Riemannian Submersion Metrics on Grassmann Manifolds with Applications to Preconditioned Optimization

Markus Schlarb

Department of Mathematics, Julius-Maximilians-Universität Würzburg, Germany
Center for Signal Analysis of Complex Systems, Ansbach University of Applied Sciences, Germany

markus.schlarb@uni-wuerzburg.de

Fátima Silva Leite

Institute of Systems and Robotics, University of Coimbra, Portugal
Department of Mathematics, University of Coimbra, Portugal

fleite@mat.uc.pt

Knut Hüper

Department of Mathematics, Julius-Maximilians-Universität Würzburg, Germany

knut.hueper@uni-wuerzburg.de

Abstract

A family of Riemannian submersion metrics on the real Grassmann manifold, parameterized by smooth maps from the Stiefel manifold to the manifold of symmetric positive definite matrices that satisfy an invariance property, is investigated. These maps are strongly related to a known family of metrics on the Stiefel manifold. Endowed with metrics from these families, the Stiefel manifold becomes a Riemannian submersion over the Grassmann manifold. An explicit formula for the projection onto the horizontal bundle is derived and horizontal lifts are investigated. This leads to explicit expressions for Riemannian gradients and Hessians of smooth functions on the Grassmann manifold. The formulas are applied to the optimization of a well-known cost on the Grassmann manifold, the generalized Rayleigh quotient. Based on an estimate of the condition number of the Riemannian Hessian at a critical point of the cost, a construction of certain Riemannian metrics adapted to the cost is proposed. This gives rise to Riemannian preconditioning schemes. In addition, all those differential geometric quantities are explicitly derived to implement a geometric conjugate gradient algorithm on the Grassmann manifold endowed with a submersion metric.

Keywords: Grassmann manifolds, Stiefel manifolds, Riemannian submersions, Riemannian preconditioning, Rayleigh quotient optimization

2020 Mathematics Subject Classification: 53B20, 53C30, 65F08, 65F15

Contents

1 Introduction

The manifold of all k-dimensional subspaces of $\mathbb{R}^n$, known as Grassmann manifold, plays an important role in applications across several areas. Indeed, this manifold has been used to encode problems from computer vision [30], signal processing [16], machine learning [13], quantum computing [11], and discrete geometry [7]. However, one of the best known applications is probably related to the real symmetric eigenvalue problem, or, to be more precise, related to the computation of an invariant subspace of a real symmetric matrix. This problem can be tackled by minimizing the generalized Rayleigh quotient which induces a smooth function on the Grassmann manifold. Historically, algorithms based on (generalized) Rayleigh quotients were not necessarily formulated via differential geometric tools, see e.g. [27] or [19]. However, some of the algorithms that have been developed for computing invariant subspaces of a symmetric matrix rely on optimizing the generalized Rayleigh quotient on the Grassmann manifold using, among others, Riemannian Newton methods, Riemannian steepest descent methods, Riemannian conjugate gradients methods or Riemannian trust-region methods. Without being exhaustive, we refer to [1, 10, 14] for more details. Also, very recently, a gradient descent method and a conjugate gradient method for solving the symmetric invariant subspace problem using a retraction involving the polar-decomposition has been studied in [2]. An essential ingredient for a Riemannian optimization method is the choice of a Riemannian metric. In general, this choice influences the performance of the optimization algorithm, see e.g [1, Thm. 4.5.6] and [24, 29]. Moreover, choosing a metric that is adapted to the cost so that the performance of the Riemannian optimization algorithms is improved, is known as Riemannian preconditioning, see e.g [24, 29]. In particular, Riemannian preconditioning is of interest in the context of symmetric eigenvalue problems, as highlighted in the recent work [28].

Motivated by the impact of the choice of a Riemannian metric on the performance of Riemannian optimization algorithms, we study a family of submersion metrics on the Grassmann manifold. In particular, specific Riemannian metrics in this family are constructed that can be viewed as a preconditioning scheme for generalized Rayleigh quotient optimization. Nevertheless, the focus of this paper is mainly on differential geometric aspects.

This paper is structured as follows: After the introduction, some notations used throughout this text are listed in Section 2. Then, in Section 3, we first revisit a family of Riemannian metrics from [29] on the Stiefel manifold $\mathrm{St}_{n,k}$. This family of metrics is parameterized by smooth maps from $\mathrm{St}_{n,k}$ taking values in the manifold of symmetric positive definite $(n \times n)$-matrices $\mathrm{SPD}(n)$. We include detailed proofs for the readers convenience.

In Section 4, we study the Grassmann manifold considered from two perspectives: as a quotient of the Stiefel manifold by the orthogonal group, i.e. $\mathrm{St}_{n,k}/\mathrm{O}(k)$, and as an embedded submanifold $\mathrm{Gr}_{n,k}$ of the real symmetric matrices $\mathbb{R}^{n\times n}_{\mathrm{sym}}$ given by projection matrices. We show that the metrics on the Stiefel manifolds considered in Section 3 induce a family of metrics on the Grassmann manifold provided that the above map satisfies a certain invariance property. In other words, the Stiefel manifold becomes a Riemannian submersion over the Grassmann manifold identified with $\mathrm{St}_{n,k}/\mathrm{O}(k)$ as well as $\mathrm{Gr}_{n,k}$. For both identifications, explicit expressions for Riemannian gradients and Riemannian Hessians of smooth functions are derived. In order to obtain these expressions, a specific horizontal bundle on $\mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ given as the orthogonal complement of the vertical bundle with respect to the Riemannian metric under consideration is studied. In particular, an explicit expression for the projection onto the horizontal bundle is obtained.

In Section 5, we consider an application of the metrics investigated in Section 4 to an optimization problem on the Grassmann manifold: the computation of an invariant subspace of a symmetric matrix by optimizing the generalized Rayleigh quotient.

Motivated by the work on Riemannian preconditioning [29] and, in particular [25, Sec. 6], we construct a metric on the Grassmann manifold considered as the quotient $\mathrm{St}_{n,k}/\mathrm{O}(k)$. This metric belongs to the family introduced in Section 4, which is adapted to the cost, i.e. the generalized Rayleigh quotient. Our construction is based on an estimate of the condition number of the Riemannian Hessian of the cost at a critical point with respect to an arbitrary submersion metric from Section 4.

We then continue by considering the geometric conjugate gradient algorithm from [1, Alg. 13] applied to the generalized Rayleigh quotient leading to Algorithm 1 in Section 5. By using a retraction based on the polar decomposition and a suitable vector transport, this algorithm induces an algorithm on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ equipped with the submersion metric that is constructed before.

A detailed study of the convergence of Algorithm 1 is out of the scope of this text. Nevertheless, we perform a numerical experiment in Section 5.4.5 that suggest the performance of the proposed algorithm may benefit by choosing properly designed submersion metrics on the Grassmann manifold.

2 Notations

These are some of the notations used throughout the paper.

