\xi_k, \tag{6.5}$$

where $\xi_k = (\xi_k^1, \ldots, \xi_k^n)$ is an i.i.d. sequence of standard Gaussian vectors in $\mathbb{R}^n$. This scheme is explicit. In order to extend it to sampling from (6.2) we must understand how to modify the Langevin equation on the Riemannian manifold $(\mathcal{S}_+^{n,p}, g)$.

First, the term $\nabla\mathcal{E}$ must be replaced by the Riemannian gradient, written as $\operatorname{grad}\mathcal{E}$. The more subtle modification of equation (6.3) concerns the noise. The natural analogy is to replace the Wiener process W_t on $\mathbb{R}^n$ with Brownian motion on the Riemannian manifold $(\mathcal{S}_+^{n,p}, g)$ at inverse temperature β, denoted $\boldsymbol{B}_t^{g,\beta}$. This yields the (formal) Riemannian Langevin equation on $(\mathcal{S}_+^{n,p}, g)$

$$\mathrm{d}\boldsymbol{X}_t = -\mathrm{grad}\,\mathcal{E}(\boldsymbol{X}_t)\mathrm{d}t + \mathrm{d}\boldsymbol{B}_t^{g,\beta}. \tag{6.6}$$

This equation is only formal because stochastic differential equations on manifolds must be defined using the Stratonovich formulation in order to ensure coordinate independence (Itô differentials do not satisfy the chain rule, while Stratonovich differentials do) [84, 85]. On the other hand, Itô differential equations are convenient for analysis as well as simulation. Thus, in formulating the Riemannian Langevin equation, it is necessary to first formulate the appropriate Stratonovich equation and then compute the deterministic Itô–Stratonovich correction. A central observation in our work is that this correction term is due to curvature and is explicitly computable for several Riemannian geometries relevant to optimization [81–83, 86].

6.3.2 The Riemannian Langevin Equation $\mathcal{S}_+^{n,p}$ with the Euclidean Metric

Let $X \in \mathcal{S}_+^{n,p}$ whose compact SVD is $X = U\Lambda U^T$ with $U \in \mathbb{R}^{n\times p}$. Equation (6.6) describes the evolution of a point $\boldsymbol{X}_t \in \mathcal{S}_+^{n,p}$ in abstract terms. We now rewrite it in a simpler equivalent form describing the evolution of the entries of the matrix entries $\{(X_t)_{\mathrm{ij}}\}_{\mathrm{i,j}=1}^n$ representing $\boldsymbol{X}_t$. Let us write $X = U\Lambda U^T$ for the compact SVD of X with the singular values $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ written in decreasing order. We suppress the subscript t in the following equations, though the reader should note that U and Λ depend on X_t.

It can be shown that the law of X_t is determined by the Itô differential equation

$$\mathrm{d}X_t = -\mathrm{grad}\,\mathcal{E}(X_t)\mathrm{d}t + \sqrt{\frac{2}{\beta}}\mathrm{d}W_t^{n,p,X_t} + \frac{1}{\beta}H(X_t)\mathrm{d}t. \tag{6.7}$$

In this equation, the stochastic forcing W_t^{n,p,X_t} is the orthogonal projection of white noise in $\mathbb{R}^{n\times n}$ onto $T_{X_t}\mathcal{S}_+^{n,p}$. Precisely, given W_t^{i} for $1 \leq \mathrm{i} \leq n$ and $W_t^{\mathrm{i,j}}$ for $1 \leq \mathrm{i} < \mathrm{j} \leq n$ independent standard one-dimensional Wiener process, we set

$$\mathrm{d}W_t^{n,p,X_t} = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \mathrm{d}W_t^1 & \cdots & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,p} & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,p+1} & \cdots & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,p} & \cdots & \mathrm{d}W_t^p & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{p,p+1} & \cdots & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{p,n} \\ \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,p+1} & \cdots & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{p,p+1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}\mathrm{d}W_t^{1,n} & \cdots & \frac{1}{\sqrt{2}}\mathrm{d}W_t^{p,n} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} U^T \\ U_\perp^T \end{bmatrix},$$

The term $H(X_t)$ is the mean curvature of the embedding $\mathcal{S}_+^{n,p} \to \mathbb{R}^{n\times n}$. We adopt the convention in geometric analysis: the mean curvature is defined as the trace of the second fundamental form of the embedding. Explicitly, we have

$$H(X_t) = \left(\sum_{\mathrm{i}=1}^{p} \frac{1}{\lambda_{\mathrm{i}}}\right) \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} 0_{p\times p} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & I_{n-p} \end{bmatrix} \begin{bmatrix} U^T \\ U_\perp^T \end{bmatrix}. \tag{6.8}$$

An important role of $H(X_t)$ in equation (6.7) is the following: the stochastic forcing is the naive projection of white noise in the ambient space $\mathbb{R}^{n\times n}$ onto $T_{X_t}\mathcal{S}_+^{n,p}$. Intuitively, when one uses the Euler-Maruyama discretization, the role of this term is to update X_t by taking unbiased random steps in any direction in the tangent space. However, Itô calculus has a subtle interplay with the geometry of the embedding, and in order to keep X_t on the manifold $\mathcal{S}_+^{n,p}$, it is necessary to include the correction term given by the mean curvature.

6.3.3 The Riemannian Langevin Equation for $\mathcal{S}_+^{n,p}$ with the Bures-Wasserstein Metric

The manifold $S_+^{n,p}$ can also be viewed as a quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. Recall that the noncompact Stiefel manifold $\mathbb{R}_*^{n\times p}$ is the total space and the natural projection is

$$\pi : \mathbb{R}_*^{n\times p} \to \mathbb{R}_*^{n\times p}/\mathcal{O}_p.$$

For any $Y \in \mathbb{R}_*^{n\times p}$, the equivalence class containing Y is

$$[Y] = \pi^{-1}(\pi(Y)) = \{YO \mid O \in \mathcal{O}_p\},$$

which is an embedded submanifold of $\mathbb{R}_*^{n\times p}$ (see e.g., [87, Prop. 3.4.4]). The tangent space of $[Y]$ at Y is a subspace of $T_Y\mathbb{R}_*^{n\times p}$ called the *vertical space* at Y, denoted by $\mathcal{V}_Y = \left\{Y\Omega \mid \Omega^T = -\Omega, \Omega \in \mathbb{R}^{p\times p}\right\}$ and $\mathcal{H}_Y$ is the horizontal space w.r.t. the Bures-Wasserstein metric g^1.

And also recall that

$$\begin{aligned} \theta : \mathbb{R}_*^{n\times p} &\to \mathcal{S}_+^{n,p} \\ Y &\mapsto YY^T. \end{aligned}$$

is invariant under the equivalence relation and induces a bijection $\tilde{\theta}$ on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ such that $\theta = \tilde{\theta} \circ \pi$. For any function $\mathcal{E}(X)$ defined on $\mathcal{S}_+^{n,p}$, there is a function F defined on $\mathbb{R}_*^{n\times p}$ that induces $\mathcal{E}$: for any $X = YY^T \in \mathcal{S}_+^{n,p}$, $F(Y) := \mathcal{E} \circ \theta(Y) = \mathcal{E}(YY^T)$. This is summarized in the diagram below:

$$\begin{tikzcd}
\mathbb{R}_*^{n\times p} \arrow[d, "\pi"] \arrow[dr, dashed, "\theta:=\tilde{\theta}\circ\pi"] & & \\
\mathbb{R}_*^{n\times p}/\mathcal{O}_p \arrow[r, leftrightarrow, "\tilde{\theta}"] & \mathcal{S}_+^{n,p} \arrow[r, "\mathcal{E}"] & \mathbb{R}
\end{tikzcd}$$

In particular, $\mathcal{S}_+^{n,p}$ is diffeomorphic to $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ under $\tilde{\theta}$, see [75]. Therefore, the Bures-Wasserstein metric g^1 defined in Chapter 2 on the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ induces a metric on $\mathcal{S}_+^{n,p}$, which we also call the Bures-Wasserstein metric and denote it by g_{BW}.

To understand the Bures-Wasserstein metric g_{BW} on $\mathcal{S}_+^{n,p}$ is via the map $\tilde{\theta}$: for any $A, B \in T_X\mathcal{S}_+^{n,p}$ with $X = YY^T$, there exists $a, b \in \mathcal{H}_Y$ such that $\mathrm{d}\tilde{\theta}(\pi(Y))[a] = A$, $\mathrm{d}\tilde{\theta}(\pi(Y))[b] = B$. Then the Bures-Wasserstein metric on $\mathcal{S}_+^{n,p}$ can be written as

$$g_{BW}(A, B) := g^1_{\pi(Y)}(a, b).$$

The Riemannian Langevin equation is now determined by the geometry of Riemannian submersion. We must obtain an Itô differential equation for Y_t, such that $X_t = Y_tY_t^T$ is a matrix that has the same law as the solution to (6.6) in $(\mathcal{S}_+^{n,p}, g_{BW})$.

In comparison with equation (6.7), we see that the natural choice for white noise driving Y_t is white noise in $\mathbb{R}^{n\times p}$. This is the stochastic differential $\mathrm{d}W_t$, where $W_t = \{W_t^{\mathrm{ij}}\}_{1\leq \mathrm{i}\leq n, 1\leq \mathrm{j}\leq p}$ consists of np independent standard one-dimensional Wiener processes. However, as in equation (6.7) we must include a deterministic correction. This correction corresponds to mean curvature again, but in a more subtle way than (6.7). The equivalence class of Y such that $X = YY^T$ is a group orbit of $\mathcal{O}_p$ embedded within $\mathbb{R}^{n\times p}$. The logarithm of the volume of this group orbit constitutes a natural Boltzmann entropy denoted by $S(Y)$. It can be shown that

$$S(Y) = \frac{1}{2}\sum_{\mathrm{i}=1}^{p}\sum_{\mathrm{j}=\mathrm{i}+1}^{p}\log(\sigma_{\mathrm{i}}^2+\sigma_{\mathrm{j}}^2) \tag{6.9}$$

where $\{\sigma_{\mathrm{i}}\}_{\mathrm{i}=1}^{p}$ are singular values of Y. It is known that $\nabla S(Y)$ is the mean curvature of the group orbit in $\mathbb{R}^{n\times p}$ [88, p.3505].

We then have the following Itô differential equation for Y_t such that $X_t = Y_tY_t^T$ has the same law as the solution to (6.6).

$$\mathrm{d}Y_{\mathrm{ij}} = -\frac{\partial\mathcal{E}(YY^T)}{\partial Y_{\mathrm{ij}}}\mathrm{d}t + \sqrt{\frac{2}{\beta}}\mathrm{d}W_t^{\mathrm{ij}} - \frac{1}{\beta}\frac{\partial S(Y)}{\partial Y_{\mathrm{ij}}}\mathrm{d}t, \qquad 1\leq \mathrm{i}\leq n, 1\leq \mathrm{j}\leq p. \tag{6.10}$$

The correction term $\frac{\partial S(Y)}{\partial Y_{\mathrm{ij}}}$ can be explicitly computed using the following Lemma.

Lemma 6.3.1. If $Y\in\mathbb{R}_*^{n\times p}$ has SVD as $Y = Q\Sigma P^T$ with singular values σ_{i}, then the gradient of the correction term S is given by $\nabla S(Y) = Q\tilde{\Sigma}P^T$ where $\tilde{\Sigma}$ is a diagonal matrix with diagonal entries $\sum_{\mathrm{j}\neq 1}\frac{\sigma_1}{\sigma_1^2+\sigma_{\mathrm{j}}^2}, \sum_{\mathrm{j}\neq 2}\frac{\sigma_2}{\sigma_2^2+\sigma_{\mathrm{j}}^2}, \cdots, \sum_{\mathrm{j}\neq p}\frac{\sigma_p}{\sigma_p^2+\sigma_{\mathrm{j}}^2}$.

6.4 The Riemannian Langevin Monte Carlo Schemes

In this section, we give two simple Riemannian Langevin Monte Carlo sampling schemes corresponding to the two Riemannian Langevin equations (6.7) and (6.10). We only consider convenient discretization and approximation methods, i.e., the Euler-Maruyama type discretization; and we use retraction to approximate the exponential map. In particular, we get the two simple Riemannian Langevin Monte Carlo schemes as follows.

6.4.1 Scheme E for the Embedded Geometry

For approximating the SDE (6.7) on $(S_+^{n,p}, g_E)$, with the retraction operator and Euler-Maruyama method for SDE, we have the following scheme

$$X_{k+1} = \mathrm{P}_{\mathcal{S}_+^{n,p}} \left[X_k - \Delta t \operatorname{grad} \mathcal{E}(X_k) + Q_k \left(\sqrt{\frac{2\Delta t}{\beta}} \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^T & 0 \end{bmatrix} + \frac{\Delta t}{\beta} \sum_{\mathrm{i}=1}^{p} \frac{1}{\lambda_{\mathrm{i}}} \begin{bmatrix} 0 & 0 \\ 0 & I_{n-p} \end{bmatrix} \right) Q_k^T \right], \tag{6.11}$$

which can be written equivalently as

$$X_{k+1} = \mathrm{P}_{S_+^{n,p}} \left(\begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Lambda - \Delta t U^T \nabla \mathcal{E}(X_k) U + \sqrt{\frac{2\Delta t}{\beta}} B_{11} & -\Delta t U^T \nabla \mathcal{E}(X_k) U_\perp + \sqrt{\frac{2\Delta t}{\beta}} B_{12} \\ -\Delta t U_\perp^T \nabla \mathcal{E}(X_k) U + \sqrt{\frac{2\Delta t}{\beta}} B_{12}^T & \frac{\Delta t}{\beta} \sum_{\mathrm{i}=1}^{p} \frac{1}{\lambda_{\mathrm{i}}} I_{n-p} \end{bmatrix} \begin{bmatrix} U^T \\ U_\perp^T \end{bmatrix} \right), \tag{6.12}$$

where $X_k = U \Lambda U^T$ is the compact SVD of $X_k \in S_+^{n,p}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$. The third term in the right hand side is the white noise term in the tangent space $T_{X_k} \mathcal{S}_+^{n,p}$. Entries of $B_{12} \in \mathbb{R}^{p \times (n-p)}$ are i.i.d drawn from $\sqrt{\frac{1}{2}} \mathcal{N}(0,1)$, and $B_{11} \in \mathbb{R}^{p \times p}$ are defined as follows.

$$B_{11} = \begin{bmatrix} \mathcal{N}(0,1) & & & \\ & \ddots & b_{\mathrm{ij}} & \\ & b_{\mathrm{ji}} & \ddots & \\ & & & \mathcal{N}(0,1) \end{bmatrix} \tag{6.13}$$

with $b_{\mathrm{ij}} = b_{\mathrm{ji}} \sim \sqrt{\frac{1}{2}} \mathcal{N}(0,1)$. The implementation details of the scheme (6.11) are given as follows in the Algorithm 16.

Algorithm 16 The Riemannian Langevin Monte Carlo scheme (6.11) for $(\mathcal{S}_+^{n,p}, g_E)$

```
Require: initial iterate $X_1 \in \mathcal{S}_+^{n,p}$; full SVD of $X_1$: $X_1 = Q_1 \Lambda_1 Q_1^T$
for $k = 1, 2, \dots, N$ do
    Compute Riemannian gradient
        $\xi_k := \operatorname{grad} \mathcal{E}(X_k)$                ▷ See Algorithm 3
    Compute noise term
        $B = \sqrt{\frac{2\Delta t}{\beta}} \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^T & 0 \end{bmatrix} + \frac{\Delta t}{\beta} \sum_{i=1}^{p} \frac{1}{\lambda_i} \begin{bmatrix} 0 & 0 \\ 0 & I_{n-p} \end{bmatrix}$
    Obtain the new iterate by retraction
        $X_{k+1} = \mathrm{P}_{\mathcal{S}_+^{n,p}}(X_k - \Delta t \xi_k + Q_k B Q_k^T)$                ▷ See Algorithm 6
end for
```

Remark 6.4.1. The mean curvature correction term is necessary for avoiding rank deficient samples in the following sense. A sampling scheme on $\mathcal{S}_+^{n,p}$ might generate a sample X with a rank numerically close to $p-1$, and the mean curvature correction term in the scheme (6.11) would be huge if $\lambda_p \to 0$, thus it will force iterate X_k to stay away from the boundary of $\mathcal{S}_+^{n,p}$.

Remark 6.4.2. Notice that the complexity of computing SVD of $X+Z$ in Algorithm 6 would be $O(n^3)$ in a naive implementation. For a Riemannian gradient method, if $Z \in T_{X_k}\mathcal{S}_+^{n,p}$, a compact implementation of computing $P_{\mathcal{S}_+^{n,p}}(X+Z)$ in [75] is only $O(np^2) + O(p^3)$, which is no longer possible for the Langevin Monte Carlo scheme (6.11) due to the mean curvature correction term in the normal space. On the other hand, if a Lanczos type algorithm is used for computing the top p eigen-components of $X+Z$, it seems possible to explore the special structure in (6.12) to find a more efficient implementation, but we do not consider a more compact implementation in this thesis.