T_pM	tangent space of a manifold M at the point $p \in M$
N_pM	normal space of a manifold M at the point $p \in M$
TM	tangent bundle of M
T^*M	cotangent bundle of M
$\mathscr{L}_V f$	Lie derivative of the function f relative to the vector field V
∇^{LC}	Levi-Civita covariant derivative
$\Gamma^\infty\big(\operatorname{End}(TM)\big)$	smooth endomorphisms of the tangent bundle of M
$\Gamma^\infty\big(\mathrm{S}^2(T^*M)\big)$	smooth sections of the symmetric 2-tensor bundle over M
$\Vert.\Vert_F$	Frobenius norm
$\mathbb{R}^{n\times n}_{\mathrm{sym}}$	manifold of symmetric matrices of order n
$\lambda_{\max}(B)$	maximal eigenvalue of $B \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$
$\lambda_{\min}(B)$	minimal eigenvalue of $B \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$
κ	condition number of a linear operator or a symmetric matrix
$\operatorname{sym}(C)$	for a square matrix C, $\operatorname{sym}(C) := \frac{1}{2}(C + C^\top)$
$\operatorname{skew}(C)$	for a square matrix C, $\operatorname{skew}(C) := \frac{1}{2}(C - C^\top)$
$\mathrm{SPD}(n)$	manifold of symmetric positive definite matrices of order n
$\mathrm{St}_{n,k}$	Stiefel manifold, $\mathrm{St}_{n,k} = \{X \in \mathbb{R}^{n\times k} \mid X^\top X = I_k\}$
$\langle\cdot,\cdot\rangle^M$	Riemannian metric on $\mathrm{St}_{n,k}$ dependent on a smooth map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$
P_X	orthogonal projection from $\mathbb{R}^{n\times k}$ onto $T_X\mathrm{St}_{n,k}$
$\mathrm{O}(k)$	orthogonal group, $\mathrm{O}(k) = \{\Theta \in \mathbb{R}^{k\times k} \mid \Theta^\top\Theta = I_k\}$
$\mathfrak{so}(k)$	Lie algebra of $\mathrm{O}(k)$, $\mathfrak{so}(k) = \{A \in \mathbb{R}^{k\times k} \mid A^\top = -A\}$
$\triangleleft$	right $\mathrm{O}(k)$-action on $\mathrm{St}_{n,k}$
$\mathrm{St}_{n,k}/\mathrm{O}(k)$	quotient representation of Grassmann manifold
$\widetilde{\langle\cdot,\cdot\rangle}^M$	Riemannian metric on $\mathrm{St}_{n,k}/\mathrm{O}(k)$
$\operatorname{Ver}(\mathrm{St}_{n,k})$	vertical bundle of $\mathrm{St}_{n,k}$
$\operatorname{Hor}(\mathrm{St}_{n,k})$	horizontal bundle of $\mathrm{St}_{n,k}$
∇F	gradient of $F\colon U \subset \mathbb{R}^{n\times k} \to \mathbb{R}$ w.r.t. Frobenius scalar product
$\operatorname{grad} f$	Riemannian gradient of $f = F\big\vert_{\mathrm{St}_{n,k}}$
$\operatorname{Hess}(f)$	Riemannian Hessian of f
$\mathrm{Gr}_{n,k}$	Grassmann manifold, $\mathrm{Gr}_{n,k} = \{P \in \mathbb{R}^{n\times n} \mid P = P^2 = P^\top, \operatorname{tr}(P) = k\}$
$\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$	Riemannian metric on $\mathrm{Gr}_{n,k}$
$\otimes$	tensor product of vector spaces or matrix Kronecker product
vec	vec operator

3 Revisiting Stiefel Manifolds

For $1 \leq k \leq n$, the Stiefel manifold is defined by

$$\mathrm{St}_{n,k} = \{X \in \mathbb{R}^{n\times k} \mid X^\top X = I_k\} \tag{3.1}$$

whose tangent space at $X \in \mathrm{St}_{n,k}$ can be characterized by

$$T_X\mathrm{St}_{n,k} = \{V \in \mathbb{R}^{n\times k} \mid X^\top V = -V^\top X\}. \tag{3.2}$$

Its dimension is given by

$$\dim(\mathrm{St}_{n,k}) = nk - \tfrac{1}{2}k(k+1). \tag{3.3}$$

3.1 A Family of Metrics on Stiefel Manifolds

In this section, we adapt some results from [29]. More precisely, we recall the definition of the family of metrics from [29] as well as the formula for orthogonal projections and Riemannian gradients. In addition, the result from [29] on the Riemannian Hessian is generalized. Below we focus on the Stiefel manifold $\mathrm{St}_{n,k}$, i.e the matrix $B \in \mathbb{R}^{n\times n}$ in [29] is chosen as $B = I_n$. Detailed proofs are included for the readers convenience.

Denote by $\mathrm{SPD}(n)$ the manifold of positive definite real symmetric matrices and fix a smooth map

$$\mathrm{St}_{n,k} \to \mathrm{SPD}(n), \quad X \mapsto M_X. \tag{3.4}$$

Denote a smooth extension of this map to some open $U \subseteq \mathbb{R}^{n\times k}$ with $\mathrm{St}_{n,k} \subseteq U$ by the same symbol. That is $U \ni X \mapsto M_X \in \mathrm{SPD}(n)$. Define the Riemannian metric $\langle\cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*U)\big)$ on U for $X \in U$ and $V, W \in T_XU \cong \mathbb{R}^{n\times k}$ point-wise by

$$\langle\cdot,\cdot\rangle^M_X \colon T_XU \times T_XU \cong \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k} \to \mathbb{R}, \quad (V,W) \mapsto \langle V,W\rangle^M_X = \mathrm{tr}\big(V^\top M_X W\big). \tag{3.5}$$

Clearly, for each $X \in U$, $\langle\cdot,\cdot\rangle^M_X$ is an inner product, i.e. symmetric and positive definite by the assumption $M_X \in \mathrm{SPD}(n)$ for all $X \in U$. Hence $\langle\cdot,\cdot\rangle^M$ is indeed a Riemannian metric on U.

Let $\iota\colon \mathrm{St}_{n,k} \ni X \mapsto X \in U$ be the canonical inclusion. Then $\langle\cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*U)\big)$ induces the Riemannian metric $\iota^*\langle\cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{St}_{n,k})\big)$ on $\mathrm{St}_{n,k}$.

Notation 3.1 From now on, we suppress the pull-back of $\langle\cdot,\cdot\rangle^M$ by $\iota\colon \mathrm{St}_{n,k} \to U$ in the notation. In other words, by abuse of notation, we denote $\iota^*\langle\cdot,\cdot\rangle^M$ by $\langle\cdot,\cdot\rangle^M$, as well. Explicitly, $\langle\cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{St}_{n,k})\big)$ is given by

$$\langle V,W\rangle^M_X = \mathrm{tr}\big(V^\top M_X W\big) \tag{3.6}$$

for $X \in \mathrm{St}_{n,k}$ and $V, W \in T_X\mathrm{St}_{n,k}$.

Remark 3.2 1. The metrics on $\mathrm{St}_{n,k}$ of the form $\langle\cdot,\cdot\rangle^M$ include the one-parameter family of metrics from [17], the so-called α-metrics in the Riemannian case. Indeed, let $\alpha \in \mathbb{R}\setminus\{-1\}$ with $-\frac{2\alpha+1}{\alpha+1} > -2$ to guarantee that (3.7) below takes values in $\mathrm{SPD}(n)$. Then,

$$\mathrm{St}_{n,k} \mapsto \mathrm{SPD}(n), \quad X \mapsto M_X = 2I_k - \tfrac{2\alpha+1}{\alpha+1}XX^\top \tag{3.7}$$

yields a well-defined smooth map. Moreover, for $X \in \mathrm{St}_{n,k}$ and $V, W \in T_X\mathrm{St}_{n,k}$, one obtains by (3.7)

$$\langle V,W\rangle^M_X = \mathrm{tr}\big(V^\top(2I_k - \tfrac{2\alpha+1}{\alpha+1}XX^\top)W\big) = 2\,\mathrm{tr}\big(V^\top W\big) - \tfrac{2\alpha+1}{\alpha+1}\,\mathrm{tr}(V^\top XX^\top W), \tag{3.8}$$

in accordance with [17, Cor. 2].