6.4.2 Scheme BW for the Bures-Wasserstein Metric

With the Euler-Maruyama discretization for SDE (6.10), and the simple retraction and Riemannian gradient of quotient manifold which are given in chapter 2, a simple Riemannian Langevin Monte Carlo scheme for approximating the Riemannian SDE (6.10) on the Riemannian manifold $(S_+^{n,p}, g_{BW})$ can be given as

$$Y_{k+1} = Y_k - \Delta t 2 \nabla \mathcal{E}(Y_k Y_k^T) Y_k + \sqrt{\frac{2\Delta t}{\beta}} B_k + \frac{\Delta t}{\beta} U \left[\sum_{\mathrm{j:j \neq i}} \frac{\sigma_{\mathrm{i}}}{\sigma_{\mathrm{i}}^2 + \sigma_{\mathrm{j}}^2} \right]_{\mathrm{ii}} V^T, \tag{6.14}$$

where B_k is n-by-p matrix with i.i.d. $\mathcal{N}(0,1)$ entries and $Y_k = U\Sigma V^T$ is the compact SVD of Y_k with singular values $\sigma_{\mathrm{i}} > 0$ for $\mathrm{i} = 1, 2, \cdots, p$.

Notice that all operations are performed in the space of size $n \times p$. For finding compact SVD of Y, one can first compute QR decomposition of Y, which costs $O(np^2) + O(p^3)$. Then compute SVD of size $p \times p$, which is $O(p^3)$. So the complexity of this scheme is $O(np^2) + O(p^3)$ for each iteration. For large n and small p, Scheme BW should be cheaper than Scheme E in each iteration, but they generate different samples for different Gibbs distributions which depend on the metric, i.e., Scheme BW cannot replace Scheme E for generating Gibbs distribution defined by embedded geometry.

6.5 Examples with Analytical Formulae

In this section, we provide a few examples with analytical formulae so that they can be used in numerical experiments for testing the two schemes (6.12) and (6.14) on the Gibbs distribution.

For the rest of this section, $X = Q\Lambda Q^T \in \mathcal{S}_+^{n,p}$ denotes the full SVD with descending eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.

6.5.1 Scalar Random Variables as the Testing Random Variable

Let X be a random variable satisfying the Gibbs distribution on $\mathcal{S}_+^{n,p}$ with dimension $N = np - \frac{p(p-1)}{2}$ under either metric g_E or g_{BW}, then X is a matrix-valued random variable, making it difficult to validate our schemes. Therefore, for convenience, we consider a scalar random variable $D = D(X)$ which is a function of $X \in \mathcal{S}_+^{n,p}$, e.g., $D = \|X\|_F$ where $\|\cdot\|_F$ is the matrix Frobenius norm.

We consider the distribution function for the scalar random variable D:

$$\Pr[D < d] = \frac{1}{Z_\beta}\int\limits_{U_d} \mathsf{e}^{-\beta\mathcal{E}}\mathrm{d}V, \quad Z_\beta = \int\limits_{\mathcal{M}} \mathsf{e}^{-\beta\mathcal{E}}\mathrm{d}V, \tag{6.15}$$

where $U_d := \{X \in \mathcal{S}_+^{n,p} | D(X) < d\}$ is the domain of the integral. For simplicity, we only consider symmetric functions such that the random variable D, the energy function $\mathcal{E}$, and the volume form are all invariant under rotations, i.e., the group action by the orthogonal group $\mathcal{O}_n$. We consider an energy function $\mathcal{E}$ satisfying $\mathcal{E}(X) = \mathcal{E}(OXO^T)$, $\forall\, O \in \mathcal{O}_n$, such that the Gibbs distribution function only depends on the spectrum of X when considering (6.15) with $D = \|X\|_F = \sqrt{\lambda_1^2 + \cdots + \lambda_p^2}$. Since $\mathcal{O}_n$ is an isometry group for both metrics g_E and g_{BW}, the volume form $\mathrm{d}V$ in the two cases is also invariant under $\mathcal{O}_n$ action.

Notice that Q and Λ can be used as coordinates of the manifold $\mathcal{S}_+^{n,p}$. The volume form expressed by coordinates Q and Λ is given by

$$\mathrm{d}V = \sqrt{\det g}(\prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_\mathrm{i})\mathrm{d}\mu_{\mathcal{O}_n},$$

where $\mu_{\mathcal{O}_n}$ is the Haar measure on $\mathcal{O}_n$, and g is the matrix of metric g_E or g_{BW} expressed under coordinate Q and λ. For g_E its determinant $\det g$ is

$$\det g = \Big(\prod_{1\le \mathrm{i}<\mathrm{j}\le p} |\lambda_\mathrm{i} - \lambda_\mathrm{j}|^2\Big)\Big(\prod_{1\le \mathrm{i}\le p} \lambda_\mathrm{i}^{2(n-p)}\Big),$$

and for g_{BW} it is

$$\det g = \Big(\prod_{1\le \mathrm{i}<\mathrm{j}\le p} \frac{|\lambda_\mathrm{i} - \lambda_\mathrm{j}|^2}{\lambda_\mathrm{i} + \lambda_\mathrm{j}}\Big)\Big(\prod_{1\le \mathrm{i}\le p} \lambda_\mathrm{i}^{(n-p)}\Big).$$

So for g_E the distribution $\Pr[D < d]$ is expressed as

$$\Pr[D < d] = \frac{1}{Z_\beta}\int\limits_{\|X\|_F < d} \mathsf{e}^{-\beta\mathcal{E}}\mathrm{d}V$$

$$\propto \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}}>0, \mathrm{i}=1,\ldots,p}} \mathsf{e}^{-\beta\mathcal{E}(\lambda_1,\ldots,\lambda_p)}\Big(\prod_{1\leq \mathrm{i}<\mathrm{j}\leq p}|\lambda_{\mathrm{i}}-\lambda_{\mathrm{j}}|\Big)\Big(\prod_{1\leq \mathrm{i}\leq p}\lambda_{\mathrm{i}}^{n-p}\Big)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_p, \tag{6.16}$$

where we have used the fact that the integrand does not depend on the coordinate $Q \in \mathcal{O}_n$, so the integral of $\mu_{\mathcal{O}_n}$ only provides a constant coefficient. As we could always renormalize $\Pr[D < d]$ by considering the quotient $\frac{\Pr[D<d]}{\Pr[D<\infty]}$, we only need the dependence of the integral on parameter d.

Similarly, for the Bures-Wasserstein metric g_{BW} we have

$$\Pr[D < d] \propto \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}}>0, \mathrm{i}=1,\ldots,p}} \mathsf{e}^{-\beta\mathcal{E}(\lambda_1,\ldots,\lambda_p)}\Big(\prod_{1\leq \mathrm{i}<\mathrm{j}\leq p}\frac{|\lambda_{\mathrm{i}}-\lambda_{\mathrm{j}}|}{\sqrt{\lambda_{\mathrm{i}}+\lambda_{\mathrm{j}}}}\Big)\Big(\prod_{1\leq \mathrm{i}\leq p}\lambda_{\mathrm{i}}^{\frac{n-p}{2}}\Big)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_p \tag{6.17}$$

6.5.2 Example I: $\mathcal{E}(X) = \frac{1}{2}\|X\|_F^2$

This is the simplest example. Applying the general expression (6.16), for embedded geometry g_E we have

$$\begin{aligned}
\Pr[D < d] \propto & \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}}>0, \mathrm{i}=1,\ldots,p}} \mathsf{e}^{-\frac{\beta}{2}\sum_{\mathrm{i}=1}^{p}\lambda_{\mathrm{i}}^2}\Big(\prod_{1\leq \mathrm{i}<\mathrm{j}\leq p}|\lambda_{\mathrm{i}}-\lambda_{\mathrm{j}}|\Big)\Big(\prod_{1\leq \mathrm{i}\leq p}\lambda_{\mathrm{i}}^{n-p}\Big)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_p \\
&= \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2}\rho^{N-1}\Big(\int\limits_{S_+^{p-1}}\prod_{1\leq \mathrm{i}<\mathrm{j}\leq p}|\omega_{\mathrm{i}}-\omega_{\mathrm{j}}|\prod_{\mathrm{i}=1}^{p}|\omega_{\mathrm{i}}|^{n-p}\prod_{\mathrm{i}=1}^{p}\mathrm{d}\omega\Big)\mathrm{d}\rho \\
&= \Big(\int\limits_{S_+^{p-1}}\prod_{1\leq \mathrm{i}<\mathrm{j}\leq p}|\omega_{\mathrm{i}}-\omega_{\mathrm{j}}|\prod_{\mathrm{i}=1}^{p}|\omega_{\mathrm{i}}|^{n-p}\prod_{\mathrm{i}=1}^{p}\mathrm{d}\omega\Big)\int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2}\rho^{N-1}\mathrm{d}\rho \\
&\propto \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2}\rho^{N-1}\mathrm{d}\rho,
\end{aligned} \tag{6.18}$$

where we have used the spherical coordinate for $(\lambda_1, ..., \lambda_p) = \rho\omega$, with $\rho = \sqrt{\sum\limits_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2}$ being the radius and $\omega \in S_+^{p-1} = S^{p-1} \cap \mathbb{R}_+^p$ being the coordinate on the positive orthant of the unit sphere.

For g_{BW}, similarly we have

$$\Pr[D < d] \propto \int\limits_0^d \mathsf{e}^{-\beta\rho^2} \rho^{\frac{N}{2}-1} \mathrm{d}\rho. \tag{6.19}$$

Now we can see that $\beta D^2 = \beta \|X\|_F^2$ is subject to $\chi^2(N)$ distribution for the embedded metric g_E, and $\chi^2(\frac{N}{2})$ distribution for the Bures-Wasserstein metric.

6.5.3 Example II: $\mathcal{E}(X) = \mathrm{Tr}(X \log X)$

The second example for the energy function is the von Neumann entropy defined as

$$\mathcal{E}(X) = \mathrm{Tr}(X \log X) = \sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}} \log \lambda_{\mathrm{i}}$$

The minimizers of $\mathcal{E}(X) = \mathrm{Tr}(X \log X)$ on $\mathcal{S}_+^{n,p}$ are matrices $X \in \mathcal{S}_+^{n,p}$ with spectrum $\lambda_1 = \cdots = \lambda_p = \mathsf{e}^{-1}$.

We still consider the scalar random variable $D = \|X\|_F$. Since $\mathcal{E}(X) = \mathrm{Tr}(X \log X) = \sum\limits_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}} \log \lambda_{\mathrm{i}}$ only depends on spectrum, the argument in the previous section about integral on $\mathcal{O}_n$ still applies. Applying (6.16), for g_E we have

$$\begin{aligned}\Pr(D < d) &= \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}} > 0, \mathrm{i}=1,\ldots,p}} \mathsf{e}^{-\beta \sum\limits_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}} \log \lambda_{\mathrm{i}}} \prod_{1 \le \mathrm{i} < \mathrm{j} \le p} |\lambda_{\mathrm{i}} - \lambda_{\mathrm{j}}| \prod_{\mathrm{i}=1}^{p} |\lambda_{\mathrm{i}}|^{n-p} \prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_{\mathrm{i}} \\ &= \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}} > 0, \mathrm{i}=1,\ldots,p}} \prod_{1 \le \mathrm{i} < \mathrm{j} \le p} |\lambda_{\mathrm{i}} - \lambda_{\mathrm{j}}| \prod_{\mathrm{i}=1}^{p} |\lambda_{\mathrm{i}}|^{n-p-\beta\lambda_{\mathrm{i}}} \prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_{\mathrm{i}},\end{aligned}$$

and for g_{BW} we have

$$\Pr(D < d) = \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < d^2 \\ \lambda_{\mathrm{i}} > 0, \mathrm{i}=1,\dots,p}} \prod_{1 \le \mathrm{i} < \mathrm{j} \le p} \frac{|\lambda_{\mathrm{i}} - \lambda_{\mathrm{j}}|}{\sqrt{\lambda_{\mathrm{i}} + \lambda_{\mathrm{j}}}} \prod_{\mathrm{i}=1}^{p} |\lambda_{\mathrm{i}}|^{\frac{n-p}{2} - \beta\lambda_{\mathrm{i}}} \prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_{\mathrm{i}}.$$

Although we do not have a closed expression for both cases, such integrals can be easily approximated by an accurate quadrature when p is small, e.g., $p \le 3$.

6.5.4 Example III: $\mathcal{E}(X) = \frac{1}{2}\|X - A\|_F^2$

In the third example, we consider a quadratic energy function $\mathcal{E}(X) = \frac{1}{2}\|X - A\|_F^2$ where $A \in \mathcal{S}_+^{n,p}$; and the scalar random variable is $D = \|X - A\|_F$. In this example, $\mathcal{O}_n$ symmetry does not hold, and we can only make an estimate of the distribution function.

The distribution function of D is evaluated through

$$\Pr(D < d) \propto \int_{U_d} \mathsf{e}^{-\frac{\beta}{2}D^2} \mathrm{d}V,$$

where $U_d = \{X \in \mathcal{S}_+^{n,p} | D(X) < d\}$. Using delta function, formally we can simplify the integral to

$$\begin{aligned}
\Pr(D < d) &\propto \int_{\mathcal{M}} \mathbf{1}_{\{D<d\}} \mathsf{e}^{-\frac{\beta}{2}D^2} \mathrm{d}V && (6.20)\\
&= \int_{\mathcal{S}_+^{n,p}} \left(\int_0^\infty \mathbf{1}_{\{\rho<d\}} \mathsf{e}^{-\frac{\beta}{2}\rho^2} \delta(D-\rho)\mathrm{d}\rho\right)\mathrm{d}V \\
&= \int_0^\infty \mathbf{1}_{\{\rho<d\}} \mathsf{e}^{-\frac{\beta}{2}\rho^2} \left(\int_{\mathcal{S}_+^{n,p}} \delta(D-\rho)\mathrm{d}V\right)\mathrm{d}\rho \\
&= \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2} \left(\int_{\mathcal{M}} \frac{\mathrm{d}}{\mathrm{d}\rho} \mathbf{1}_{\{D-\rho\}} \mathrm{d}V\right)\mathrm{d}\rho \\
&= \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2} \frac{\mathrm{d}}{\mathrm{d}\rho} \left(\int_{\mathcal{S}_+^{n,p}} \mathbf{1}_{\{D-\rho\}} \mathrm{d}V\right)\mathrm{d}\rho \\
&= \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2} \frac{\mathrm{d}}{\mathrm{d}\rho} V_D(\rho)\mathrm{d}\rho
\end{aligned}$$

(6.21)

where $V_D(\rho) = \int_{\mathcal{M}} \mathbf{1}_{\{D<\rho\}} \mathrm{d}V = \int_{D<\rho} \mathrm{d}V$.

In general, it is difficult to calculate $V_D(\rho)$. But we can intuitively replace it with some approximations. $V_D(\rho)$ is intuitively the volume of the intersection of $\mathcal{S}_+^{n,p}$ and the ball centered at A of radius ρ: $B_A^{n,p}(\rho) := B_A(\rho) \cap \mathcal{S}_+^{n,p}$, where

$$B_A(\rho) = \Big\{ X \in \mathcal{S}^{n\times n} : \|X - A\|_F < \rho \Big\}.$$

Therefore, when the eigenvalues of A are large, we have the following approximation:

$$V_D(\rho) \approx \alpha \rho^N, \tag{6.22}$$

where α is a constant that does not depend on r, N is the dimension of $\mathcal{S}_+^{n,p}$. For g_E, α is exactly the volume of unit ball in $\mathbb{R}^N$, while for g_{BW}, α depends on dimension N and $A \in \mathcal{S}_+^{n,p}$.

For the g_{BW} metric, following similar arguments, we can get the same approximation (6.22). Putting all this together, when A has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p \gg 1$, we have the following

$$\Pr(D < d) \propto \int\limits_{D<d} \mathsf{e}^{-\frac{\beta}{2}D^2} \mathrm{d}V = \int\limits_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2} \frac{\mathrm{d}}{\mathrm{d}\rho}\Big(V_D(\rho)\Big)\mathrm{d}\rho \mathrel{\underset{\sim}{\propto}} \int\limits_0^t \mathsf{e}^{-\frac{\beta}{2}\rho^2} \rho^{N-1} \mathrm{d}\rho, \tag{6.23}$$

where $\underset{\sim}{\propto}$ stands for *being approximately proportional to.*

6.5.5 MCMC Numerical Integration

It is well known that MCMC can be used for integrating a function numerically, and that one of the main advantages is that the convergence rate is independent of the dimension. Both schemes in this chapter are MCMC type sampling schemes on the manifold. Suppose we have generated samples $X_{\rm i}$ satisfying the Gibbs distribution on the manifold, e.g.,

$$X_{\rm i} \sim \frac{1}{Z_\beta} \mathrm{e}^{-\beta \mathcal{E}(X)} \mathrm{d}V_g,$$

where $Z_\beta = \int\limits_{S_+^{n,p}} \mathsf{e}^{-\beta\mathcal{E}(X)} \mathrm{d}V$ is an unknown normalization factor and $\mathrm{d}V$ is the volume form depending on the metric. Then for approximating the integral of a nice function $f(X)$ on the same manifold $\int\limits_{S_+^{n,p}} f(X)\mathrm{d}V$, we can use

$$\frac{1}{m}\sum_{\mathrm{i}=1}^{m} f(X_\mathrm{i})\mathsf{e}^{\beta\mathcal{E}(X_\mathrm{i})} \approx \frac{\int\limits_{S_+^{n,p}} f(X)\mathrm{d}V}{\int\limits_{S_+^{n,p}} \mathsf{e}^{-\beta\mathcal{E}(X)}\mathrm{d}V} = \frac{1}{Z_\beta}\int\limits_{S_+^{n,p}} f(X)\mathrm{d}V, \tag{6.24}$$

because each $f(X_\mathrm{i})\mathsf{e}^{\beta\mathcal{E}(X_\mathrm{i})}$ is a random variable with expectation

$$\mathbb{E}\left[f(X_\mathrm{i})\mathsf{e}^{\beta\mathcal{E}(X_\mathrm{i})}\right] = \frac{1}{Z_\beta}\int\limits_{S_+^{n,p}} f(X_\mathrm{i})\mathsf{e}^{\beta\mathcal{E}(X_\mathrm{i})}\mathsf{e}^{-\beta\mathcal{E}(X_\mathrm{i})}\mathrm{d}V,$$

and the left hand side is a random variable with expectation

$$\mathbb{E}\left[\frac{1}{m}\sum_{\mathrm{i}=1}^{m} f(X_\mathrm{i})\mathsf{e}^{\beta E(X_\mathrm{i})}\right] = \frac{1}{m}\sum_{\mathrm{i}=1}^{m}\mathbb{E}\left[f(X_\mathrm{i})\mathsf{e}^{\beta E(X_\mathrm{i})}\right] = \frac{1}{Z_\beta}\int\limits_{S_+^{n,p}} f(X)\mathrm{d}V,$$

where the expectation $\mathbb{E}[\cdot]$ is taken w.r.t. the Gibbs distribution under the corresponding metric.

So using the generated samples X_i, we can approximate the integral $\int\limits_{S_+^{n,p}} f(X)\mathrm{d}V$ up to a constant Z_β that does not depend on $f(X)$. Notice that the additional advantage of Monte Carlo type quadrature on a manifold is that we do not need to know what $\mathrm{d}V$ is. On the other hand, Z_β cannot be approximated by the same approach. Though we do not consider any specific application for numerical integration, equation (6.24) can be used as one way to validate the Riemannian Langevin Monte Carlo schemes.

For the following special functions, it is possible to calculate the *exact integrals*. For the energy function $\mathcal{E}(X) = \frac{1}{2}\|X\|_F^2$, and a special integrand $f(X) = \|X\|_F^k \mathsf{e}^{-\frac{\alpha}{m}\|X\|_F^m}$ with $k > -N, m > 2, \alpha > 0$, using the results in Section 6.5.2, the distribution of $D = \|X\|_F$ obtains the following explicit forms:

$$\text{for metric } g_E: \quad \Pr[D < d] \propto \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2}\rho^{N-1}\mathrm{d}\rho, \tag{6.25}$$

$$\text{for metric } g_{BW}: \quad \Pr[D<d] \propto \int_0^d \mathsf{e}^{-\frac{\beta}{2}\rho^2}\rho^{\frac{N}{2}-1}\mathrm{d}\rho, \tag{6.26}$$

so the integral on the manifold could be expressed by the expectation of a random variable, which leads to

$$\begin{aligned}
\text{for } g_E :&\frac{1}{Z_\beta}\int_{\mathcal{S}_+^{n,p}} f(X)\mathrm{d}V = \mathbb{E}[f(X)\mathsf{e}^{\frac{\beta}{2}\|X\|_F^2}] = \mathbb{E}[D^k \mathsf{e}^{-\frac{\alpha}{m}D^m}\mathsf{e}^{\frac{\beta}{2}D^2}] \\
&= \frac{\int_0^\infty \rho^k \mathsf{e}^{-\frac{\alpha}{m}\rho^m+\frac{\beta}{2}\rho^2}\rho^{N-1}\mathsf{e}^{-\frac{\beta}{2}\rho^2}\mathrm{d}\rho}{\int_0^\infty \rho^{N-1}\mathsf{e}^{-\frac{\beta}{2}\rho^2}\mathrm{d}\rho} = \frac{\frac{1}{m}(\alpha/m)^{-\frac{k+N}{m}}\Gamma((k+N)/m)}{\frac{1}{2}(\beta/2)^{-N/2}\Gamma(N/2)} \qquad (6.27)\\
\text{for } g_{BW} :&\frac{1}{Z_\beta}\int_{\mathcal{S}_+^{n,p}} f(X)\mathrm{d}V = \mathbb{E}[f(X)\mathsf{e}^{\frac{\beta}{2}\|X\|_F^2}] = \mathbb{E}[D^k \mathsf{e}^{-\frac{\alpha}{m}D^m}\mathsf{e}^{\frac{\beta}{2}D^2}] \\
&= \frac{\int_0^\infty \rho^k \mathsf{e}^{-\frac{\alpha}{m}\rho^m+\frac{\beta}{2}\rho^2}\rho^{\frac{N}{2}-1}\mathsf{e}^{-\frac{\beta}{2}\rho^2}\mathrm{d}\rho}{\int_0^\infty \rho^{\frac{N}{2}-1}\mathsf{e}^{-\frac{\beta}{2}\rho^2}\mathrm{d}\rho} = \frac{\frac{1}{m}(\alpha/m)^{-\frac{k+N/2}{m}}\Gamma((k+N/2)/m)}{\frac{1}{2}(\beta/2)^{-N/4}\Gamma(N/4)}. \qquad (6.28)
\end{aligned}$$

6.6 Numerical Experiments

In this section, we test the samples generated by the two Riemannian Langevin Monte Carlo schemes (6.12) and (6.14) on the examples constructed in the previous section. The samples are generated by the following procedure: we run the iterative schemes (6.12) or (6.14) for sufficiently many $\tilde{m}$ iterations then take the last m iterates as the samples for the Gibbs distribution. Both $\tilde{m}$ and m should be chosen such that the $(\tilde{m}-m)$-th iterate has already reached equilibrium e.g., $\tilde{m}$ is $6,000,000$ and m is $5,000,000$ for specially chosen energy functions and parameters β.

Now suppose we have generated samples $X_{\mathrm{i}} \in \mathcal{S}_+^{n,p}$ $(\mathrm{i}=1,\cdots,m)$ for either metric. In order to test or show the numerical convergence to the Gibbs distribution, we will consider two kinds of numerical tests.

The first kind of tests is to test on the scalar random variable $D(X)=\|X\|_F$ or $D(X)=\|X-A\|_F$ as described in Section 6.5. Then we compare the cumulative distribution function (CDF) of the random variable D with its empirical CDF calculated from the MCMC samples.

Denote the true CDF of D by $F_D(t) := \Pr(D \leq t)$. The empirical CDF of samples is

$$\hat{F}_D(t) := \frac{1}{m}\sum_{\mathrm{i}=1}^{m} \mathbb{1}_{D(X_\mathrm{i})\leq t},$$

where $\mathbb{1}_{D(X_\mathrm{i})\leq t}$ takes value 1 if $D(X_\mathrm{i}) \leq t$, and value 0 if otherwise. The KolmogorovSmirnov test statistic (K-S statistic) is defined by

$$KS_D := \sup_t \left| F_D(t) - \hat{F}_D(t) \right|. \tag{6.29}$$

In our numerical tests, we compute the KS statistic by taking the maximum difference of F_D and $\hat{F}_D$ at 100 equally spaced points in the interval $[0, t_{max}]$ where $F_D(t_{max}) \approx 1$.

The second kind of tests is on the integral examples in Section 6.5.5, let X be a random variable satisfying Gibbs distribution on the manifold $\mathcal{S}_+^{n,p}$ under either metric. Define

$$\mu := \mathrm{E}\, f(X)\mathrm{e}^{\beta\mathcal{E}(X)} = \frac{1}{Z_\beta}\int\limits_{\mathcal{S}_+^{n,p}} f(X)\mathrm{d}V.$$

Given m samples $X_\mathrm{i} \in \mathcal{S}_+^{n,p}$, we define

$$\hat{\mu}_m := \frac{1}{m}\sum_{\mathrm{i}=1}^{m} f(X_\mathrm{i})\mathrm{e}^{\beta\mathcal{E}(X_\mathrm{i})}. \tag{6.30}$$

Notice that samples generated by MCMC are not independent. If we assume

$$\sigma^2 := \mathrm{var}\left(f(X_1)\mathrm{e}^{\beta\mathcal{E}(X_1)}\right) + 2\sum_{k=1}^{\infty}\mathrm{cov}\left(f(X_1)\mathrm{e}^{\beta\mathcal{E}(X_1)}, f(X_{1+k})\mathrm{e}^{\beta\mathcal{E}(X_{1+k})}\right) < \infty,$$

then by the Markov Chain Central Limit Theorem[89, 90], as $m \to \infty$, we have

$$\sqrt{m}(\hat{\mu}_m - \mu) \to \mathcal{N}(0, \sigma^2) \tag{6.31}$$

where the convergence is in the sense of distribution. Thus if $m \gg 1$, $\frac{\hat{\mu}_m - \mu}{\mu}$ roughly follows the distribution $\mathcal{N}(0, O(\frac{1}{m}))$ and the relative error term $\left|\frac{\hat{\mu}_m - \mu}{\mu}\right|$ roughly follows the folded normal distribution with mean $O(\frac{1}{\sqrt{m}})$ and variance $O(\frac{1}{m})$. Hence we can use $\hat{\mu}_m$ defined in (6.30) to estimate $\mu = \frac{1}{Z_\beta} \int\limits_{\mathcal{S}_+^{n,p}} f(X) \mathrm{d}V$, and the relative error is $O(\frac{1}{\sqrt{m}})$.

6.6.1 Numerical Validation of the CDF of the Scalar Variable $D(X)$

The manifold $\mathcal{S}_+^{n,p}$ has dimension $N = np - p(p-1)/2$. For both metrics, we consider three examples in Section 6.5 with special energy functions $\mathcal{E}$ in the Gibbs distribution $\mathsf{e}^{-\beta\mathcal{E}}$ and the CDF for the scalar variable $D(X)$:

1. Example I: $\mathcal{E}(X) = \frac{1}{2}\|X\|_F^2$ with the CDF for $D(X) = \|X\|_F$:

$$\text{For } g_E: \quad F_D(t) = \Pr(\|X\|_F \leq t) \propto \int\limits_0^t \mathsf{e}^{-\frac{\beta}{2}\rho^2} \rho^{N-1} \mathrm{d}\rho,$$

$$\text{For } g_{BW}: \quad F_D(t) = \Pr(\|X\|_F \leq t) \propto \int\limits_0^t \mathsf{e}^{-\frac{\beta}{2}\rho^2} \rho^{N/2-1} \mathrm{d}\rho.$$

2. Example II: $\mathcal{E}(X) = \operatorname{tr}(X \log X)$ with the CDF $F_D(t) = \Pr(\|X\|_F \leq t)$ for $D(X) = \|X\|_F$:

$$\text{For } g_E: \quad F_D(t) \propto \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < t^2 \\ \lambda_{\mathrm{i}} > 0, \mathrm{i}=1,\dots,p}} \prod_{1 \leq \mathrm{i} < \mathrm{j} \leq p} |\lambda_{\mathrm{i}} - \lambda_{\mathrm{j}}| \prod_{\mathrm{i}=1}^{p} |\lambda_{\mathrm{i}}|^{n-p-\beta\lambda_{\mathrm{i}}} \prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_{\mathrm{i}},$$

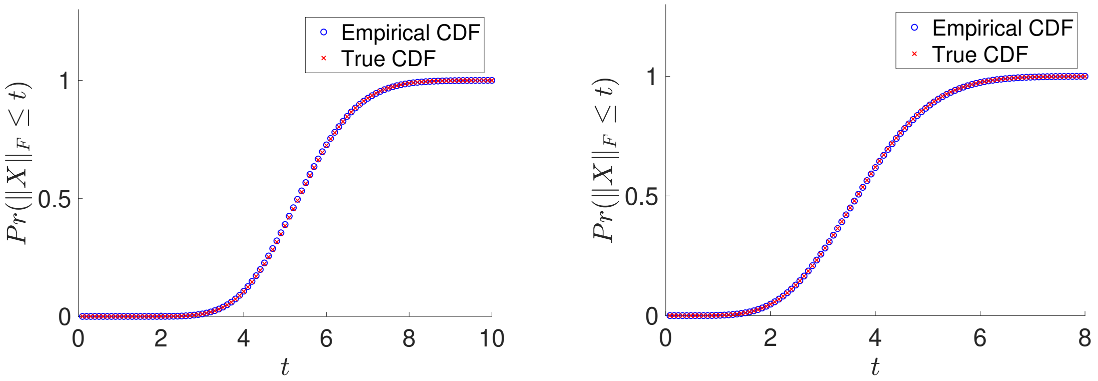
$$\text{For } g_{BW}: \quad F_D(t) \propto \int\limits_{\substack{\sum_{\mathrm{i}=1}^{p} \lambda_{\mathrm{i}}^2 < t^2 \\ \lambda_{\mathrm{i}} > 0, \mathrm{i}=1,\dots,p}} \prod_{1 \leq \mathrm{i} < \mathrm{j} \leq p} \frac{|\lambda_{\mathrm{i}} - \lambda_{\mathrm{j}}|}{\sqrt{\lambda_{\mathrm{i}} + \lambda_{\mathrm{j}}}} \prod_{\mathrm{i}=1}^{p} |\lambda_{\mathrm{i}}|^{\frac{n-p-1}{2} - \beta\lambda_{\mathrm{i}}} \prod_{\mathrm{i}=1}^{p} \mathrm{d}\lambda_{\mathrm{i}}.$$

which is a p-fold integral and can be approximated accurately by quadrature such as Simpson's rule for relatively small values of p, e.g., $p = 2, 3$.

3. Example III: $\mathcal{E}(X) = \frac{1}{2}\|X - A\|_F^2$ where $A \in \mathcal{S}_+^{n,p}$ has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p \gg 1$, with the CDF for $D(X) = \|X - A\|_F$:

$$\text{For both } g_E \text{ and } g_{BW}: \quad F_D(t) = \Pr(\|X - A\|_F \leq t) \propto \int_0^t \mathrm{e}^{-\frac{\beta}{2}\rho^2}\rho^{N-1}\mathrm{d}\rho.$$

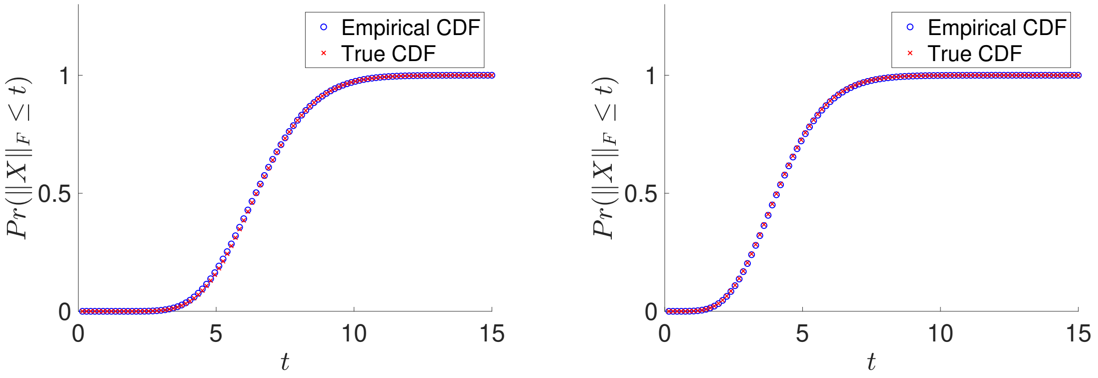
In the implementation of the scheme, the step size Δt and β in schemes (6.12) and (6.14) are two parameters that need to be tuned to reach equilibrium with reasonable computing time. We first use a numerically stable Δt then adjust β so that the noise term has reasonable variance. And of course one needs a sufficiently large number of iterations for schemes (6.12) and (6.14) to reach their equilibrium state, and a sufficiently large number m of samples to observe numerical convergence toward the Gibbs distribution through the scalar random variable D, e.g., the KS statistic (6.29) should be small. See Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4 for the numerical results.



(a) Scheme E (6.12) on $(\mathcal{S}_+^{n,p}, g_E)$ with $\Delta t = 0.001$ and $\beta = 0.4$. The error between two CDFs is $KS = 0.0054$.

(b) Scheme BW (6.14) on $(\mathcal{S}_+^{n,p}, g_{BW})$ with $\Delta t = 0.001$ and $\beta = 0.4$. The error between two CDFs is $KS = 0.0023$.