2. The metrics on $\mathrm{St}_{n,k}$ of the form $\langle \cdot, \cdot \rangle^M$ include the metrics considered in [5]. Indeed, define the constant map $\mathrm{St}_{n,k} \ni X \mapsto M_X = A \in \mathrm{SPD}(n)$, where A is some fixed positive definite diagonal matrix. Then, $\langle \cdot, \cdot \rangle^M$ is a metric of the form considered in [5, Eq. (2)]. However, the scope of this reference is distinct form the objective of the present paper.

3.2 Orthogonal Projection onto $T_X\mathrm{St}_{n,k}$

For fixed $X \in \mathrm{St}_{n,k}$, we now recall an explicit expression for the orthogonal projection $P_X \colon T_X\mathbb{R}^{n\times k} \cong \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ with respect to $\langle \cdot, \cdot \rangle^M_X$. Following [29, p. 7] closely, we first characterize the normal bundle of $\mathrm{St}_{n,k} \subseteq U$ with respect to $\langle \cdot, \cdot \rangle^M$.

Lemma 3.3 *Let $X \in \mathrm{St}_{n,k}$. The normal space $N_X\mathrm{St}_{n,k}$ of $\mathrm{St}_{n,k}$ at X with respect to $\langle \cdot, \cdot \rangle^M_X$ is given by*

$$N_X\mathrm{St}_{n,k} = \big(T_X\mathrm{St}_{n,k}\big)^{\perp} = \big\{M_X^{-1}XS \mid S = S^\top \in \mathbb{R}^{k\times k}_{\mathrm{sym}}\big\}. \tag{3.9}$$

Proof: Let $S = S^\top \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$ and set $W = M_X^{-1}XS$. Moreover, let $V \in T_X\mathrm{St}_{n,k}$. Then,

$$\langle V, W\rangle^M_X = \mathrm{tr}\big(V^\top M_X(M_X^{-1}XS)\big) = \mathrm{tr}\big(V^\top XS\big) = 0, \tag{3.10}$$

where the last equality is valid due to $V^\top X \in \mathfrak{so}(k)$. Hence, the inclusion $\big\{M_X^{-1}XS \mid S \in \mathbb{R}^{k\times k}_{\mathrm{sym}}\big\} \subseteq N_X\mathrm{St}_{n,k}$ is proven. In addition, $\big\{M_X^{-1}XS \mid S \in \mathbb{R}^{k\times k}_{\mathrm{sym}}\big\}$ is a subspace of $\mathbb{R}^{n\times k}$ of dimension $k(k+1)/2$. This yields the desired result. □

Next, we define a linear map φ_{X,M_X} depending on the symmetric positive definite matrix $M_X \in \mathrm{SPD}(n)$ and $X \in \mathrm{St}_{n,k}$. We set

$$\varphi_{X,M_X} \colon \mathbb{R}^{k\times k}_{\mathrm{sym}} \to \mathbb{R}^{k\times k}_{\mathrm{sym}}, \quad S \mapsto (X^\top M_X^{-1}X)S + S(X^\top M_X^{-1}X). \tag{3.11}$$

Lemma 3.4 *The map $\varphi_{X,M_X} \colon \mathbb{R}^{k\times k}_{\mathrm{sym}} \to \mathbb{R}^{k\times k}_{\mathrm{sym}}$ defined in (3.11) is a linear isomorphism. For $T \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$, its inverse $S = \varphi^{-1}_{X,M_X}(T)$ is given by the unique solution $S \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$ of the Sylvester equation*

$$(X^\top M_X^{-1}X)S + S(X^\top M_X^{-1}X) = T. \tag{3.12}$$

Proof: Because $X^\top M_X^{-1}X \in \mathrm{SPD}(k)$ is positive definite, the assertion follows by [20, Thm. 5.2.2] □

After this preparation, we obtain a formula for the orthogonal projection which is a reformulation of [29, Lem. 3.1].

Lemma 3.5 *Let $X \in \mathrm{St}_{n,k}$. The orthogonal projection onto $T_X\mathrm{St}_{n,k}$ with respect to $\langle \cdot, \cdot \rangle^M_X$ is given by*

$$P_X \colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}, \quad V \mapsto V - 2M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top V)\big)\Big). \tag{3.13}$$

Proof: We first show that $P_X\big|_{T_X\mathrm{St}_{n,k}} = \mathrm{id}_{T_X\mathrm{St}_{n,k}}$. To this end, let $V \in T_X\mathrm{St}_{n,k}$. Then $X^\top V = -V^\top X \in \mathfrak{so}(k)$ holds implying $\mathrm{sym}(X^\top V) = 0$. Because $\varphi_{X,M_X} \colon \mathbb{R}^{k\times k}_{\mathrm{sym}} \to \mathbb{R}^{k\times k}_{\mathrm{sym}}$ is a linear isomorphism, we obtain $\varphi^{-1}_{X,M_X}(0) = 0$ leading to

$$P_X(V) = V - 2M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}(\mathrm{sym}(X^\top V))\Big) = V. \tag{3.14}$$

It remains to prove $P_X\big|_{N_X \mathrm{St}_{n,k}} = 0$. Let $V \in N_X \mathrm{St}_{n,k}$. Using Lemma 3.3, we write $V = M_X^{-1} X S$ with some suitable $S = S^\top \in \mathbb{R}^{k \times k}_{\mathrm{sym}}$. Then,

$$P_X(V) = M_X^{-1} X S - 2 M_X^{-1} X \Big(\varphi_{X, M_X}^{-1} \big(\tfrac{1}{2} \big(X^\top M_X^{-1} X S + S X^\top M_X^{-1} X \big) \big) \Big). \tag{3.15}$$

In order to determine the symmetric matrix $T = \varphi_{X,M_X}^{-1} \big(\frac{1}{2} (X^\top M_X^{-1} X S + S X^\top M_X^{-1} X) \big)$ from (3.15), we consider the equation

$$(X^\top M_X^{-1} X) T + T (X^\top M_X^{-1} X) = \tfrac{1}{2} (X^\top M_X^{-1} X S + S X^\top M_X^{-1} X). \tag{3.16}$$

Obviously, $T = \frac{1}{2} S$ is a solution, which is unique by Lemma 3.4. Plugging $T = \frac{1}{2} S$ into (3.15) yields

$$P_X(M_X^{-1} X S) = M_X^{-1} X S - 2 M_X^{-1} X \big(\tfrac{1}{2} S \big) = 0 \tag{3.17}$$

as desired. □

3.3 Riemannian Gradients and Hessians

Throughout this section, let $f \colon \mathrm{St}_{n,k} \to \mathbb{R}$ be smooth and let $F \colon U \subseteq \mathbb{R}^{n \times k} \to \mathbb{R}$ be a smooth extension of f, i.e. $f = F\big|_{\mathrm{St}_{n,k}}$. We denote the gradient of F at $X \in U$ with respect to the Frobenius scalar product by $\nabla F(X)$.

3.3.1 Riemannian Gradients

Lemma 3.6 *The gradient of $f \colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle \cdot, \cdot \rangle^M$ is given by*

$$\begin{aligned} \operatorname{grad} f(X) &= P_X \big(M_X^{-1} \nabla F(X) \big) \\ &= M_X^{-1} \nabla F(X) - 2 M_X^{-1} X \Big(\varphi_{X,M_X}^{-1} \big(\operatorname{sym}(X^\top M_X^{-1} \nabla F(X)) \big) \Big) \end{aligned} \tag{3.18}$$

for all $X \in \mathrm{St}_{n,k}$, where $\varphi_{X,M_X}^{-1} \colon \mathbb{R}^{k \times k}_{\mathrm{sym}} \to \mathbb{R}^{k \times k}_{\mathrm{sym}}$ is given by Lemma 3.4.

Proof: Obviously, the gradient of $F \colon U \to \mathbb{R}$ at $X \in U$ with respect to $\langle \cdot, \cdot \rangle^M \in \Gamma^\infty \big(\mathrm{S}^2 (T^* U) \big)$ is given by $\operatorname{grad} F(X) = M_X^{-1} \nabla F(X)$. In fact, one has for all $V \in \mathbb{R}^{n \times k}$

$$\langle V, \operatorname{grad} F(X) \rangle_X^M = \operatorname{tr} \big(V^\top M_X (M_X^{-1} \nabla F(X) \big) = \operatorname{tr} \big(V^\top \nabla F(X) \big) = \mathrm{D} F(X) V. \tag{3.19}$$

Using the well-known identity $\operatorname{grad} f(X) = P_X(\operatorname{grad} F(X))$ yields the desired result because $\big(\mathrm{St}_{n,k}, \langle \cdot, \cdot \rangle^M \big)$ is a Riemannian submanifold of $\big(\mathbb{R}^{n \times k}, \langle \cdot, \cdot \rangle^M \big)$. □

Notation 3.7 In the sequel, we always denote by $\operatorname{grad} f \in \Gamma^\infty(T \mathrm{St}_{n,k})$ the gradient of $f \colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle \cdot, \cdot \rangle^M$, no matter how the map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ is chosen. This map will always be clear from the context.

Corollary 3.8 *Let $\mathrm{St}_{n,k}$ be endowed with the metric $\langle \cdot, \cdot \rangle^M$ defined by the constant map $\mathrm{St}_{n,k} \ni X \mapsto M_X = I_n \in \mathrm{SPD}(n)$, i.e. with the Euclidean metric. Then the Riemannian gradient of $f \colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle \cdot, \cdot \rangle^M$ is given by*

$$\operatorname{grad} f(X) = \nabla F(X) - \tfrac{1}{2} X X^\top \nabla F(X) - \tfrac{1}{2} X \big(\nabla F(X) \big)^\top X. \tag{3.20}$$

Proof: For the specific choice $M_X = I_n$ for all $X \in \mathrm{St}_{n,k}$, the orthogonal projection from Lemma 3.5 simplifies for $V \in \mathbb{R}^{n \times k}$ to

$$P_X(V) = V - X \operatorname{sym}(X^\top V) = V - \tfrac{1}{2} X X^\top V - \tfrac{1}{2} X V^\top X, \tag{3.21}$$

see also [29, Eq. (3.17)]. Combining (3.21), Lemma 3.6, and $M_X = I_n$ for all $X \in \mathrm{St}_{n,k}$ proves (3.20) as desired. □

3.3.2 Riemannian Hessians

Next we consider the Riemannian Hessian of a smooth function $f\colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle\cdot,\cdot\rangle^M$. Let ∇^{LC} denote the Levi-Civita covariant derivative on $\big(\mathrm{St}_{n,k}, \nabla^{\mathrm{LC}}\big)$. By exploiting

$$\mathrm{Hess}(f)\big|_X(V) = \nabla^{\mathrm{LC}}_V \operatorname{grad} f\big|_X, \quad X \in \mathrm{St}_{n,k}, V \in T_X\mathrm{St}_{n,k}, \tag{3.22}$$

see e.g [12, Prop. 8.1] or [1, Def. 5.5.1], we derive a rather explicit expression for $\mathrm{Hess}(f) \in \Gamma^\infty\big(\operatorname{End}(T\mathrm{St}_{n,k})\big)$, the Riemannian Hessian of f considered as a section of the endomorphism bundle of $T\mathrm{St}_{n,k}$.

For the rest of this subsection, for simplifying notations, we use the same capital letters for tangent vectors and vector fields. However, the difference will be clear from the context. As a preparation, let

$$\widetilde{\nabla^{\mathrm{LC}}}\colon \Gamma^\infty(TU) \times \Gamma^\infty(TU) \to \Gamma^\infty(TU) \tag{3.23}$$

denote the Levi-Civita covariant derivative on U defined by $\langle\cdot,\cdot\rangle^M$. Moreover, let $V, W \in \Gamma^\infty\big(T\mathrm{St}_{n,k}\big)$ be vector fields on $\mathrm{St}_{n,k}$ and denote smooth extensions of V and W to U by $\widetilde{V}, \widetilde{W} \in \Gamma^\infty(TU)$, respectively. Then, see e.g. [26, Chap. 4, Lem. 3], the Levi-Civita covariant derivative $\nabla^{\mathrm{LC}}\colon \Gamma^\infty\big(T\mathrm{St}_{n,k}\big) \times \Gamma^\infty\big(T\mathrm{St}_{n,k}\big) \to \Gamma^\infty\big(T\mathrm{St}_{n,k}\big)$ on the Riemannian submanifold $\big(\mathrm{St}_{n,k}, \langle\cdot,\cdot\rangle^M\big)$ is given by

$$\nabla^{\mathrm{LC}}_V W\big|_X = P_X\big(\widetilde{\nabla^{\mathrm{LC}}}_{\widetilde{V}}\widetilde{W}\big|_X\big), \quad X \in \mathrm{St}_{n,k}. \tag{3.24}$$

We now take a closer look at $\widetilde{\nabla^{\mathrm{LC}}}$. Let $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,k\}$. Denote by $E_{ij} \in \mathbb{R}^{n\times k}$ the matrix whose entries satisfy $(E_{ij})_{f\ell} = \delta_{if}\delta_{j\ell}$. Obviously, for $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,k\}$, the vector fields

$$U \ni X \mapsto (X, E_{ij}) \in TU \tag{3.25}$$

denoted by E_{ij}, as well, form a local frame of TU. Moreover, denote by $\widetilde{V}, \widetilde{W} \in \Gamma^\infty(TU)$ vector fields on U and identify them with smooth maps $\widetilde{V}, \widetilde{W}\colon U \to \mathbb{R}^{n\times k}$. In addition, we write $\widetilde{W}_{ab}, \widetilde{V}_{ab}\colon U \to \mathbb{R}$ for the smooth functions given by the (a,b)-entry of $\widetilde{V}\colon U \to \mathbb{R}^{n\times k}$ and $\widetilde{W}\colon U \to \mathbb{R}^{n\times k}$, respectively, where $a \in \{1,\ldots,n\}$ and $b \in \{1,\ldots,k\}$. Then, by using [22, Prop. 4.6], we obtain for $\widetilde{\nabla^{\mathrm{LC}}}$ evaluated at $\widetilde{V}, \widetilde{W} \in \Gamma^\infty(TU)$