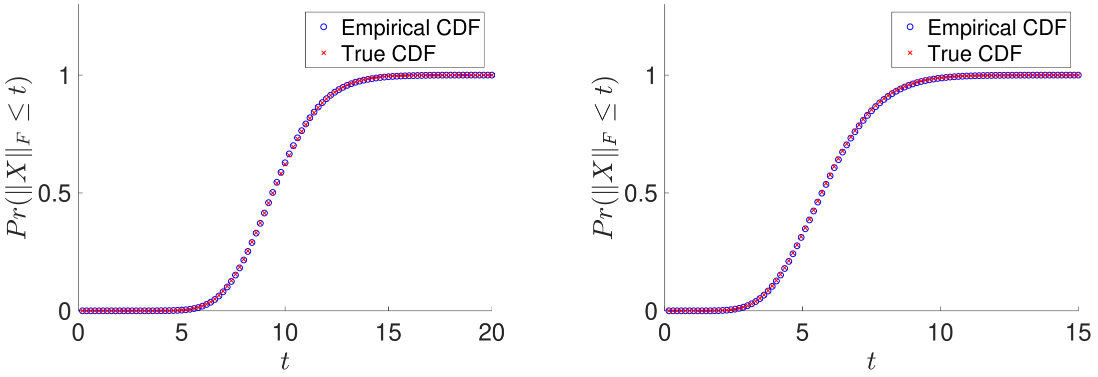
Figure 6.1. Example I: $\mathcal{E}(X) = \frac{1}{2}\|X\|_F^2, n = 5, p = 3$ and manifold dimension is $N = 12$. The empirical CDF is computed by $5E6$ MCMC samples generated after $6E6$ iterations of the Riemannian Langevin Monte Carlo schemes. Both CDFs of scheme E and scheme BW are evaluated at 100 equally spaced points on $[0, 10]$ and $[0, 8]$, respectively, and the difference can be measured by the KS statistic (6.29).



(a) Scheme E (6.12) on $(\mathcal{S}_+^{n,p}, g_E)$ with $\Delta t = 0.001$ and $\beta = 0.5$. The error between two CDFs is $KS = 0.0096$

(b) Scheme BW (6.14) on $(\mathcal{S}_+^{n,p}, g_{BW})$ with $\Delta t = 0.001$ and $\beta = 0.5$. The error between two CDFs is $KS = 0.0043$.

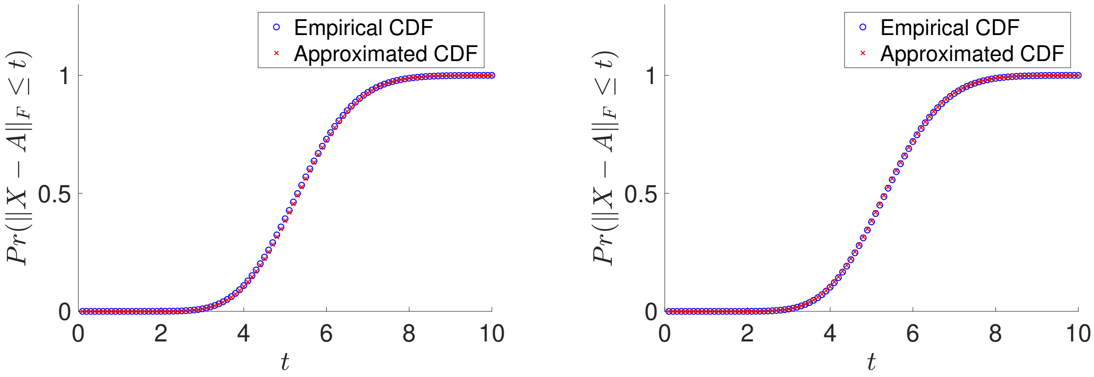
Figure 6.2. Example II: $\mathcal{E}(X) = \mathrm{tr}(X \log X)$, $n = 5, p = 3$ and manifold dimension is $N = 12$. The empirical CDF is computed by $5E6$ MCMC samples generated after $6E6$ iterations of the Riemannian Langevin Monte Carlo schemes. Both CDFs are evaluated at 100 equally spaced points on $[0, 15]$, and the difference can be measured by the KS statistic (6.29).



(a) Scheme E (6.12) on $(\mathcal{S}_+^{n,p}, g_E)$ with $\Delta t = 0.001$ and $\beta = 0.5$. The error between two CDFs is $KS = 0.006$.

(b) Scheme BW (6.14) on $(\mathcal{S}_+^{n,p}, g_{BW})$ with $\Delta t = 0.001$ and $\beta = 0.5$. The error between two CDFs is $KS = 0.0043$.

Figure 6.3. Example II: $\mathcal{E}(X) = \mathrm{tr}(X \log X)$, $n = 10, p = 2$ and manifold dimension is $N = 19$. The empirical CDF is computed by $5E6$ MCMC samples generated after $6E6$ iterations of the Riemannian Langevin Monte Carlo schemes. Both CDFs of scheme E and scheme BW are evaluated at 100 equally spaced points on $[0, 20]$ and $[0, 15]$,respectively, and the difference can be measured by the KS statistic (6.29).



(a) Scheme E (6.12) on $(\mathcal{S}_+^{n,p}, g_E)$ with $\Delta t = 0.001$ and $\beta = 0.4$. The error between two CDFs is $KS = 0.0084$.

(b) Scheme BW (6.14) on $(\mathcal{S}_+^{n,p}, g_{BW})$ with Δt =2E-7 and $\beta = 0.4$. The error between two CDFs is $KS = 0.0052$.

Figure 6.4. Example III: $\mathcal{E}(X) = \frac{1}{2}\|X - A\|_F^2$, $n = 5, p = 3$ and manifold dimension is $N = 12$. The nonzero eigenvalues of A are equally spaced between 10000 and 20000. The empirical CDF is computed by $5E6$ MCMC samples generated after $6E6$ iterations of the Riemannian Langevin Monte Carlo schemes. Both CDFs are evaluated at 100 equally spaced points on $[0, 10]$, and the difference can be measured by the KS statistic (6.29).

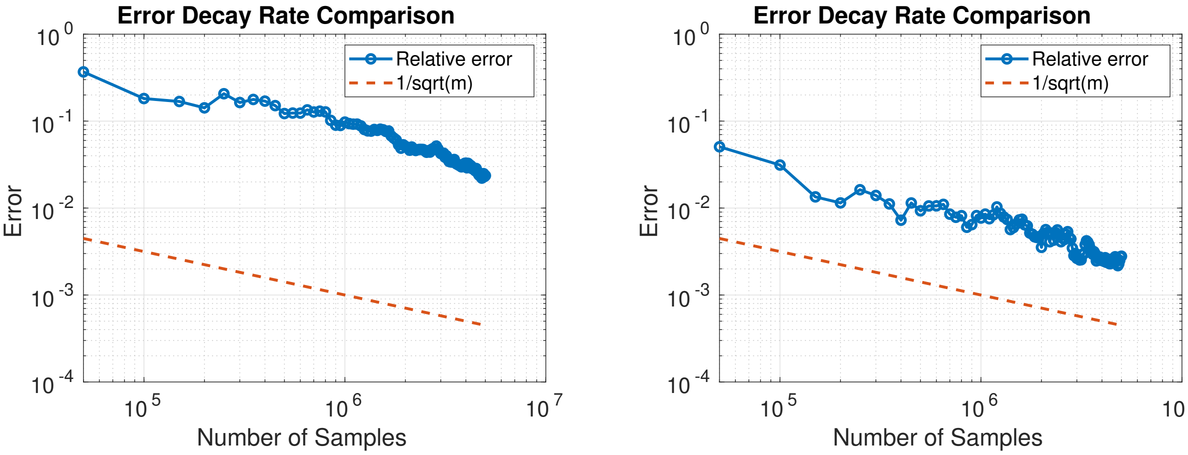
6.6.2 Validation using MCMC Numerical Integration

We consider special cases $k = 0, m = 2$ in the examples (6.27) and (6.28), then (6.27) reduces to $(\frac{\beta}{\alpha})^{N/2}$ and (6.28) reduces to $(\frac{\beta}{\alpha})^{N/4}$. In other words, we may verify the numerical convergence of samples X_{i} to Gibbs distribution by verifying

$$\text{For } g_E : \qquad \frac{1}{m}\sum_{\mathrm{i}=1}^{m} \mathrm{e}^{-\frac{\alpha-\beta}{2}\|X_{\mathrm{i}}\|_F^2} \to \left(\frac{\beta}{\alpha}\right)^{N/2}, \tag{6.32}$$

$$\text{For } g_{BW} : \qquad \frac{1}{m}\sum_{\mathrm{i}=1}^{m} \mathrm{e}^{-\frac{\alpha-\beta}{2}\|X_{\mathrm{i}}\|_F^2} \to \left(\frac{\beta}{\alpha}\right)^{N/4}. \tag{6.33}$$

In Figure 6.5 we indeed observe the $O(1/\sqrt{m})$ for the relative error of numerical integration.



(a) Integration on $(\mathcal{S}_+^{n,p}, g_E)$ via samples generated by Scheme E (6.12) with $\Delta t = 0.001$ and $\beta = 0.4$.

(b) Integration on $(\mathcal{S}_+^{n,p}, g_{BW})$ via samples generated by Scheme BW (6.14) with $\Delta t = 0.001$ and $\beta = 0.4$.

Figure 6.5. Convergence rate of the relative error of $\left|\frac{\hat{\mu}_m-\mu}{\mu}\right|$ MCMC integration on the manifold with $n = 10, p = 2$ and dimension $N = 19$. Parameters are $\alpha = 0.75, \beta = 0.4$, for which it is a numerical integration of the function $\mathcal{E}(X) = \frac{1}{2}\|X\|_F^2$ on the manifold $\mathcal{S}_+^{n,p}$. The error shown is the averaged one of 12 independent runs.

6.7 Concluding Remarks

We have constructed two efficient Riemannian Langevin Monte Carlo schemes for sampling PSD matrices of fixed rank from the Gibbs distribution on the manifold $\mathcal{S}_+^{n,p}$ equipped with two fundamental metrics: the embedded metric and the Bures-Wasserstein metric. We have also provided several examples for which these sampling schemes can be numerically validated for correctness.

7. SUMMARY AND FUTURE WORKS

This dissertation develops a unified geometric framework for optimization and sampling over the manifold of fixed-rank positive semidefinite (PSD) matrices. These types of constraints arise naturally in a variety of applications, including matrix completion, phase retrieval, eigenvalue problems, and Bayesian inference. Motivated by both practical needs and theoretical challenges, we investigate efficient algorithms that exploit the underlying Riemannian manifold structure of the constraint set.

We begin by analyzing the manifold geometry of Hermitian PSD matrices of fixed rank, considering both embedded and quotient representations. This leads to three different but closely related formulations of Riemannian optimization over PSD matrices. We show how these formulations relate to each other, and we study their computational and theoretical implications.

Next, we analyze the convergence and performance of Riemannian optimization algorithms within this framework. Particular attention is paid to orthogonalization-free methods and the impact of rank-deficiency on the conditioning of the problem. We derive and validate condition number estimates for Riemannian Hessians and justify the performance of numerical algorithms.

In addition to optimization, we extend this framework to the stochastic setting by developing Riemannian Langevin Monte Carlo algorithms designed for sampling over fixed-rank PSD manifolds. We propose two sampling schemes: one based on the embedded geometry and one on quotient geometry and validate their correctness and efficiency through numerical experiments.

The theoretical insights and algorithmic developments in this thesis are supported by extensive numerical results across a variety of problem domains. Collectively, the contributions provide a rigorous and practical foundation for computation over low-rank PSD matrix manifolds, bridging the gap between abstract geometric tools and real-world applications.

For future works, there remains an unanswered question in the work [75]: how to show the convergence of our Riemannian gradient-based method for the rank-deficient case, and in what sense does the convergence occur? We assume that the iterates generated by RCG on solving (3.2) will converge to a minimizer of (1.1), but such a convergence has not been rigorously justified. In fact, if the minimizer $\hat{X}$ is rank-deficient, $\hat{X}$ is not even in the fixed rank constraint set. One of my future interests aims to answer this question rigorously.

There have been several works in literature related to such a convergence question, e.g., [91] proposed a preconditioned gradient descent whose function value still has linear convergence for the rank-deficient case. However, it does not consider any manifold structure. Even though the preconditioning itself can be understood as a consequence of some Riemannian metric, it does not imply directly the convergence of the Riemannian gradient-based method w.r.t. that metric.

Instead of considering the fixed-rank manifold, there has been increasing interest in considering stratified sets [92, 93], which consider $\mathcal{S}_{\leq p}^{n\times n}$, the set of n-by-n PSD matrices of rank $\leq p$. $\mathcal{S}_{\leq p}^{n\times n}$ is no longer a manifold, but a collection of fixed-rank manifolds. Each fixed-rank manifold behaves like a stratum, conceptually suggesting the highest rank manifold stacks on top of lower-rank manifolds. Similar geometric manifold concepts, such as tangent spaces and gradients, can be generalized to stratified sets. If the constraint is the stratified set, optimization naturally happens to consider all rank scenarios and the convergence, if proved, would be more rigorous than the current one.

REFERENCES

[1] J. Bien and R. J. Tibshirani, “Sparse estimation of a covariance matrix,” *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.

[2] G. Meyer, S. Bonnabel, and R. Sepulchre, “Regression on fixed-rank positive semidefinite matrices: a Riemannian approach,” *The Journal of Machine Learning Research*, vol. 12, pp. 593–625, 2011.

[3] S. Burer and R. D. Monteiro, “Local minima and convergence in low-rank semidefinite programming,” *Mathematical programming*, vol. 103, no. 3, pp. 427–444, 2005.

[4] E. Massart and P.-A. Absil, “Quotient Geometry with Simple Geodesics for the Manifold of Fixed-Rank Positive-Semidefinite Matrices,” en, *SIAM Journal on Matrix Analysis and Applications*, vol. 41, no. 1, pp. 171–198, Jan. 2020, ISSN: 0895-4798, 1095-7162. (visited on 09/22/2021).

[5] B. Vandereycken, P. A. Absil, and S. Vandewalle, “A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank,” en, *IMA Journal of Numerical Analysis*, vol. 33, no. 2, pp. 481–514, Apr. 2013, ISSN: 0272-4979, 1464-3642. DOI: 10.1093/imanum/drs006. [Online]. Available: https://academic.oup.com/imajna/article-lookup/doi/10.1093/imanum/drs006 (visited on 09/22/2021).

[6] S. Bonnabel, G. Meyer, and R. Sepulchre, “Adaptive filtering for estimation of a low-rank positive semidefinite matrix,” in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*, 2010.

[7] B. Vandereycken and S. Vandewalle, “A Riemannian Optimization Approach for Computing Low-Rank Solutions of Lyapunov Equations,” en, *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2553–2579, Jan. 2010, ISSN: 0895-4798, 1095-7162. DOI: 10.1137/090764566. [Online]. Available: http://epubs.siam.org/doi/10.1137/090764566 (visited on 09/22/2021).

[8] E. J. Candes, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[9] E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via wirtinger flow: Theory and algorithms,” *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[10] V. Jugnon and L. Demanet, “Interferometric inversion: A robust approach to linear inverse problems,” in *SEG International Exposition and Annual Meeting*, SEG, 2013.

[11] L. Demanet and V. Jugnon, “Convex recovery from interferometric measurements,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 282–295, 2017.

[12] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan, “Underdamped Langevin MCMC: A non-asymptotic analysis,” in *Conference on learning theory*, PMLR, 2018, pp. 300–323.

[13] X. Cheng, D. Yin, P. Bartlett, and M. Jordan, “Stochastic gradient and Langevin processes,” in *International Conference on Machine Learning*, PMLR, 2020, pp. 1810–1819.

[14] Y. Du and I. Mordatch, “Implicit generation and modeling with energy based models,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[15] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, “Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics,” *ChemPhysChem*, vol. 6, no. 9, pp. 1809–1814, 2005.

[16] G. Ciccotti, T. Lelievre, and E. Vanden-Eijnden, “Projection of diffusions on submanifolds: Application to mean force computation,” *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 61, no. 3, pp. 371–408, 2008.

[17] M. Girolami and B. Calderhead, “Riemann manifold langevin and hamiltonian monte carlo methods,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 2, pp. 123–214, 2011.

[18] M. Brubaker, M. Salzmann, and R. Urtasun, “A Family of MCMC Methods on Implicitly Defined Manifolds,” in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, N. D. Lawrence and M. Girolami, Eds., ser. Proceedings of Machine Learning Research, vol. 22, La Palma, Canary Islands: PMLR, Apr. 2012, pp. 161–172.

[19] S. Byrne and M. Girolami, “Geodesic Monte Carlo on embedded manifolds,” *Scandinavian Journal of Statistics*, vol. 40, no. 4, pp. 825–845, 2013.

[20] E. Zappa, M. Holmes-Cerfon, and J. Goodman, “Monte Carlo on manifolds: sampling densities and integrating functions,” *Communications on Pure and Applied Mathematics*, vol. 71, no. 12, pp. 2609–2647, 2018.

[21] A. Moitra and A. Risteski, “Fast Convergence for Langevin with Matrix Manifold Structure,” in *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

[22] R. Ge, H. Lee, J. Lu, and A. Risteski, “Efficient sampling from the bingham distribution,” in *Algorithmic Learning Theory*, PMLR, 2021, pp. 673–685.

[23] J. Leake, C. McSwiggen, and N. K. Vishnoi, “Sampling matrices from harish-chandra–itzykson–zuber densities with applications to quantum inference and differential privacy,” in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 1384–1397.

[24] M. B. Li and M. A. Erdogdu, *Riemannian langevin algorithm for solving semidefinite programs*, 2023. arXiv: 2010.11176 `[stat.ML]`.

[25] P. Xu, J. Chen, D. Zou, and Q. Gu, *Global convergence of langevin dynamics based algorithms for nonconvex optimization*, 2020. arXiv: 1707.06618 `[stat.ML]`.

[26] J. S. Liu and J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer, 2001, vol. 75.

[27] M. Coste, “AN INTRODUCTION TO SEMIALGEBRAIC GEOMETRY,” en, *Citeseer*, p. 26, 2000.