$$\widetilde{\nabla^{\mathrm{LC}}}_{\widetilde{V}}\widetilde{W} = \sum_{a=1}^{n}\sum_{b=1}^{k}\Big(\big(\mathscr{L}_{\widetilde{V}}\widetilde{W}_{ab}\big) + \sum_{c,f=1}^{n}\sum_{d,g=1}^{k}\widetilde{V}_{cd}\widetilde{W}_{fg}\Gamma^{(a,b)}_{(c,d),(f,g)}\Big)E_{ab}, \tag{3.26}$$

where $\Gamma^{(a,b)}_{(c,d),(f,g)}$ denotes the Christoffel symbols given for $a,c,f \in \{1,\ldots,n\}$ and $b,d,g \in \{1,\ldots,k\}$ by

$$\widetilde{\nabla^{\mathrm{LC}}}_{E_{cd}}E_{fg} = \sum_{a=1}^{n}\sum_{b=1}^{k}\Gamma^{(a,b)}_{(c,d),(f,g)}E_{ab}, \tag{3.27}$$

see e.g [22, Eq. (4.8)]. Next, define the smooth map $\Gamma\colon U \ni X \mapsto \Gamma_X \in \mathrm{S}((\mathbb{R}^{n\times k})^*)\otimes\mathbb{R}^{n\times k}$ for $X \in U$ and $\widetilde{V}, \widetilde{W} \in \mathbb{R}^{n\times k}$ by

$$\Gamma_X(\widetilde{V}, \widetilde{W}) = \sum_{a,c,f=1}^{n}\sum_{b,d,g=1}^{k}\widetilde{V}_{cd}\widetilde{W}_{fg}\Gamma^{(a,b)}_{(c,d),(f,g)}E_{ab}, \tag{3.28}$$

where $\Gamma^{(a,b)}_{(c,d),(f,g)}$ are the Christoffel symbols from (3.27). By this notation, we rewrite (3.26) for $X \in U$

$$\widetilde{\nabla^{\mathrm{LC}}}_{\widetilde{V}}\widetilde{W}\big|_X = \mathrm{D}\,\widetilde{W}(X)\widetilde{V}\big|_X + \Gamma_X\big(\widetilde{V}\big|_X, \widetilde{W}\big|_X\big). \tag{3.29}$$

In the sequel, we denote some smooth extension of $\operatorname{grad} f \in \Gamma^\infty(T\mathrm{St}_{n,k})$ to U by $\operatorname{grad} f$, as well. Then, for $X \in \mathrm{St}_{n,k}$ and $V \in T_X\mathrm{St}_{n,k}$ we get

$$\begin{aligned} \operatorname{Hess}(f)\big|_X(V) &\overset{(3.22)}{=} \nabla^{\mathrm{LC}}_V \operatorname{grad} f\big|_X \\ &\overset{(3.24),(3.29)}{=} P_X\Big(\mathrm{D}\big(\operatorname{grad} f\big)(X)V + \Gamma_X\big(V, \operatorname{grad} f(X)\big)\Big). \end{aligned} \tag{3.30}$$

In view of (3.30), one would need a rather explicit expression for $\Gamma_X(V,W)$ in order to obtain an explicit formula for $\operatorname{Hess}(f)\big|_X(V)$. However, our main interest in Section 5.2 is the Hessian of f at a critical point $X_* \in \mathrm{St}_{n,k}$, i.e. $\operatorname{grad} f(X_*) = 0$. Note that $\Gamma_{X_*}\big(V, \operatorname{grad} f(X_*)\big) = 0$ holds for all $V \in T_{X_*}\mathrm{St}_{n,k}$ because of (3.28). Hence (3.30) yields, for a critical point $X_* \in \mathrm{St}_{n,k}$ of f,

$$\operatorname{Hess}(f)\big|_{X_*}(V) = P_{X_*}\Big(\mathrm{D}\big(\operatorname{grad} f(X_*)\big)V\Big). \tag{3.31}$$

So we omit a detailed investigation of $\Gamma_X(V,W)$ defined in (3.28).

The next theorem gives an explicit expression for the Riemannian Hessian of $f\colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle\cdot,\cdot\rangle^M$ up to an explicit expression for Γ from (3.28). Note that this theorem generalizes [29, App. A.2], see also Corollary 3.11 below.

Theorem 3.9 *Let $X \in \mathrm{St}_{n,k}$, $V \in T_X\mathrm{St}_{n,k}$, and $\operatorname{grad} f$ be the gradient of f from Lemma 3.6 with respect to $\langle\cdot,\cdot\rangle^M$, as usual. In addition, let $\varphi_{X,M_X}\colon \mathbb{R}^{k\times k}_{\mathrm{sym}} \to \mathbb{R}^{k\times k}_{\mathrm{sym}}$ be the linear isomorphism defined in (3.11) and $\Gamma_X(V, \operatorname{grad} f(X))$ be given by (3.28). Using this notation, the Riemannian Hessian of f with respect to $\langle\cdot,\cdot\rangle^M$, defined by the smooth map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$, considered as a section of the endomorphism bundle $\mathrm{End}(T\mathrm{St}_{n,k}) \to \mathrm{St}_{n,k}$ is given by*

$$\begin{aligned} &\operatorname{Hess}(f)\big|_X(V) \\ &= P_X\Big(M_X^{-1}\,\mathrm{D}(\nabla F)(X)V - M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}\nabla F(X)\Big) \\ &\quad + P_X\bigg(\Big(M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}X - M_X^{-1}V\Big)\Big(X^\top \nabla F(X) - X^\top M_X \operatorname{grad} f(X)\Big)\bigg) \\ &\quad + P_X\Big(\Gamma_X(V, \operatorname{grad} f(X))\Big). \end{aligned} \tag{3.32}$$

Here $\mathrm{D}\,M(X)V$ denotes the tangent map of $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ at $X \in \mathrm{St}_{n,k}$ evaluated at $V \in T_X\mathrm{St}_{n,k}$, as usual.

Proof: In contrast to the proof in [29, App. A.2] which relies on the Weingarten map and where $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ is assumed to be constant, we compute $\operatorname{Hess}(f)$ directly by using (3.30). As a preparation, we consider the map

$$P\colon \mathrm{St}_{n,k} \times \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}, \quad (X,V) \mapsto P(X,V) = P_X(V), \tag{3.33}$$

where $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k} \subseteq \mathbb{R}^{n\times k}$ is given by Lemma 3.5. Then, by abuse of notation, we define for some fixed but arbitrary $V \in \mathbb{R}^{n\times k}$

$$P_V\colon \mathrm{St}_{n,k} \to \mathbb{R}^{n\times k}, \quad X \mapsto P_V(X) = P(X,V) = P_X(V). \tag{3.34}$$

Since, in the sequel, points in $\mathrm{St}_{n,k}$ are usually denoted by X, it will be always clear by the context whether (3.34) or the orthogonal projection from Lemma 3.5 is meant. Using this notation, and defining $\eta = M_X^{-1}\nabla F(X)$, we obtain, by plugging $\operatorname{grad} f$ from Lemma 3.6 into (3.30),

$$\begin{aligned}\operatorname{Hess}(f)\big|_X(V) &= P_X\Big(\Big(\mathrm{D}\,\Big(P_{(\cdot)}\big(M_{(\cdot)}^{-1}\nabla F(\cdot)\big)\Big)(X)V\Big) + \Gamma_X(V, \operatorname{grad} f(X))\Big)\\ &= P_X\Big(\,\mathrm{D}\,P_\eta(X)V\Big)\\ &\quad + P_X\Big(M_X^{-1}\,\mathrm{D}(\nabla F)(X)V - M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}\nabla F(X)\Big)\\ &\quad + P_X\Big(\Gamma_X(V, \operatorname{grad} f(X)\Big),\end{aligned} \tag{3.35}$$

where we used the chain-rule and exploited $P_X^2 = P_X$. In more detail, the tangent map of P defined in (3.33) at $(X, V) \in \mathrm{St}_{n,k} \times \mathbb{R}^{n\times k}$ evaluated at $(W, Y) \in T_{(X,V)}(\mathrm{St}_{n,k} \times \mathbb{R}^{n\times k}) \cong T_X\mathrm{St}_{n,k} \times \mathbb{R}^{n\times k}$ is given by

$$\mathrm{D}\,P(X, V)(W, Y) = \mathrm{D}\,P_V(X)W + \mathrm{D}\,P_X(V)Y = \mathrm{D}\,P_V(X)W + P_X(Y), \tag{3.36}$$

where the second equality follows by the linearity of $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k} \subseteq \mathbb{R}^{n\times k}$. Next we define the smooth map

$$g\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k} \times \mathbb{R}^{n\times k}, \quad X \mapsto g(X) = (X, M_X^{-1}\nabla F(X)). \tag{3.37}$$

Using

$$\mathrm{D}\,\big(\big(M_{(\cdot)}\big)^{-1}\big)(X)V = -M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}, \tag{3.38}$$

the tangent map of g at $X \in \mathrm{St}_{n,k}$ becomes for $V \in T_X\mathrm{St}_{n,k}$

$$\mathrm{D}\,g(X)V = \Big(V, M_X^{-1}\,\mathrm{D}(\nabla F)(X)V - M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}\nabla F(X)\Big). \tag{3.39}$$

Combining (3.33) and (3.37), we observe that

$$\operatorname{grad} f(X) = P_X\big(M_X^{-1}\nabla F(X)\big) = (P \circ g)(X) \tag{3.40}$$

holds for all $X \in \mathrm{St}_{n,k}$. Recalling the definition $\eta = M_X^{-1}\nabla F(X)$ and using (3.40) as well as the chain-rule, we now obtain for the Riemannian Hessian of f

$$\begin{aligned}\operatorname{Hess}(f)\big|_X(V) &= P_X\big(\,\mathrm{D}(\operatorname{grad} f)(X)V + \Gamma_X(V, \operatorname{grad} f(X))\big)\\ &= P_X\Big(\,\mathrm{D}(P \circ g)(X)V + \Gamma_X(V, \operatorname{grad} f(X))\Big)\\ &= P_X\Big(\,\mathrm{D}\,P(g(X))\big(\mathrm{D}\,g(X)V\big) + \Gamma_X(V, \operatorname{grad} f(X))\Big)\\ &\overset{(3.36),(3.34),(3.39)}{=} P_X\Big(\,\mathrm{D}\,P_\eta(X)V\Big)\\ &\quad + P_X\Big(M_X^{-1}\,\mathrm{D}(\nabla F)(X)V - M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}\nabla F(X)\Big)\\ &\quad + P_X\Big(\Gamma_X(V, \operatorname{grad} f(X))\Big),\end{aligned} \tag{3.41}$$

where we also used the definition $\eta = M_X^{-1}\nabla F(X)$ and $P_X^2 = P_X$ to obtain the last equality, i.e the second equality of (3.35) is proven.

Next we simplify the first summand of the second equality of (3.35). To this end, we first investigate $\mathrm{D}\,P_\eta(X)V$. Let $\gamma\colon I \to \mathrm{St}_{n,k}$ be a curve with $\gamma(0) = X$ and $\dot\gamma(0) = V \in T_X\mathrm{St}_{n,k}$.

Using the expression for $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ from Lemma 3.5 and the definition of $P_\eta\colon \mathrm{St}_{n,k} \to \mathbb{R}^{n\times k}$ from (3.34), we compute

$$\begin{aligned}
\mathrm{D}\, P_\eta(X)V &= \tfrac{\mathrm{d}}{\mathrm{d}t} P\big(\gamma(t), \eta\big)\big|_{t=0} \\
&= \tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\eta - 2M_{\gamma(t)}^{-1}\gamma(t)\Big(\varphi^{-1}_{\gamma(t),M_{\gamma(t)}}\big(\mathrm{sym}(\gamma(t)^\top \eta)\big)\Big)\Big)\big|_{t=0} \\
&= 2M_X^{-1}\big(\mathrm{D}\, M(X)V\big)M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad - 2M_X^{-1}V\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad - 2M_X^{-1}X\Big(\Big(\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi^{-1}_{\gamma(t),M_{\gamma(t)}}\big|_{t=0}\Big)\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad - 2M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(V^\top\eta)\big)\Big),
\end{aligned} \tag{3.42}$$

where (3.38) is exploited to get the first summand of the third equality. To simplify the third summand in (3.42) we compute

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi^{-1}_{\gamma(t),M_{\gamma(t)}}\big|_{t=0} = -\varphi^{-1}_{\gamma(0),M_{\gamma(0)}} \circ \Big(\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi_{\gamma(t),M_{\gamma(t)}}\big|_{t=0}\Big) \circ \varphi^{-1}_{\gamma(0),M_{\gamma(0)}} \tag{3.43}$$

by the chain rule, where we also used the well-known formula for the tangent map of the map $\mathrm{inv}\colon \mathrm{GL}(\mathbb{R}^{k\times k}_{\mathrm{sym}}) \ni \varphi \mapsto \varphi^{-1} \in \mathrm{GL}(\mathbb{R}^{k\times k}_{\mathrm{sym}})$, see e.g. [23, Cor. 4.3]. More explicitly, we obtain for $S \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$

$$\begin{aligned}
&\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi_{\gamma(t),M_{\gamma(t)}}(S)\big|_{t=0} \\
&= \tfrac{\mathrm{d}}{\mathrm{d}t}\Big(\big(\gamma(t)^\top M_{\gamma(t)}^{-1}\gamma(t)\big)S + S\big(\gamma(t)^\top M_{\gamma(t)}^{-1}\gamma(t)\big)\Big)\big|_{t=0} \\
&= \Big(V^\top M_X^{-1}X - X^\top\big(M_X^{-1}\big(\mathrm{D}\, M(X)V\big)M_X^{-1}\big)X + X^\top M_X^{-1}V\Big)S \\
&\quad + S\Big(V^\top M_X^{-1}X - X^\top\big(M_X^{-1}\big(\mathrm{D}\, M(X)V\big)M_X^{-1}\big)X + X^\top M_X^{-1}V\Big) \\
&=: \psi_{X,M_X,V}(S).
\end{aligned} \tag{3.44}$$