[28] C. G. Gibson, *Singular points of smooth mappings* (Research notes in mathematics 25), en. London ; San Francisco: Pitman, 1979, ISBN: 978-0-273-08410-5.

[29] J. M. Lee, *Introduction to Smooth Manifolds* (Graduate Texts in Mathematics), en. New York, NY: Springer New York, 2012, vol. 218, ISBN: 978-1-4419-9981-8 978-1-4419-9982-5.

[30] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, en. Princeton, N.J. ; Woodstock: Princeton University Press, 2008, OCLC: ocn174129993, ISBN: 978-0-691-13298-3.

[31] F. Brickell and R. S. Clark, *Differentiable Manifolds: an introduction*, en. Van Nostrand Reinhold, 1970, Google-Books-ID: 25EJNQEACAAJ, ISBN: 978-0-442-01051-5.

[32] E. Massart, J. M. Hendrickx, and P.-A. Absil, “Curvature of the manifold of fixed-rank positive-semidefinite matrices endowed with the Bures–Wasserstein metric,” in *Geometric Science of Information: 4th International Conference, GSI 2019, Toulouse, France, August 27–29, 2019, Proceedings*, Springer, 2019, pp. 739–748.

[33] J. v. Oostrum, “Bures–Wasserstein geometry for positive-definite Hermitian matrices and their trace-one subset,” *Information Geometry*, vol. 5, no. 2, pp. 405–425, 2022.

[34] R. Bhatia, T. Jain, and Y. Lim, “On the Bures–Wasserstein distance between positive definite matrices,” *Expositiones Mathematicae*, vol. 37, no. 2, pp. 165–191, 2019.

[35] A. Han, B. Mishra, P. K. Jawanpuria, and J. Gao, “On Riemannian optimization over positive definite matrices with the Bures-Wasserstein geometry,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 8940–8953, 2021.

[36] W. Huang, K. A. Gallivan, and X. Zhang, "Solving PhaseLift by Low-Rank Riemannian Optimization Methods for Complex Semidefinite Constraints," en, *SIAM Journal on Scientific Computing*, vol. 39, no. 5, B840–B859, Jan. 2017, ISSN: 1064-8275, 1095-7197. (visited on 09/22/2021).

[37] W. Huang, "Optimization algorithms on Riemannian manifolds with applications," Ph.D. dissertation, The Florida State University, 2013.

[38] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization—extended version," *arXiv:1209.3834 [math]*, Sep. 2012, arXiv: 1209.3834. [Online]. Available: http://arxiv.org/abs/1209.3834.

[39] P.-A. Absil and J. Malick, "Projection-like Retractions on Matrix Manifolds," en, *SIAM Journal on Optimization*, vol. 22, no. 1, pp. 135–158, Jan. 2012, ISSN: 1052-6234, 1095-7189. DOI: 10.1137/100802529. [Online]. Available: http://epubs.siam.org/doi/10.1137/100802529 (visited on 09/22/2021).

[40] S. Burer and R. D. Monteiro, "Local Minima and Convergence in Low-Rank Semidefinite Programming," en, *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, Jul. 2005, ISSN: 0025-5610, 1436-4646. DOI: 10.1007/s10107-004-0564-1. [Online]. Available: http://link.springer.com/10.1007/s10107-004-0564-1 (visited on 09/22/2021).

[41] B. Vandereycken, "Low-Rank Matrix Completion by Riemannian Optimization," en, *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1214–1236, Jan. 2013, ISSN: 1052-6234, 1095-7189. DOI: 10.1137/110845768. [Online]. Available: http://epubs.siam.org/doi/10.1137/110845768 (visited on 09/22/2021).

[42] D. Kressner, M. Steinlechner, and B. Vandereycken, "Low-rank tensor completion by Riemannian optimization," en, *BIT Numerical Mathematics*, vol. 54, no. 2, pp. 447–468, Jun. 2014, ISSN: 0006-3835, 1572-9125. (visited on 11/17/2021).

[43] B. Vandereycken, P.-A. Absil, and S. Vandewalle, "Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank," en, in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, Cardiff, United Kingdom: IEEE, Aug. 2009, pp. 389–392, ISBN: 978-1-4244-2709-3. DOI: 10.1109/SSP.2009.5278558. [Online]. Available: http://ieeexplore.ieee.org/document/5278558/ (visited on 09/22/2021).

[44] W. Huang, K. A. Gallivan, and P.-A. Absil, "A Broyden Class of Quasi-Newton Methods for Riemannian Optimization," en, *SIAM Journal on Optimization*, vol. 25, no. 3, pp. 1660–1685, Jan. 2015, ISSN: 1052-6234, 1095-7189. DOI: 10.1137/140955483. [Online]. Available: http://epubs.siam.org/doi/10.1137/140955483 (visited on 09/22/2021).

[45] N. Boumal, V. Voroninski, and A. S. Bandeira, "Deterministic Guarantees for Burer-Monteiro Factorizations of Smooth Semidefinite Programs," *Communications on Pure and Applied Mathematics*, vol. 73, no. 3, pp. 581–608, 2020.

[46] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[47] J. Cheeger, "A lower bound for the smallest eigenvalue of the Laplacian," in *Problems in analysis*, Princeton University Press, 2015, pp. 195–200.

[48] W. E. Donath and A. J. Hoffman, "Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices," *IBM Technical Disclosure Bulletin*, vol. 15, no. 3, pp. 938–944, 1972.

[49] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak mathematical journal*, vol. 23, no. 2, pp. 298–305, 1973.

[50] J. Lu and H. Yang, "Preconditioning orbital minimization method for planewave discretization," *Multiscale Modeling & Simulation*, vol. 15, no. 1, pp. 254–273, 2017.

[51] J. Lu and H. Yang, "A cubic scaling algorithm for excited states calculations in particle–particle random phase approximation," *Journal of Computational Physics*, vol. 340, pp. 297–308, 2017, ISSN: 0021-9991.

[52] Z. Wang, Y. Li, and J. Lu, "Coordinate descent full configuration interaction," *Journal of chemical theory and computation*, vol. 15, no. 6, pp. 3558–3569, 2019.

[53] Q. Pang and H. Yang, "A distributed block chebyshev-davidson algorithm for parallel spectral clustering," *arXiv preprint arXiv:2212.04443*, 2022.

[54] R. R. Coifman *et al.*, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proceedings of the national academy of sciences*, vol. 102, no. 21, pp. 7426–7431, 2005.

[55] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *International Conference on Machine Learning*, PMLR, 2017, pp. 1724–1732.

[56] W. Gao, Y. Li, and B. Lu, "Global Convergence of Triangularized Orthogonalization-free Method," *arXiv preprint arXiv:2110.06212*, 2021.

[57] W. Gao, Y. Li, and B. Lu, "Triangularized orthogonalization-free method for solving extreme eigenvalue problems," *Journal of Scientific Computing*, vol. 93, no. 3, pp. 1–28, 2022.

[58] A. V. Knyazev, "Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method," *SIAM Journal on Scientific Computing*, vol. 23, no. 2, pp. 517–541, 2001.

[59] Y. Zhou, Y. Saad, M. L. Tiago, and J. R. Chelikowsky, “Self-consistent-field calculations using chebyshev-filtered subspace iteration,” *Journal of Computational Physics*, vol. 219, no. 1, pp. 172–184, 2006, ISSN: 0021-9991.

[60] K. Neymeyr, “A geometric theory for preconditioned inverse iteration IV: On the fastest convergence cases,” *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 114–139, 2006, Special Issue on Large Scale Linear and Nonlinear Eigenvalue Problems, ISSN: 0024-3795.

[61] E. S. Coakley and V. Rokhlin, “A fast divide-and-conquer algorithm for computing the spectra of real symmetric tridiagonal matrices,” *Applied and Computational Harmonic Analysis*, vol. 34, no. 3, pp. 379–414, 2013, ISSN: 1063-5203.

[62] H. M. Aktulga, L. Lin, C. Haine, E. G. Ng, and C. Yang, “Parallel eigenvalue calculation based on multiple shift–invert Lanczos and contour integral based spectral projection method,” *Parallel Computing*, vol. 40, no. 7, pp. 195–212, 2014, 7th Workshop on Parallel Matrix Algorithms and Applications, ISSN: 0167-8191.

[63] R. Li, Y. Xi, E. Vecharynski, C. Yang, and Y. Saad, “A Thick-Restart Lanczos Algorithm with Polynomial Filtering for Hermitian Eigenvalue Problems,” *SIAM Journal on Scientific Computing*, vol. 38, no. 4, A2512–A2534, 2016.

[64] E. Polizzi, “Density-matrix-based algorithm for solving eigenvalue problems,” *Physical Review B—Condensed Matter and Materials Physics*, vol. 79, no. 11, p. 115 112, 2009.

[65] T. Sakurai and H. Tadano, “CIRR: a Rayleigh-Ritz type method with contour integral for generalized eigenvalue problems,” *Hokkaido Mathematical Journal*, vol. 36, no. 4, pp. 745–757, 2007.

[66] Y. Xi and Y. Saad, “Computing partial spectra with least-squares rational filters,” *SIAM Journal on Scientific Computing*, vol. 38, no. 5, A3020–A3045, 2016.

[67] X. Ye, J. Xia, R. H. Chan, S. Cauley, and V. Balakrishnan, “A Fast Contour-Integral Eigensolver for Non-Hermitian Matrices,” *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 4, pp. 1268–1297, 2017.

[68] S. Zheng, H. Yang, and X. Zhang, “On the convergence of orthogonalization-free conjugate gradient method for extreme eigenvalues of hermitian matrices: A riemannian optimization interpretation,” *Journal of Computational and Applied Mathematics*, vol. 451, p. 116 053, 2024, ISSN: 0377-0427. DOI: https://doi.org/10.1016/j.cam.2024.116053. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377042724003030.

[69] F. Corsetti, “The orbital minimization method for electronic structure calculations with finite-range atomic basis sets,” *Computer Physics Communications*, vol. 185, no. 3, pp. 873–883, 2014, ISSN: 0010-4655.

[70] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999.

[71] U. Hetmaniuk and R. Lehoucq, "Basis selection in LOBPCG," *Journal of Computational Physics*, vol. 218, no. 1, pp. 324–332, 2006, ISSN: 0021-9991.

[72] J. A. Duersch, M. Shao, C. Yang, and M. Gu, "A robust and efficient implementation of LOBPCG," *SIAM Journal on Scientific Computing*, vol. 40, no. 5, pp. C655–C676, 2018.

[73] X. Liu, Z. Wen, and Y. Zhang, "An efficient Gauss–Newton algorithm for symmetric low-rank product matrix approximations," *SIAM Journal on Optimization*, vol. 25, no. 3, pp. 1571–1608, 2015.

[74] Y. Li, J. Lu, and Z. Wang, "Coordinatewise descent methods for leading eigenvalue problem," *SIAM Journal on Scientific Computing*, vol. 41, no. 4, A2681–A2716, 2019.

[75] S. Zheng, W. Huang, B. Vandereycken, and X. Zhang, "Riemannian optimization using three different metrics for hermitian psd fixed-rank constraints," *Computational Optimization and Applications*, vol. 91, no. 3, pp. 1135–1184, Jul. 2025, ISSN: 1573-2894. DOI: 10.1007/s10589-025-00687-8. [Online]. Available: https://doi.org/10.1007/s10589-025-00687-8.

[76] H. Sato and T. Iwai, "A new, globally convergent Riemannian conjugate gradient method," *Optimization*, vol. 64, no. 4, pp. 1011–1031, 2015.

[77] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on optimization*, vol. 2, no. 1, pp. 21–42, 1992.

[78] D. H. Gutman and N. Ho-Nguyen, *Coordinate Descent Without Coordinates: Tangent Subspace Descent on Riemannian Manifolds*, arXiv:1912.10627 [math], Jun. 2020. DOI: 10.48550/arXiv.1912.10627. [Online]. Available: http://arxiv.org/abs/1912.10627.

[79] T. Yu, S. Zheng, J. Lu, G. Menon, and X. Zhang, *Riemannian langevin monte carlo schemes for sampling psd matrices with fixed rank*, 2023. arXiv: 2309.04072 `[math.NA]`. [Online]. Available: https://arxiv.org/abs/2309.04072.

[80] B. Vandereycken, P.-A. Absil, and S. Vandewalle, "Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank," in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, IEEE, 2009, pp. 389–392.

[81] D. Inauen and G. Menon, *Stochastic nash evolution*, 2024. arXiv: 2312.06541 `[math.PR]`. [Online]. Available: https://arxiv.org/abs/2312.06541.

[82] G. Menon and T. Yu, *The Riemannian Langevin equation and conic programs*, 2023. arXiv: 2302.11653 `[math.PR]`.

[83] G. Menon and T. Yu, *Siegel brownian motion*, 2023. arXiv: 2309.04299 `[math.PR]`. [Online]. Available: https://arxiv.org/abs/2309.04299.

[84] E. P. Hsu, *Stochastic analysis on manifolds* (Graduate Studies in Mathematics). American Mathematical Society, Providence, RI, 2002, vol. 38, pp. xiv+281, ISBN: 0-8218-0802-8.

[85] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes* (North-Holland Mathematical Library), Second. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989, vol. 24, pp. xvi+555, ISBN: 0-444-87378-3.

[86] C.-P. Huang, D. Inauen, and G. Menon, "Motion by mean curvature and Dyson Brownian motion," *Electron. Commun. Probab.*, vol. 28, pp. 1–10, 2023, ISSN: 1083-589X. DOI: 10.1214/23-ECP540.

[87] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[88] T. Pacini, "Mean curvature flow, orbits, moment maps," *Trans. Amer. Math. Soc.*, vol. 355, no. 8, pp. 3343–3357, 2003, ISSN: 0002-9947. DOI: 10.1090/S0002-9947-03-03307-5. [Online]. Available: https://doi.org/10.1090/S0002-9947-03-03307-5.

[89] G. L. Jones, "On the Markov chain central limit theorem," *Probability Surveys*, vol. 1, no. none, pp. 299–320, 2004. DOI: 10.1214/154957804100000051. [Online]. Available: https://doi.org/10.1214/154957804100000051.

[90] C. J. Geyer, "Markov chain monte carlo lecture notes," *Course notes, Spring Quarter*, vol. 80, 1998.

[91] G. Zhang, S. Fattahi, and R. Y. Zhang, "Preconditioned gradient descent for overparameterized nonconvex burer–monteiro factorization with global optimality certification," *Journal of Machine Learning Research*, vol. 24, no. 163, pp. 1–55, 2023.

[92] G. Olikier, K. A. Gallivan, and P. .-. Absil, *First-order optimization on stratified sets*, 2023. arXiv: 2303.16040 `[math.OC]`. [Online]. Available: https://arxiv.org/abs/2303.16040.

[93] R. Schneider and A. Uschmajew, "Convergence results for projected line-search methods on varieties of low-rank matrices via łojasiewicz inequality," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 622–646, 2015. DOI: 10.1137/140957822. eprint: https://doi.org/10.1137/140957822. [Online]. Available: https://doi.org/10.1137/140957822.

A. Derivatives

See A.5 in [30] for more details in this section.

A.0.1 Fréchet Derivatives

For any two finite-dimensional inner product vector spaces $\mathcal{U}$ and $\mathcal{V}$ over $\mathbb{R}$, a mapping $F : \mathcal{U} \to \mathcal{V}$ is *Fréchet differentiable* at $x \in \mathcal{U}$ if there exists a linear operator

$$\begin{aligned} \mathrm{D}F(x) : \ \ \mathcal{U} &\to \mathcal{V} \\ h &\mapsto \mathrm{D}F(x)[h] \end{aligned}$$

such that

$$F(x+h) = F(x) + \mathrm{D}F(x)[h] + o(\|h\|).$$

The operator $\mathrm{D}F(x)$ is called the *Fréchet differential* and $\mathrm{D}F(x)[h]$ is called the *directional derivative* of F at x along h. The derivative satisfies the chain rule:

$$\mathrm{D}(f \circ g)(x)[h] = \mathrm{D}f(g(x))[\mathrm{D}g(x)[h]].$$

For a smooth real-valued function $f : \mathcal{U} \to \mathbb{R}$, the *Fréchet gradient* of f at x, denoted by $\nabla f(x)$, is the unique element in $\mathcal{U}$ satisfying

$$\langle \nabla f(x), h \rangle_{\mathcal{U}} = \mathrm{D}f(x)[h], \quad \forall h \in \mathcal{U},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ is the inner product in $\mathcal{U}$.