Obviously, $\psi_{X,M_X,V}\colon \mathbb{R}^{k\times k}_{\mathrm{sym}} \ni S \mapsto \psi_{X,M_X,V}(S) \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$, where $\psi_{X,M_X,V}(S)$ is defined by the last line of (3.44), is a linear map. Using this notation, we simplify (3.42). One obtains

$$\begin{aligned}
\mathrm{D}\, P_\eta(X)V &= 2M_X^{-1}\big(\mathrm{D}\, M(X)V\big)M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad - 2M_X^{-1}V\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad + 2M_X^{-1}X\Big(\big(\varphi^{-1}_{X,M_X} \circ \psi_{X,M_X,V} \circ \varphi^{-1}_{X,M_X}\big)\big(\mathrm{sym}(X^\top\eta)\big)\Big) \\
&\quad - 2M_X^{-1}X\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(V^\top\eta)\big)\Big).
\end{aligned} \tag{3.45}$$

Observe that the third and fourth summand on the right-hand side of (3.45) belong to the normal space $N_X\mathrm{St}_{n,k}$ by Lemma 3.3. Indeed, they are both of the form $M_X^{-1}XS$ with some symmetric matrix $S \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$. Thus, applying the orthogonal projection $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ to (3.45) yields

$$\begin{aligned}
&P_X\Big(\mathrm{D}\, P_\eta(X)V\Big) \\
&= P_X\bigg(\Big(2M_X^{-1}\big(\mathrm{D}\, M(X)V\big)M_X^{-1}X - 2M_X^{-1}V\Big)\Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top\eta)\big)\Big)\bigg).
\end{aligned} \tag{3.46}$$

Next, plugging (3.46) into (3.35) yields

$$
\begin{aligned}
&\operatorname{Hess}(f)\big|_X(V) \\
&= P_X\Big(M_X^{-1}\operatorname{D}(\nabla F)(X)V - M_X^{-1}\big(\operatorname{D}M(X)V\big)M_X^{-1}\nabla F(X)\Big) \\
&\quad + P_X\bigg(\Big(2M_X^{-1}\big(\operatorname{D}M(X)V\big)M_X^{-1}X - 2M_X^{-1}V\Big)\Big(\varphi_{X,M_X}^{-1}\big(\operatorname{sym}(X^\top\eta)\big)\Big)\bigg) \\
&\quad + P_X\Big(\Gamma_X(V, \operatorname{grad} f(X))\Big).
\end{aligned}
\tag{3.47}
$$

Let $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ be the orthogonal projection from Lemma 3.5, as usual. We now compute for $\eta = M_X^{-1}\nabla F(X) \in \mathbb{R}^{n\times k}$

$$
\begin{aligned}
&X^\top\nabla F(X) - X^\top M_X \operatorname{grad} f(X) \\
&= X^\top M_X\Big(M_X^{-1}\nabla F(X) - P_X\big(M_X^{-1}\nabla F(X)\big)\Big) \\
&= X^\top M_X\bigg(M_X^{-1}\nabla F(X) - \bigg(M_X^{-1}\nabla F(X) - 2M_X^{-1}X\Big(\varphi_{X,M_X}^{-1}\big(\operatorname{sym}(X^\top M_X^{-1}\nabla F(X))\big)\Big)\bigg)\bigg) \\
&= 2\varphi_{X,M_X}^{-1}\big(\operatorname{sym}(X^\top\eta)\big).
\end{aligned}
\tag{3.48}
$$

Plugging (3.48) into (3.47) yields the desired result. □

The next corollary provides an explicit formula for the Hessian of f at a critical point $X_* \in \mathrm{St}_{n,k}$. This expression will be used in Section 5.2 below.

Corollary 3.10 *Let $X_* \in \mathrm{St}_{n,k}$ be a critical point of f. Then the Riemannian Hessian of f at X_* is given by*

$$
\operatorname{Hess}(f)\big|_{X_*}(V) = P_{X_*}\Big(M_{X_*}^{-1}\operatorname{D}(\nabla F)(X_*)V - M_{X_*}^{-1}VX_*^\top\nabla F(X_*)\Big) \tag{3.49}
$$

for $V \in T_{X_*}\mathrm{St}_{n,k}$.

Proof: By exploiting $\operatorname{grad} f(X_*) = 0$ and $\Gamma_X(V, 0) = 0$, where $\Gamma_X(V, 0)$ is given by (3.28), Theorem 3.9 yields

$$
\begin{aligned}
\operatorname{Hess}(f)\big|_{X_*}(V) &= P_{X_*}\Big(M_{X_*}^{-1}\operatorname{D}(\nabla F)(X_*)V - M_{X_*}^{-1}\big(\operatorname{D}M(X_*)V\big)M_{X_*}^{-1}\nabla F(X_*)\Big) \\
&\quad + P_{X_*}\Big(\Big(M_{X_*}^{-1}\big(\operatorname{D}M(X_*)V\big)M_{X_*}^{-1}X_* - M_{X_*}^{-1}V\Big)X_*^\top\nabla F(X_*)\Big).
\end{aligned}
\tag{3.50}
$$

We next prove that (3.50) is equivalent to (3.49) by showing that $X_*X_*^\top\nabla F(X_*) = \nabla F(X_*)$. By Corollary 3.8, the gradient of f with respect to the Euclidean metric is given by

$$
\operatorname{grad} f(X) = \nabla F(X) - \tfrac{1}{2}XX^\top\nabla F(X) - \tfrac{1}{2}X\big(\nabla F(X)\big)^\top X, \quad X \in \mathrm{St}_{n,k}. \tag{3.51}
$$

Moreover, $\operatorname{grad} f(X_*) = 0$ holds because $X_* \in \mathrm{St}_{n,k}$ is a critical point of f. Therefore (3.51) yields at X_*

$$
\nabla F(X_*) = \tfrac{1}{2}\big(X_*X_*^\top\nabla F(X_*) + X_*\big(\nabla F(X_*)\big)^\top X_*\big) \tag{3.52}
$$

Multiplying (3.52) by $X_*X_*^\top$ from the left yields, due to $X_*^\top X_* = I_k$,

$$
X_*X_*^\top\nabla F(X_*) = \tfrac{1}{2}X_*X_*^\top\big(X_*X_*^\top\nabla F(X_*) + X_*\big(\nabla F(X_*)\big)^\top X_*\big) = \nabla F(X_*). \tag{3.53}
$$

The desired result follows. □

As a special case of Theorem 3.9, we obtain the following expression for the Riemannian Hessian that was already derived in [29, App. A.2].