In particular, regard $\mathcal{U} = \mathbb{C}^{n \times n}$ as a vector space over $\mathbb{R}$ with the standard inner product $\langle X, Y \rangle_{\mathbb{C}^{n\times n}} = \mathrm{Re}(\mathrm{tr}(X^*Y))$. Regard X as $(\mathrm{Re}(X), \mathrm{Im}(X))$ and regard $f(X)$ as $f(\mathrm{Re}(X), \mathrm{Im}(X))$. By the multivariate Taylor theorem of the function $f(\mathrm{Re}(X), \mathrm{Im}(X))$, we get

$$\begin{aligned} &|f(X+h) - f(X) - \langle \nabla f(X), h \rangle_{\mathbb{C}^{n\times n}}| = \\ &|f\left(\mathrm{Re}(X) + \mathrm{Re}(h), \mathrm{Im}(X) + \mathrm{Im}(h)\right) - f(\mathrm{Re}(X), \mathrm{Im}(X)) - \end{aligned}$$

$$\left(\left\langle \frac{\partial f}{\partial \operatorname{Re}(X)}, \operatorname{Re}(h) \right\rangle_{\mathbb{R}^{n\times n}} + \left\langle \frac{\partial f}{\partial \operatorname{Im}(X)}, \operatorname{Im}(h) \right\rangle_{\mathbb{R}^{n\times n}}\right)\Bigg| \\ = o(\|h\|_{\mathbb{C}^{n\times n}}).$$

Notice

$$\left\langle \frac{\partial f}{\partial \operatorname{Re}(X)}, \operatorname{Re}(h) \right\rangle_{\mathbb{R}^{n\times n}} + \left\langle \frac{\partial f}{\partial \operatorname{Im}(X)}, \operatorname{Im}(h) \right\rangle_{\mathbb{R}^{n\times n}} = \left\langle \frac{\partial f(X)}{\partial \operatorname{Re}_X} + \mathbb{i}\frac{\partial f(X)}{\partial \operatorname{Im}_X}, h \right\rangle_{\mathbb{C}^{n\times n}}.$$

Thus the expression

$$\nabla f(X) = \frac{\partial f(X)}{\partial \operatorname{Re}_X} + \mathbb{i}\frac{\partial f(X)}{\partial \operatorname{Im}_X}$$

coincides with the Fréchet gradient for $f(X)$ under the real inner product (2.17).

Proposition A.0.1. Regard $\mathcal{U} = \mathbb{C}^{n\times n}$ as a vector space over $\mathbb{R}$ with the standard inner product $\langle X, Y\rangle_{\mathbb{C}^{n\times n}} = \operatorname{Re}(\operatorname{tr}(X^*Y))$. Let $X \in \mathbb{C}^{n\times n}$. If $X = X^*$, then $\nabla f(X) = (\nabla f(X))^*$.

Proof. Let $g : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n} : X \mapsto X^*$. Then it is straightforward to verify that

$$\operatorname{D} g(X)[h] = h^*, \forall h \in \mathbb{C}^{n\times n}.$$

Therefore for any $f : \mathbb{C}^{n\times n} \to \mathbb{R}$, by chain rule we have $\forall h \in \mathbb{C}^{n\times n}$

$$\begin{aligned}
\operatorname{D}(f\circ g)(X)[h] &= \operatorname{D} f(g(X))[\operatorname{D} g(X)[h]] \\
&= \operatorname{D} f(X^*)[h^*] \\
&= \langle \nabla f(X^*), h^*\rangle_{\mathbb{C}^{n\times n}} \\
&= \langle (\nabla f(X^*))^*, h\rangle_{\mathbb{C}^{n\times n}}.
\end{aligned}$$

Therefore we have

$$\nabla(f\circ g)(X) = (\nabla f(X^*))^*.$$

So by the definition of Fréchet derivative of $f \circ g$ at $X \in \mathbb{C}^{n\times n}$, we have the following.

$$(f\circ g)(X+h) = (f\circ g)(X) + \langle (\nabla f(X^*))^*, h\rangle_{\mathbb{C}^{n\times n}} + o(\|h\|), \quad \forall h \in \mathbb{C}^{n\times n}.$$

Let $\mathcal{H}^{n\times n} = \{X \in \mathbb{C}^{n\times n} : X^* = X\}$. Then $\mathcal{H}^{n\times n}$ is a linear subspace of the vector space $\mathbb{C}^{n\times n}$. Hence we can restrict f to $\mathcal{H}^{n\times n}$ and define its Fréchet gradient in $\mathcal{H}^{n\times n}$. Let $\nabla^{\mathcal{H}} f$ denote the Fréchet gradient of f in $\mathcal{H}^{n\times n}$. In particular, consider $X, h \in \mathcal{H}^{n\times n}$, then the above equality turns into

$$f(X+h) = f(X) + \langle (\nabla f(X))^*, h\rangle_{\mathbb{C}^{n\times n}} + o(\|h\|), \quad \forall h \in \mathcal{H}^{n\times n}.$$

Hence we have

$$\forall X \in \mathcal{H}^{n\times n}, \nabla^{\mathcal{H}} f(X) = (\nabla f(X))^*. \tag{A.1}$$

On the other hand, by the definition of Fréchet derivative of f, we have

$$f(X+h) = f(X) + \langle \nabla f(X), h\rangle_{\mathbb{C}^{n\times n}} + o(\|h\|), \quad \forall h \in \mathbb{C}^{n\times n}.$$

In particular consider $X, h \in \mathcal{H}^{n\times n}$, then the above equality turns into

$$f(X+h) = f(X) + \langle \nabla f(X), h\rangle_{\mathbb{C}^{n\times n}} + o(\|h\|), \quad \forall h \in \mathcal{H}^{n\times n}.$$

This gives us

$$\nabla^{\mathcal{H}} f(X) = \nabla f(X). \tag{A.2}$$

Combining (A.1) and (A.2), we obtain the desired result.

Proposition A.0.2. Let $\beta : \mathbb{C}^{n\times p} \to \mathbb{C}^{n\times n} : Y \mapsto YY^*$ and the inner product on $\mathbb{C}^{n\times p}$ as the standard inner product $\langle A, B\rangle_{\mathbb{C}^{n\times p}} = \mathrm{Re}(\mathrm{tr}(A^*B))$. Then the Fréchet gradient of $F := f \circ \beta$ satisfies

$$\nabla F(Y) = 2\nabla f(YY^*)Y. \tag{A.3}$$

Proof. Indeed, by the chain rule of Fréchet derivative we have

$$\mathrm{D}\, F(Y)[h] = \mathrm{D}\, f(\beta(Y))\left[\mathrm{D}\, \beta(Y)[h]\right], \quad \forall h \in \mathbb{C}^{n\times p}.$$

Hence

$$\langle \nabla F(Y), h\rangle_{\mathbb{C}^{n\times p}} = \langle \nabla f(YY^*), \mathrm{D}\, \beta(Y)[h]\rangle_{\mathbb{C}^{n\times n}}.$$

One can check by definition that $\mathrm{D}\,\beta(Y)[h] = Yh^* + hY^*$. Hence

$$\langle \nabla F(Y), h\rangle_{\mathbb{C}^{n\times p}} = \langle \nabla f(YY^*), Yh^* + hY^*\rangle_{\mathbb{C}^{n\times n}} = \langle 2\nabla f(YY^*)Y, h\rangle_{\mathbb{C}^{n\times p}}.$$

This proves (A.3).

Proposition A.0.3. If f takes the form of $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_F^2$ for a linear operator $\mathcal{A}$, then the Fréchet gradient of $f(X)$ is given by

$$\nabla f(X) = \mathcal{A}^*(\mathcal{A}(X) - b),$$

where $\mathcal{A}^*$ is the conjugate operator of $\mathcal{A}$.

Proof. We know by the definition of Fréchet gradient that

$$\nabla f(X) = \frac{\partial f}{\partial \operatorname{Re}_X} + \mathbb{i}\frac{\partial f}{\partial \operatorname{Im}_X},$$

Now for $f(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|^2 = \frac{1}{2}\langle \mathcal{A}(X) - b, \mathcal{A}(X) - b\rangle$, by the linearity of $\mathcal{A}$, we have

$$\begin{aligned}
\nabla f(X) &= \frac{1}{2}\frac{\partial}{\partial X}\langle \mathcal{A}(X) - b,\ \mathcal{A}(\Delta) - b\rangle|_{\Delta=X} + \frac{1}{2}\frac{\partial}{\partial \Delta}\langle \mathcal{A}(X) - b,\ \mathcal{A}(\Delta) - b\rangle|_{\Delta=X} \\
&= \frac{1}{2}\frac{\partial}{\partial X}\langle \mathcal{A}(X) - b,\ \mathcal{A}(\Delta) - b\rangle_{\mathbb{C}^{n\times n}}|_{\Delta=X} + \frac{1}{2}\frac{\partial}{\partial \Delta}\langle \mathcal{A}(\Delta) - b,\ \mathcal{A}(X) - b\rangle_{\mathbb{C}^{n\times n}}|_{\Delta=X} \\
&= \frac{1}{2}\frac{\partial}{\partial X}\langle X,\ \mathcal{A}^*(\mathcal{A}(\Delta) - b)\rangle_{\mathbb{C}^{n\times n}}|_{\Delta=X} + \frac{1}{2}\frac{\partial}{\partial \Delta}\langle \Delta,\ \mathcal{A}^*(\mathcal{A}(X) - b)\rangle_{\mathbb{C}^{n\times n}}|_{\Delta=X} \\
&= \frac{1}{2}\frac{\partial}{\partial X}\left(\langle \operatorname{Re}(X),\ \operatorname{Re}(\mathcal{A}^*(\mathcal{A}(\Delta) - b))\rangle + \langle \operatorname{Im}(X),\ \operatorname{Im}(\mathcal{A}^*(\mathcal{A}(\Delta) - b))\rangle\right)|_{\Delta=X} \\
&\quad + \frac{1}{2}\frac{\partial}{\partial \Delta}\left(\langle \operatorname{Re}(\Delta),\ \operatorname{Re}(\mathcal{A}^*(\mathcal{A}(X) - b))\rangle + \langle \operatorname{Im}(\Delta),\ \operatorname{Im}(\mathcal{A}^*(\mathcal{A}(X) - b))\rangle\right)|_{\Delta=X} \\
&= \frac{1}{2}\left(\frac{\partial}{\partial \operatorname{Re}(X)} + \mathbb{i}\frac{\partial}{\partial \operatorname{Im}(X)}\right)\left(\langle \operatorname{Re}(X),\ \operatorname{Re}(\mathcal{A}^*(\mathcal{A}(\Delta) - b))\rangle\right. \\
&\quad \left. + \langle \operatorname{Im}(X),\ \operatorname{Im}(\mathcal{A}^*(\mathcal{A}(\Delta) - b))\rangle\right)|_{\Delta=X} \\
&\quad + \frac{1}{2}\left(\frac{\partial}{\partial \operatorname{Re}(\Delta)} + \mathbb{i}\frac{\partial}{\partial \operatorname{Im}(\Delta)}\right)\left(\langle \operatorname{Re}(\Delta),\ \operatorname{Re}(\mathcal{A}^*(\mathcal{A}(X) - b))\rangle\right. \\
&\quad \left. + \langle \operatorname{Im}(\Delta),\ \operatorname{Im}(\mathcal{A}^*(\mathcal{A}(X) - b))\rangle\right)|_{\Delta=X} \\
&= \frac{1}{2}\left(\operatorname{Re}(\mathcal{A}^*(\mathcal{A}(\Delta) - b)) + \mathbb{i}\operatorname{Im}(\mathcal{A}^*(\mathcal{A}(\Delta) - b))\right)|_{\Delta=X}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\left(\mathrm{Re}(\mathcal{A}^*(\mathcal{A}(X)-b)) + \mathbb{i}\,\mathrm{Im}(\mathcal{A}^*(\mathcal{A}(X)-b))\right)|_{\Delta=X} \\
= \; & \mathcal{A}^*(\mathcal{A}(X)-b).
\end{aligned}
$$

A.0.2 Fréchet Hessian

For a Euclidean space $\mathcal{E}$ and a twice-differentiable, real-valued function f on $\mathcal{E}$, the *Fréchet Hessian operator* of f at x is the unique symmetric operator $\nabla^2 f(x) : \mathcal{E} \to \mathcal{E}$ defined by

$$\nabla^2 f(x)[h] = \mathrm{D}\,(\nabla f)(x)[h]$$

for all $h \in \mathcal{E}$.

Proposition A.0.4. Regard $\mathcal{E} = \mathbb{C}^{n\times n}$ as a Euclidean space over $\mathbb{R}$ with the standard inner product $\langle X, Y\rangle_{\mathbb{C}^{n\times n}} = \mathrm{Re}(\mathrm{tr}(X^*Y))$. If $X = X^*$ and $h = h^*$, then

$$\nabla^2 f(X)[h] = (\nabla^2 f(X)[h])^*.$$

Proof. Let $g : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n} : X \mapsto X^*$. Consider the Fréchet Hessian of $f \circ g$. By the definition of Fréchet Hessian we have

$$\nabla(f\circ g)(X+h) = \nabla(f\circ g)(X) + \nabla^2(f\circ g)(X)[h] + o(\|h\|^2).$$

We know from the proof of Proposition A.0.1 that

$$\nabla(f\circ g)(X) = (\nabla f(X^*))^*.$$

Hence

$$(\nabla f(X^* + h^*))^* = (\nabla f(X^*))^* + \nabla^2(f\circ g)(X)[h] + o(\|h\|^2).$$

Therefore

$$\nabla^2(f\circ g)(X)[h] = (\nabla^2 f(X^*)[h^*])^*.$$

Now restrict $f \circ g$ on the subspace $\mathcal{H}^{n\times n}$, we have $f \circ g|_{\mathcal{H}^{n\times n}} = f|_{\mathcal{H}^{n\times n}}$. Hence the Fréchet Hessian operator of f on $\mathcal{H}^{n\times n}$ is $(\nabla^2 f(X^*)[h^*])^* = (\nabla^2 f(X)[h])^*$. On the other hand, the Fréchet Hessian operator of f on $\mathcal{H}^{n\times n}$ is denoted as $\nabla^2 f(X)[h]$. Hence if $X, h \in \mathcal{H}^{n\times n}$, we have

$$\nabla^2 f(X)[h] = (\nabla^2 f(X)[h])^*.$$

This proves the proposition.

A.0.3 Taylor's Formula

Let $\mathcal{E}$ be finite-dimensional Euclidean space. Let f be a twice-differentiable real-valued function on an open convex domain $\Omega \subset \mathcal{E}$. Then for all x and $x + h \in \Omega$,

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle_{\mathcal{E}} + \frac{1}{2} \left\langle \nabla^2 f(x)[h], h \right\rangle_{\mathcal{E}} + O\left(\|h\|_{\mathcal{E}}^3 \right).$$

B. Embedded manifold $\mathcal{H}_+^{n,p}$

The geometry of the real case, i.e., $\mathcal{S}_+^{n,p}$ has been explored in [43]. However, it is not straightforward to extend these results directly to the complex case. Although the methods of proofs of the complex case turn out to be similar to the real case, we still need to provide them. Recall that a complex matrix manifold is viewed as a manifold over $\mathbb{R}$ instead of $\mathbb{C}$. One way is to identify a complex matrix with the pair of its real and imaginary part; another way is to identify the matrix with its *realification*.

Definition B.0.1 (Realification). The realification is an injective mapping $\mathcal{R} : \mathbb{C}^{n\times n} \to \mathbb{R}^{2n\times 2n}$ defined by replacing each entry a_{ij} of $A = (a_{\mathrm{ij}})_{n\times n} \in \mathbb{C}^{n\times n}$ by the 2×2 matrix

$$\begin{bmatrix} \mathrm{Re}(a_{\mathrm{ij}}) & -\mathrm{Im}(a_{\mathrm{ij}}) \\ \mathrm{Im}(a_{\mathrm{ij}}) & \mathrm{Re}(a_{\mathrm{ij}}) \end{bmatrix}.$$

It can be shown that $\mathcal{R}$ preserves the algebraic structure:

- $\mathcal{R}(A+B) = \mathcal{R}(A) + \mathcal{R}(B)$
- $\mathcal{R}(AB) = \mathcal{R}(A)\mathcal{R}(B)$
- $\mathcal{R}(aA) = a\mathcal{R}(A) \quad \forall a \in \mathbb{R}$
- $\mathcal{R}(I) = I$
- $\mathcal{R}(A^*) = (\mathcal{R}(A))^T$

Hence $A \in \mathbb{C}^{n\times n}$ is invertible if and only if $\mathcal{R}(A)$ is invertible. [1]

Lemma B.0.1. Let $\mathrm{GL}(n,\mathbb{C})$ be the general linear group viewed as a real Lie group. Then it is a semialgebraic set.

Proof. Recall that a subset of $\mathbb{R}^m$ is a *semialgebraic* set if it can be obtained by finitely many intersections, union and set differences starting from sets of the from $\{x \in \mathbb{R}^m : P(x) > 0\}$

[1] ↑. See for example https://www.maths.tcd.ie/pub/coursework/424/LieGroups.pdf

with P a polynomial on $\mathbb{R}^m$ [28, Appendix B]. Since $\mathrm{GL}(n,\mathbb{C})$ is viewed as a real Lie group, $\mathrm{GL}(n,\mathbb{C})$ is understood as a subset of $\mathrm{GL}(2n,\mathbb{R})$ through realification. It can be shown that

$$\mathrm{GL}(n,\mathbb{C}) = \{X \in \mathrm{GL}(2n,\mathbb{R}) : XJ = JX\}, \quad \text{with } J = \mathcal{R}(\mathbb{i}I).$$

We know that $\mathrm{GL}(2n,\mathbb{R})$ is a semialgebraic set since it is the non-vanishing points of determinant; and $\{X \in \mathbb{R}^{2n\times 2n} : XJ = JX\}$ is also a semialgebraic set by definition. Hence $\mathrm{GL}(n,\mathbb{C})$ is a semialgebraic set.

B.0.1 Calculations for the Riemannian Hessian

Let f be a smooth real-valued function on $\mathcal{H}_+^{n,p}$. In this section, we derive the Riemannian Hessian operator of f.

By [39, section 4] we know that the retraction R defined in (2.20) is a second-order retraction.

Proposition 5.5.5 in [30] states that if R is a second-order retraction, then the Riemannian Hessian of f can be computed in the following nice way:

$$\operatorname{Hess} f(X) = \operatorname{Hess}\,(f \circ R_X)(0_X).$$

Notice that now $f \circ R_X$ is a smooth function defined on a vector space. Hence, we obtain

$$g_X\left(\operatorname{Hess} f(X)[\xi_X], \xi_X\right) = \frac{d^2}{dt^2} f(R_X(t\xi_X))|_{t=0}.$$

However, it is difficult to obtain a second-order derivative of $f \circ R_X$ using the retraction R_X defined in (2.20). The references [7] and [41] proposed a method to compute $\operatorname{Hess} f(X)$ by constructing a second-order retraction $R^{(2)}$ that has a second-order series expansion which makes it simple to derive a series expansion of $f \circ R_X^{(2)}$ up to second order and thus obtain the Hessian of f. We will summarize the derivation below.

Lemma B.0.2. For any $X \in \mathcal{H}_+^{n,p}$ with $X^\dagger$ the pseudoinverse, the mapping $R_X^{(2)} : T_X\mathcal{H}_+^{n,p} \to \mathcal{H}_+^{n,p}$ given by

$$\xi_X \mapsto wX^\dagger w^*, \text{ with } w = X + \frac{1}{2}\xi_X^s + \xi_X^p - \frac{1}{8}\xi_X^s X^\dagger \xi_X^s - \frac{1}{2}\xi_X^p X^\dagger \xi_X^s,$$

is a second-order retraction on $\mathcal{H}_+^{n,p}$, where $\xi_X^s = P_X^s(\xi_X)$ and $\xi_X^p = P_X^p(\xi_X)$ as defined in (2.7). Moreover, we have

$$R_X^{(2)}(\xi_X) = X + \xi_X + \xi_X^p X^\dagger \xi_X^p + O(\|\xi_X\|^3).$$

Proof. It follows the same proof of [7, Proposition 5.10] .

From this, the Riemannian Hessian operator of f can be computed in essentially the same way as in [38, Section A.2] but applied to the general cost function $f(X)$ instead of a least squared cost function. Consider the Taylor expansion of $\hat{f}_X^{(2)} := f \circ R_X^{(2)}$, which is a real-valued function on a vector space. We get

$$\begin{aligned}
\hat{f}_X^{(2)}(\xi_X) &= f(R_X^{(2)}(\xi_X)) \\
&= f\left(X + \xi_X + \xi_X^p X^\dagger \xi_X^p + O(\|\xi_X\|^3)\right) \\
&= f(X) + \left\langle \nabla f(X), \xi_X + \xi_X^p X^\dagger \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}} \\
&\quad + \frac{1}{2}\left\langle \nabla^2 f(X)[\xi_X + \xi_X^p X^\dagger \xi_X^p], \xi_X + \xi_X^p X^\dagger \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}} + O(\|\xi_X\|^3) \\
&= f(X) + \langle \nabla f(X), \xi_X \rangle_{\mathbb{C}^{n\times n}} + \left\langle \nabla f(X), \xi_X^p X^\dagger \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}} \\
&\quad + \frac{1}{2}\left\langle \nabla^2 f(X)[\xi_X], \xi_X \right\rangle_{\mathbb{C}^{n\times n}} + O(\|\xi_X\|^3).
\end{aligned}$$

We can immediately recognize the first order term and the second order term that contribute to the Riemannian gradient and Hessian, respectively. That is,

$$g_X\left(\operatorname{grad} f(X), \xi_X\right) = \langle \nabla f(X), \xi_X \rangle_{\mathbb{C}^{n\times n}},$$

$$g_X\left(\operatorname{Hess} f(X)[\xi_X], \xi_X\right) = \underbrace{2\left\langle \nabla f(X), \xi_X^p X^\dagger \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}}}_{f_1 := \langle \mathcal{H}_1(\xi_X), \xi_X \rangle_{\mathbb{C}^{n\times n}}} + \underbrace{\left\langle \nabla^2 f(X)[\xi_X], \xi_X \right\rangle_{\mathbb{C}^{n\times n}}}_{f_2 := \langle \mathcal{H}_2(\xi_X), \xi_X \rangle_{\mathbb{C}^{n\times n}}}.$$

The first equation immediately gives us

$$\operatorname{grad} f(X) = P_X^t(\nabla f(X)).$$

For the second equation, the inner product of the Riemannian Hessian consists of the sum of f_1 and f_2; and the Riemannian Hessian operator is the sum of two operators $\mathcal{H}_1$ and $\mathcal{H}_2$. Since ξ_X is already separated in f_2, the contribution to the Riemannian Hessian from $\mathcal{H}_2$ is readily given by

$$\mathcal{H}_2(\xi_X) = P_X^t(\nabla^2 f(X)[\xi_X]).$$

Now, we still need to separate ξ_X in f_1 to see the contribution to Riemannian Hessian from $\mathcal{H}_1$. Since we can choose to bring over $\xi_X^p X^\dagger$ or $X^\dagger \xi_X^p$ to the first position of $\langle .,. \rangle_{\mathbb{C}^{n\times n}}$, we write $\mathcal{H}_1(\xi_X)$ as the linear combination of both:

$$f_1 = 2c\left\langle \nabla f(X)(X^\dagger \xi_X^p)^*, \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}} + 2(1-c)\left\langle (\xi_X^p X^\dagger)^* \nabla f(X), \xi_X^p \right\rangle_{\mathbb{C}^{n\times n}}.$$

Operator $\mathcal{H}_1$ is clearly linear. Since $\mathcal{H}_1$ is symmetric, we must have $\langle \mathcal{H}_1(\xi_X), \nu_X \rangle_{\mathbb{C}^{n\times n}} = \langle \nu_X, \mathcal{H}_1(\xi_X) \rangle_{\mathbb{C}^{n\times n}}$ for all tangent vector ν_X. Hence we must have $c = \frac{1}{2}$ and we obtain

$$\mathcal{H}_1(\xi_X) = P_X^p\left(\nabla f(X)(X^\dagger \xi_X^p)^* + (\xi_X^p X^\dagger)^* \nabla f(X)\right).$$

Putting $\mathcal{H}_1$ and $\mathcal{H}_2$ together, we obtain

$$\operatorname{Hess} f(X)[\xi_X] = P_X^t(\nabla^2 f(X)[\xi_X]) + P_X^p\left(\nabla f(X)(X^\dagger \xi_X^p)^* + (\xi_X^p X^\dagger)^* \nabla f(X)\right).$$

C. Quotient Manifold $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$

C.0.1 Calculations for the Riemannian Hessian

In this section, we outline the computations of the Riemannian Hessian operators of the cost function h defined on $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ under the three different metrics g^{i}.

Definition C.0.1. [30, Definition 5.5.1] Given a real-valued function f on a Riemannian manifold $\mathcal{M}$, the Riemannian Hessian of f at a point x in $\mathcal{M}$ is the linear mapping $\operatorname{Hess} f(x)$ of $T_x\mathcal{M}$ into itself defined by

$$\operatorname{Hess} f(x)[\xi_x] = \nabla_{\xi_x} \operatorname{grad} f(x)$$

for all ξ_x in $T_x\mathcal{M}$, where ∇ is the Riemannian connection on $\mathcal{M}$.

Lemma C.0.1. The Riemannian Hessian of $h : \mathbb{C}_*^{n\times p}/\mathcal{O}_p \mapsto \mathbb{R}$ is related to the Riemannian Hessian of $F : \mathbb{C}_*^{n\times p} \mapsto \mathbb{R}$ in the following way:

$$\overline{\left(\operatorname{Hess} h(\pi(Y))[\xi_{\pi(Y)}]\right)}_Y = P_Y^{\mathcal{H}}\left(\operatorname{Hess} F(Y)[\overline{\xi}_Y]\right),$$

where $\overline{\xi}_Y$ is the horizontal lift of $\xi_{\pi(Y)}$ at Y.

Proof. The result follows from [30, Proposition 5.3.3] and the definition of the Riemannian Hessian.

Riemannian Hessian for the Metric g^1

Using the Riemannian metric g^1, $\mathbb{C}_*^{n\times p}$ is a Riemannian submanifold of a Euclidean space. By [30, Proposition 5.3.2], the Riemannian connection on $\mathbb{C}_*^{n\times p}$ is the classical directional derivative:

$$\nabla_{\eta_Y}\xi = \mathrm{D}\,\xi(Y)[\eta_Y].$$

Recall that for g^1, $\operatorname{grad} F(Y) = 2\nabla f(YY^*)Y$. Hence, the Riemannian Hessian of F at Y is given by

$$\operatorname{Hess} F(Y)[\xi_Y] \quad = \quad \nabla_{\xi_Y} \operatorname{grad} F$$

$$\begin{aligned} &= \mathrm{D}\,\mathrm{grad}\, F(Y)[\xi_Y] \\ &= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y + 2\nabla f(YY^*)\xi_Y. \end{aligned}$$

The last line is by the product rule and the chain rule of differentiation. Therefore, we obtain

$$\overline{\left(\mathrm{Hess}\, h(\pi(Y))[\xi_{\pi(Y)}]\right)}_Y = P_Y^{\mathcal{H}^1}\left(2\nabla^2 f(YY^*)[Y\bar{\xi}_Y^* + \bar{\xi}_Y Y^*]Y + 2\nabla f(YY^*)\bar{\xi}_Y\right).$$

Riemannian Hessian Under Metric g^2

First, for any Riemannian metric g, g satisfies the Koszul formula

$$\begin{aligned} 2g_x(\nabla_{\xi_x}\lambda, \eta_x) &= \xi_x g(\lambda,\eta) + \lambda_x g(\eta,\xi) - \eta_x g(\xi,\lambda) \\ &\quad - g_x(\xi_x, [\lambda,\eta]_x) + g_x(\lambda_x, [\eta,\xi]_x) + g_x(\eta, [\xi,\lambda]_x) \\ &= \mathrm{D}\, g(\lambda,\eta)(x)[\xi_x] + \mathrm{D}\, g(\eta,\xi)(x)[\lambda_x] - \mathrm{D}\, g(\xi,\lambda)(x)[\eta_x] \\ &\quad - g_x(\xi_x, [\lambda,\eta]_x) + g_x(\lambda_x, [\eta,\xi]_x) + g_x(\eta, [\xi,\lambda]_x), \end{aligned}$$

where the *Lie bracket* $[\cdot,\cdot]$ is defined in [30].

In particular, for g^2 the above Koszul formula turns into

$$\begin{aligned} 2g_Y^2(\nabla_{\xi_Y}\lambda, \eta_Y) &= \mathrm{D}\, g^2(\lambda,\eta)(Y)[\xi_Y] + \mathrm{D}\, g^2(\eta,\xi)(Y)[\lambda_Y] - \mathrm{D}\, g^2(\xi,\lambda)(Y)[\eta_Y] \\ &\quad - g_Y^2(\xi_Y, [\lambda,\eta]_Y) + g_Y^2(\lambda_Y, [\eta,\xi]_Y) + g_Y^2(\eta, [\xi,\lambda]_Y). \end{aligned}$$

Recall that $g^2(\lambda,\eta)(Y) = \mathrm{Re}(\mathrm{tr}(Y^*Y\lambda_Y^*\eta_Y))$. Hence, the first term in the above sum equals

$$\begin{aligned} \mathrm{D}\, g^2(\lambda,\eta)(Y)[\xi_Y] &= g_Y^2(\mathrm{D}\,\lambda(Y)[\xi_Y], \eta_Y) + g_Y^2(\lambda_Y, \mathrm{D}\,\eta(Y)[\xi_Y]) \\ &\quad + \mathrm{Re}(\mathrm{tr}(\xi_Y^* Y\lambda_Y^*\eta_Y)) + \mathrm{Re}(\mathrm{tr}(Y^*\xi_Y\lambda_Y^*\eta_Y)). \end{aligned}$$

Following [30, Section 5.3.4], since $\mathbb{C}_*^{n\times p}$ is an open subset of $\mathbb{C}^{n\times p}$, we also have

$$[\lambda,\eta]_Y = \mathrm{D}\,\eta(Y)[\lambda_Y] - \mathrm{D}\,\lambda(Y)[\eta_Y].$$

Summarizing, we get

$$\begin{aligned}
2g^2_Y(\nabla_{\xi_Y}\lambda, \eta_Y) &= \mathrm{D}\, g^2(\lambda,\eta)(Y)[\xi_Y] + \mathrm{D}\, g^2(\eta,\xi)(Y)[\lambda_Y] - \mathrm{D}\, g^2(\xi,\lambda)(Y)[\eta_Y] \\
&\quad -g^2(\xi_Y, \mathrm{D}\,\eta(Y)[\lambda_Y] - \mathrm{D}\,\lambda(Y)[\eta_Y]) \\
&\quad +g^2(\lambda_Y, \mathrm{D}\,\xi(Y)[\eta_Y] - \mathrm{D}\,\eta(Y)[\xi_Y]) \\
&\quad +g^2(\eta_Y, \mathrm{D}\,\lambda(Y)[\xi_Y] - \mathrm{D}\,\xi(Y)[\lambda_Y]) \\
&= 2g^2_Y(\eta_Y, \mathrm{D}\,\lambda(Y)[\xi_Y]) \\
&\quad + \mathrm{Re}(\mathrm{tr}(\eta_Y^*(\lambda_Y(\xi_Y^*Y + Y^*\xi_Y) + \xi_Y(Y^*\lambda_Y + \lambda_Y^*Y) - Y\lambda_Y^*\xi_Y - Y\xi_Y^*\lambda_Y))) \\
&= 2g^2_Y(\eta_Y, \mathrm{D}\,\lambda(Y)[\xi_Y]) \\
&\quad +g^2_Y(\eta_Y, (\lambda_Y(\xi_Y^*Y + Y^*\xi_Y) + \\
&\quad \xi_Y(Y^*\lambda_Y + \lambda_Y^*Y) - Y\lambda_Y^*\xi_Y - Y\xi_Y^*\lambda_Y)(Y^*Y)^{-1}).
\end{aligned}$$

We therefore obtain a closed-form expression for Riemannian connection on $\mathbb{C}_*^{n\times p}$ for g^2:

$$\nabla_{\xi_Y}\lambda = \mathrm{D}\,\lambda(Y)[\xi_Y] + \frac{1}{2}\left(\lambda_Y(\xi_Y^*Y + Y^*\xi_Y) + \xi_Y(Y^*\lambda_Y + \lambda_Y^*Y) - Y\lambda_Y^*\xi_Y - Y\xi_Y^*\lambda_Y\right)(Y^*Y)^{-1}.$$

Recall that for the Riemannian metric g^2, we have $\mathrm{grad}\, F(Y) = 2\nabla f(YY^*)Y(Y^*Y)^{-1}$. Hence, we have

$$\begin{aligned}
\mathrm{Hess}\, F(Y)[\xi_Y] &= \nabla_{\xi_Y}\mathrm{grad}\, F \\
&= \mathrm{D}_Y\mathrm{grad}\, F(Y)[\xi_Y] \\
&\quad +\frac{1}{2}\{\mathrm{grad}\, F(Y)(\xi_Y^*Y + Y^*\xi_Y) + \xi_Y(Y^*\mathrm{grad}\, F(Y) + \mathrm{grad}\, F(Y)^*Y) - \\
&\quad Y\mathrm{grad}\, F(Y)^*\xi_Y - Y\xi_Y^*\mathrm{grad}\, F(Y)\}(Y^*Y)^{-1} \\
&= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} + 2\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} \\
&\quad -2\nabla f(YY^*)Y(Y^*Y)^{-1}(Y^*\xi_Y + \xi_Y^*Y)(Y^*Y)^{-1} \\
&\quad +\nabla f(YY^*)Y(Y^*Y)^{-1}(Y^*\xi_Y + \xi_Y^*Y)(Y^*Y)^{-1} \\
&\quad +\xi_Y\{Y^*\nabla f(YY^*)Y(Y^*Y)^{-1} + (Y^*Y)^{-1}Y^*\nabla f(YY^*)Y\}(Y^*Y)^{-1} \\
&\quad -\{Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\xi_Y + Y\xi_Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\}(Y^*Y)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} + 2\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} \\
&\quad -\nabla f(YY^*)Y(Y^*Y)^{-1}(Y^*\xi_Y + \xi_Y^* Y)(Y^*Y)^{-1} \\
&\quad +\xi_Y\{Y^*\nabla f(YY^*)Y(Y^*Y)^{-1} + (Y^*Y)^{-1}Y^*\nabla f(YY^*)Y\}(Y^*Y)^{-1} \\
&\quad -\{Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\xi_Y + Y\xi_Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\}(Y^*Y)^{-1} \\
&= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} + 2\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} \\
&\quad -\nabla f(YY^*)P_Y\xi_Y(Y^*Y)^{-1} - \nabla f(YY^*)Y(Y^*Y)^{-1}\xi_Y^* Y(Y^*Y)^{-1} \\
&\quad +\xi_Y Y^*\nabla f(YY^*)Y(Y^*Y)^{-2} + \xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1} \\
&\quad -P_Y\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} - Y\xi_Y^*\nabla f(YY^*)Y(Y^*Y)^{-2} \\
&= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} \\
&\quad +\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} - \nabla f(YY^*)P_Y\xi_Y(Y^*Y)^{-1} \\
&\quad +\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} - P_Y\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} \\
&\quad +2skew(\xi_Y Y^*)\nabla f(YY^*)Y(Y^*Y)^{-2} \\
&\quad +2skew\{\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\}Y(Y^*Y)^{-1} \\
&= 2\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} \\
&\quad +\nabla f(YY^*)P_Y^{\perp}\xi_Y(Y^*Y)^{-1} + P_Y^{\perp}\nabla f(YY^*)\xi_Y(Y^*Y)^{-1} \\
&\quad +2skew(\xi_Y Y^*)\nabla f(YY^*)Y(Y^*Y)^{-2} \\
&\quad +2skew\{\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\}Y(Y^*Y)^{-1}.
\end{aligned}
$$

To conclude, we obtain

$$
\begin{aligned}
\overline{\left(\mathrm{Hess}\, h(\pi(Y))[\eta_{\pi(Y)}]\right)}_Y &= P_Y^{\mathcal{H}^2}\Big\{2\nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*]Y(Y^*Y)^{-1} \\
&\quad +\nabla f(YY^*)P_Y^{\perp}\overline{\xi}_Y(Y^*Y)^{-1} + P_Y^{\perp}\nabla f(YY^*)\overline{\xi}_Y(Y^*Y)^{-1} \\
&\quad +2skew(\overline{\xi}_Y Y^*)\nabla f(YY^*)Y(Y^*Y)^{-2} \\
&\quad + 2skew\{\overline{\xi}_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\}Y(Y^*Y)^{-1}\Big\}.
\end{aligned}
$$

Riemannian Hessian Under Metric g^3

Recall that the Riemannian metric g^3 on $\mathbb{C}_*^{n\times p}$ satisfies

$$\begin{aligned} g^3_Y(\xi_Y,\eta_Y) &= \tilde{g}_Y(\xi_Y,\eta_Y)+g^2_Y(P^{\mathcal{V}}_Y(\xi_Y),P^{\mathcal{V}}_Y(\eta_Y)) \\ &= 2\,\mathrm{Re}(\mathrm{tr}(Y^*\xi_Y Y^*\eta_Y+Y^*Y\xi_Y^*\eta_Y))+\mathrm{Re}(\mathrm{tr}(YP^{\mathcal{V}}_Y(\xi_Y)^*P^{\mathcal{V}}_Y(\eta_Y)Y^*)) \end{aligned}$$

where

$$\tilde{g}_Y(\xi_Y,\eta_Y)=\langle Y\xi_Y^*+\xi_Y Y^*, Y\eta_Y^*+\eta_Y Y^*\rangle_{\mathbb{C}^{n\times n}}\,.$$

$$P^{\mathcal{V}}_Y(\lambda_Y)=Y skew((Y^*Y)^{-1}Y^*\lambda_Y).$$

Hence

$$\begin{aligned} &\mathrm{D}\,g^3(\lambda,\eta)(Y)[\xi_Y] \\ =\;& \tilde{g}_Y(\mathrm{D}\,\lambda(Y)[\xi_Y],\eta_Y)+\tilde{g}(\lambda_Y,D\eta(Y)[\xi_Y]) \\ &+2\,\mathrm{Re}(\mathrm{tr}(\xi_Y^*\lambda_Y Y^*\eta_Y+Y^*\lambda_Y\xi_Y^*\eta_Y+\xi_Y^*Y\lambda_Y^*\eta_Y+Y^*\xi_Y\lambda_Y^*\eta_Y)) \\ &+g^2_Y(P^{\mathcal{V}}_Y(\lambda_Y),DP^{\mathcal{V}}_Y(\eta_Y)[\xi_Y])+g^2(\mathrm{D}\,P^{\mathcal{V}}_Y(\lambda_Y)[\xi_Y],P^{\mathcal{V}}_Y(\eta_Y)) \\ &+\mathrm{Re}(\mathrm{tr}(\xi_Y P^{\mathcal{V}}_Y(\lambda_Y)^*P^{\mathcal{V}}_Y(\eta_Y)Y^*+YP^{\mathcal{V}}_Y(\lambda_Y)^*P^{\mathcal{V}}_Y(\eta_Y)\xi_Y^*)). \end{aligned}$$

Suppose λ,η and ξ are horizontal vector fields, then many terms in the above equation vanish:

$$\begin{aligned} \mathrm{D}\,g^3(\lambda,\eta)(Y)[\xi_Y] &= \tilde{g}_Y(\mathrm{D}\,\lambda(Y)[\xi_Y],\eta_Y)+\tilde{g}_Y(\lambda_Y,\mathrm{D}\,\eta_Y[\xi_Y]) \\ &\quad+2\,\mathrm{Re}(\mathrm{tr}(\xi_Y^*\lambda_Y Y^*\eta_Y+Y^*\lambda_Y\xi_Y^*\eta_Y+\xi_Y^*Y\lambda_Y^*\eta_Y+Y^*\xi_Y\lambda_Y^*\eta_Y)). \end{aligned}$$

Combining the above equation and the Koszul formula with ξ,η,λ horizontal vector fields, we obtain

$$\begin{aligned} &2g^3_Y(\nabla_{\xi_Y}\lambda,\eta_Y) \\ =\;& \mathrm{D}\,g^3(\lambda,\eta)(Y)[\xi_Y]+\mathrm{D}\,g^3(\eta,\xi)(Y)[\lambda_Y]-\mathrm{D}\,g^3(\xi,\lambda)(Y)[\eta_Y] \end{aligned}$$

$$
\begin{aligned}
& -g_Y^3(\xi_Y, \mathrm{D}\,\eta(Y)[\lambda_Y] - \mathrm{D}\,\lambda(Y)[\eta_Y]) \\
& +g_Y^3(\lambda_Y, \mathrm{D}\,\xi(Y)[\eta_Y] - \mathrm{D}\,\eta(Y)[\xi_Y]) \\
& +g_Y^3(\eta_Y, \mathrm{D}\,\lambda(Y)[\xi_Y] - \mathrm{D}\,\xi(Y)[\lambda_Y]) \\
= \;& \tilde{g}_Y(\mathrm{D}\,\lambda(Y)[\xi_Y], \eta_Y) + \tilde{g}_Y(\lambda_Y, \mathrm{D}\,\eta(Y)[\xi_Y]) \\
& +2\,\mathrm{Re}(\mathrm{tr}(\xi_Y^* \lambda_Y Y^* \eta_Y + Y^* \lambda_Y \xi_Y^* \eta_Y + \xi_Y^* Y \lambda_Y^* \eta_Y + Y^* \xi_Y \lambda_Y^* \eta_Y)) \\
& +\tilde{g}_Y(\mathrm{D}\,\eta(Y)[\lambda_Y], \xi_Y) + \tilde{g}_Y(\eta_Y, \mathrm{D}\,\xi(Y)[\lambda_Y]) \\
& +2\,\mathrm{Re}(\mathrm{tr}(\lambda_Y^* \eta_Y Y^* \xi_Y + Y^* \eta_Y \lambda_Y^* \xi_Y + \lambda_Y^* Y \eta_Y^* \xi_Y + Y^* \lambda_Y \eta_Y^* \xi_Y)) \\
& -\tilde{g}_Y(\mathrm{D}\,\xi(Y)[\eta_Y], \lambda_Y) - \tilde{g}_Y(\xi_Y, \mathrm{D}\,\lambda(Y)[\eta_Y]) \\
& -2\,\mathrm{Re}(\mathrm{tr}(\eta_Y^* \xi_Y Y^* \lambda_Y + Y^* \xi_Y \eta_Y^* \lambda_Y + \eta_Y^* Y \xi_Y^* \lambda_Y + Y^* \eta_Y \xi_Y^* \lambda_Y)) \\
& -\tilde{g}_Y(\xi_Y, \mathrm{D}\,\eta(Y)[\lambda_Y]) + \tilde{g}_Y(\xi_Y, \mathrm{D}\,\lambda(Y)[\eta_Y]) \\
& +\tilde{g}_Y(\lambda_Y, \mathrm{D}\,\xi(Y)[\eta_Y]) - \tilde{g}_Y(\lambda_Y, \mathrm{D}\,\eta(Y)[\xi_Y]) \\
& +\tilde{g}_Y(\eta_Y, \mathrm{D}\,\lambda(Y)[\xi_Y]) - \tilde{g}_Y(\eta_Y, \mathrm{D}\,\xi(Y)[\lambda_Y]) \\
= \;& 2\tilde{g}_Y(\mathrm{D}\,\lambda(Y)[\xi_Y], \eta_Y) + 4\,\mathrm{Re}(\mathrm{tr}(Y^* \xi_Y \lambda_Y^* \eta_Y + Y^* \lambda_Y \xi_Y^* \eta_Y)).
\end{aligned}
$$

It follows that

$$
g_Y^3(\nabla_{\xi_Y} \lambda, \eta_Y) = \tilde{g}_Y(\mathrm{D}\,\lambda(Y)[\xi_Y], \eta_Y) + 2\,\mathrm{Re}(\mathrm{tr}(Y^* \xi_Y \lambda_Y^* \eta_Y + Y^* \lambda_Y \xi^* \eta_Y)).
$$

By definition, we have $\mathrm{Hess}\, F(Y)[\xi_Y] = \nabla_{\xi_Y} \mathrm{grad}\, F$. By Lemma (C.0.1), it suffices to assume that ξ_Y is a horizontal vector in order to obtain the Hessian operator of h. Therefore,

$$
\begin{aligned}
& g_Y^3(\mathrm{Hess}\, F(Y)[\xi_Y], \eta_Y) \\
= \;& g_Y^3(\nabla_{\xi_Y} \mathrm{grad}\, F, \eta_Y) \\
= \;& \tilde{g}(\eta_Y, \mathrm{D}\,\mathrm{grad}\, F(Y)[\xi_Y]) + 2\,\mathrm{Re}(\mathrm{tr}(Y^* \xi_Y \mathrm{grad}\, F(Y)^* \eta_Y + Y^* \mathrm{grad}\, F(Y) \xi_Y^* \eta_Y)) \\
= \;& \tilde{g}(\eta_Y, \mathrm{D}\,\mathrm{grad}\, F(Y)[\xi_Y]) + \mathrm{Re}(\mathrm{tr}((Y \eta_Y^* + \eta_Y Y^*)(\mathrm{grad}\, F(Y) \xi_Y^* + \xi_Y \mathrm{grad}\, F(Y)^*))) \\
= \;& \tilde{g}(\eta_Y, \mathrm{D}\,\mathrm{grad}\, F(Y)[\xi_Y]) \\
& +\tilde{g}\left(\eta_Y, \left(I - \frac{1}{2} P_Y\right)(\mathrm{grad}\, F(Y) \xi_Y^* + \xi_Y \mathrm{grad}\, F(Y)^*) Y (Y^* Y)^{-1}\right).
\end{aligned}
$$

Recall that for Riemannian metric g^3, we have

$$\operatorname{grad} F(Y) = \left(I - \frac{1}{2}P_Y\right)\nabla f(YY^*)Y(Y^*Y)^{-1}.$$

Hence,

$$\begin{aligned}
\mathrm{D}\operatorname{grad} F(Y)[\xi_Y] &= \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} \\
&\quad -\frac{1}{2}(\mathrm{D}\,(P_Y)[\xi_Y])\nabla f(YY^*)Y(Y^*Y)^{-1} \\
&\quad +\left(I - \frac{1}{2}P_Y\right)\nabla f(YY^*)\mathrm{D}\,(Y(Y^*Y)^{-1})[\xi_Y],
\end{aligned}$$

where we have

$$\begin{aligned}
\mathrm{D}\,(P_Y)[\xi_Y] &= \mathrm{D}\,(Y(Y^*Y)^{-1}Y^*)[\xi_Y] \\
&= \xi_Y(Y^*Y)^{-1}Y^* - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}Y^* + Y(Y^*Y)^{-1}\xi_Y^*
\end{aligned}$$

and

$$\mathrm{D}\,(Y(Y^*Y)^{-1})[\xi_Y] = \xi_Y(Y^*Y)^{-1} - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}.$$

Combining these equations we have

$$\begin{aligned}
&g_Y^3(\operatorname{Hess} F(Y)[\xi_Y], \eta_Y) \\
= \;& \tilde{g}\left(\eta_Y, \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1}\right) \\
& -\tilde{g}\left(\eta_Y, \frac{1}{2}(\xi_Y(Y^*Y)^{-1}Y^* - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}Y^* + Y(Y^*Y)^{-1}\xi_Y^*)\right. \\
& \left.\nabla f(YY^*)Y(Y^*Y)^{-1}\right) \\
& +\tilde{g}\left(\eta_Y, \left(I - \frac{1}{2}P_Y\right)\nabla f(YY^*)\left(\xi_Y(Y^*Y)^{-1} - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}\right)\right) \\
& +\tilde{g}\left(\eta_Y, \left(I - \frac{1}{2}P_Y\right)\left(\left(I - \frac{1}{2}P_Y\right)\nabla f(YY^*)Y(Y^*Y)^{-1}\xi_Y^*\right.\right. \\
& \left.\left. +\ \xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)\left(I - \frac{1}{2}P_Y\right)\right)Y(Y^*Y)^{-1}\right) \\
= \;& \tilde{g}\left(\eta_Y, \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
& -\tilde{g}\Big(\eta_Y, \frac{1}{2}(\xi_Y(Y^*Y)^{-1}Y^* - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}Y^* + Y(Y^*Y)^{-1}\xi_Y^*) \\
& \nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big) \\
& \nabla f(YY^*)\Big(\xi_Y(Y^*Y)^{-1} - Y(Y^*Y)^{-1}(\xi_Y^*Y + Y^*\xi_Y)(Y^*Y)^{-1}\Big)\Big) \\
& +\tilde{g}\Big(\eta_Y, \Big(I - \frac{3}{4}P_Y\Big)\nabla f(YY^*)Y(Y^*Y)^{-1}\xi_Y^*Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \frac{1}{2}\Big(I - \frac{1}{2}P_Y\Big)\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
= \;& \tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1}\Big) \\
& -\tilde{g}\Big(\eta_Y, \frac{1}{2}\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
& -\tilde{g}\Big(\eta_Y, \frac{1}{2}Y(Y^*Y)^{-1}\xi_Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \frac{1}{2}Y(Y^*Y)^{-1}\xi_Y^*P_Y\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \frac{1}{2}P_Y\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla f(YY^*)\Big((I - P_Y)\xi_Y(Y^*Y)^{-1} - Y(Y^*Y)^{-1}\xi_Y^*Y(Y^*Y)^{-1}\Big)\Big) \\
& +\tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla f(YY^*)Y(Y^*Y)^{-1}\xi_Y^*Y(Y^*Y)^{-1} - \\
& \frac{1}{4}P_Y\nabla f(YY^*)Y(Y^*Y)^{-1}\xi_Y^*Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \frac{1}{2}\left(I - P_Y\right)\xi_Y Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1} + \\
& \frac{1}{4}P_Y\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big) \\
= \;& \tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, (I - P_Y)\nabla f(YY^*)(I - P_Y)\xi_Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, \frac{1}{2}Y skew\Big((Y^*Y)^{-1}Y\xi_Y(Y^*Y)^{-1}Y^*\nabla f(YY^*)Y(Y^*Y)^{-1}\Big)\Big) \\
& +\tilde{g}\Big(\eta_Y, Y skew\Big((Y^*Y)^{-1}Y^*\nabla f(YY^*)(I - P_Y)\xi_Y(Y^*Y)^{-1}\Big)\Big) \\
= \;& \tilde{g}\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1}\Big) \\
& +\tilde{g}\Big(\eta_Y, (I - P_Y)\nabla f(YY^*)(I - P_Y)\xi_Y(Y^*Y)^{-1}\Big) \\
= \;& g_Y^3\Big(\eta_Y, \Big(I - \frac{1}{2}P_Y\Big)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} + \\
& (I - P_Y)\nabla f(YY^*)(I - P_Y)\xi_Y(Y^*Y)^{-1}\Big)
\end{aligned}$$

Hence for $\xi_Y \in \mathcal{H}_Y$, we have

$$\begin{aligned}\operatorname{Hess} F(Y)[\xi_Y] &= \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\xi_Y^* + \xi_Y Y^*]Y(Y^*Y)^{-1} \\ &\quad +(I - P_Y)\nabla f(YY^*)(I - P_Y)\xi_Y(Y^*Y)^{-1}\end{aligned}$$

To summarize, we obtain

$$\begin{aligned}\overline{\left(\operatorname{Hess} h(\pi(Y))[\eta_{\pi(Y)}]\right)}_Y &= P_Y^{\mathcal{H}^3}(\operatorname{Hess} F(Y)[\overline{\xi}_Y]) \\ &= \left(I - \frac{1}{2}P_Y\right)\nabla^2 f(YY^*)[Y\overline{\xi}_Y^* + \overline{\xi}_Y Y^*]Y(Y^*Y)^{-1} \\ &\quad +(I - P_Y)\nabla f(YY^*)(I - P_Y)\overline{\xi}_Y(Y^*Y)^{-1}.\end{aligned}$$