Corollary 3.11 *Let* $\mathrm{St}_{n,k} \ni X \mapsto M_X = M \in \mathrm{SPD}(n)$ *be constant. The Riemannian Hessian of* $f\colon \mathrm{St}_{n,k} \to \mathbb{R}$ *with respect to* $\langle\cdot,\cdot\rangle^M$ *at* $X \in \mathrm{St}_{n,k}$ *evaluated at* $V \in T_X\mathrm{St}_{n,k}$ *is given by*

$$\begin{aligned}&\mathrm{Hess}(f)\big|_X(V)\\&= P_X\Big(M^{-1}\,\mathrm{D}(\nabla F)(X)V - M^{-1}V\Big(X^\top \nabla F(X) - X^\top M \,\mathrm{grad}\, f(X)\Big)\Big).\end{aligned} \tag{3.54}$$

Proof: Using $M = M_X$ for all $X \in \mathrm{St}_{n,k}$, the Riemannian metric $\langle\cdot,\cdot\rangle^M$ on $\mathrm{St}_{n,k}$ is induced by the scalar product on $\mathbb{R}^{n\times k}$ given for $V, W \in \mathbb{R}^{n\times k}$ point-wise by $\langle V, W\rangle^M = \mathrm{tr}\big(V^\top M W\big)$. By using the formula for the Christoffel symbols of $\widetilde{\nabla^{\mathrm{LC}}}$ from [22, Cor. 5.11], see also [29, App. A.2], for $\big(\mathrm{St}_{n,k}, \langle\cdot,\cdot\rangle^M\big)$ with constant $X \mapsto M_X = M$, we obtain $\Gamma_X(V,W) = 0$ for all $V, W \in \mathbb{R}^{n\times k}$, where $\Gamma_X(V,W)$ is given by (3.28). Moreover, the tangent map of $X \mapsto M_X$ vanishes, i.e. $\mathrm{D}\,M(X)V = 0$ holds for all $X \in \mathrm{St}_{n,k}$ and $V \in T_X\mathrm{St}_{n,k}$. Using these observations, the desired result follows from (3.32) in Theorem 3.9.□

4 Riemannian Submersion Metrics on the Grassmann Manifold

Consider the $\mathrm{O}(k)$-action on $\mathrm{St}_{n,k}$ via matrix multiplication from the right, i.e.

$$\triangleleft\colon \mathrm{St}_{n,k} \times \mathrm{O}(k) \to \mathrm{St}_{n,k}, \quad (X,R) \mapsto X \triangleleft R = XR. \tag{4.1}$$

For fixed $R \in \mathrm{O}(k)$, we write

$$\triangleleft_R\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}, \quad X \mapsto X\triangleleft_R = XR. \tag{4.2}$$

Remark 4.1 Obviously, (4.2) is a diffeomorphism induced by the $\mathrm{O}(k)$-action (4.1). Hence its tangent map at $X \in \mathrm{St}_{n,k}$, given by

$$\mathrm{D}(\triangleleft_R)(X)\colon T_X\mathrm{St}_{n,k} \to T_{XR}\mathrm{St}_{n,k}, \quad V \mapsto VR, \tag{4.3}$$

is a linear isomorphism. In particular, $T_{XR}\mathrm{St}_{n,k} = \big(T_X\mathrm{St}_{n,k}\big)R$ holds.

The canonical projection associated with the action (4.1) is denoted by

$$\mathrm{pr}\colon \mathrm{St}_{n,k} \ni X \mapsto \mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k). \tag{4.4}$$

Moreover, $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ is an $\mathrm{O}(k)$-principal fiber bundle, see e.g. [23, Sec. 18.5]. In addition, for fixed $X \in \mathrm{St}_{n,k}$, we define the map

$$X_\triangleleft\colon \mathrm{O}(k) \to \mathrm{St}_{n,k}, \quad R \mapsto X_\triangleleft R = XR. \tag{4.5}$$

It is well-known that $\mathrm{St}_{n,k}/\mathrm{O}(k)$ can be identified with the Grassmann manifold, i.e. the manifold of all k-dimensional subspaces of $\mathbb{R}^n$, see e.g. [3, Sec. 2.4].

In the remainder of this section, $\mathrm{St}_{n,k}$ is equipped with the Riemannian metric $\langle\cdot,\cdot\rangle^M \in \Gamma^\infty(\mathrm{S}^2(T^*\mathrm{St}_{n,k}))$. Under an additional assumption (Assumption 4.3) on the map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ defining $\langle\cdot,\cdot\rangle^M$ given by (3.6), we prove that the orthogonal complement of $\mathrm{Ver}(\mathrm{St}_{n,k}) = \ker(\mathrm{D}\,\mathrm{pr}) \subseteq T\mathrm{St}_{n,k}$ gives rise to a principal connection on $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$. This principal connection is determined explicitly in Section 4.1.

Afterwards, still imposing Assumption 4.3, Riemannian metrics on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ are defined such that $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ is a Riemannian submersion. In addition, explicit expressions for Riemannian gradients and Riemannian Hessians of smooth functions on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ with respect to these metrics are considered. In Section 4.3, similar differential geometric quantities are studied on the Grassmann manifold realized by projection matrices.

We start by the following lemma.

Lemma 4.2 *Let* $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ *be smooth such that for all* $X \in \mathrm{St}_{n,k}$ *and* $R \in \mathrm{O}(k)$
$$M_{XR} = M_X \tag{4.6}$$
is satisfied. Then the following assertions are fulfilled:

1. *The* $\mathrm{O}(k)$*-action defined in* (4.1) *is isometric.*
2. *Denote by* $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ *the equivalence class represented by* $X \in \mathrm{St}_{n,k}$. *Then the definition*
$$\check{M}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \ni \mathrm{pr}(X) \mapsto \check{M}_{\mathrm{pr}(X)} = M_X \in \mathrm{SPD}(n) \tag{4.7}$$
yields a well-defined smooth map.

Proof: Let $R \in \mathrm{O}(k)$, $X \in \mathrm{St}_{n,k}$ and $V, W \in T_X\mathrm{St}_{n,k}$. By the assumption $M_X = M_{XR}$, we obtain
$$\langle VR, WR\rangle^M_{XR} = \mathrm{tr}\left(R^\top V^\top M_{XR} W R\right) = \mathrm{tr}\left(V^\top M_X W\right) = \langle V, W\rangle^M_X \tag{4.8}$$
proving $(\triangleleft_R)^*\langle\cdot,\cdot\rangle^M = \langle\cdot,\cdot\rangle^M$ for all $R \in \mathrm{O}(k)$ as desired. It remains to prove the second claim. Because of
$$\check{M}_{\mathrm{pr}(XR)} = M_{XR} = M_X = \check{M}_{\mathrm{pr}(X)}, \tag{4.9}$$
the map $\check{M}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathrm{SPD}(n)$ defined in (4.7) is well-defined. In addition, the map $M\colon \mathrm{St}_{n,k} \to \mathrm{SPD}(n)$ is smooth by assumption and fulfills $M = \check{M} \circ \mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{SPD}(n)$. Hence $\check{M}$ is smooth by [21, Thm. 4.29] because $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ is a surjective submersion. □

Assumption 4.3 From now on, we always assume that the map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ fulfills $M_X = M_{XR}$ for all $X \in \mathrm{St}_{n,k}$ and $R \in \mathrm{O}(k)$.

4.1 A Principal Connection and the Orthogonal Projection onto the Horizontal Bundle

In the sequel, we consider $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ as $\mathrm{O}(k)$-principal fiber bundle. We recall that its vertical bundle is given point-wise by
$$\mathrm{Ver}(\mathrm{St}_{n,k})_X = \ker(\mathrm{D}\,\mathrm{pr}(X)) = \left\{X\Psi \mid \Psi \in \mathfrak{so}(k)\right\}, \quad X \in \mathrm{St}_{n,k}, \tag{4.10}$$
see e.g. [3, Sec. 2.4], and define the horizontal bundle $\mathrm{Hor}(\mathrm{St}_{n,k}) \subseteq T\mathrm{St}_{n,k}$ for $X \in \mathrm{St}_{n,k}$ point-wise by
$$\mathrm{Hor}(\mathrm{St}_{n,k})_X = \left(\mathrm{Ver}(\mathrm{St}_{n,k})_X\right)^\perp \subseteq T_X\mathrm{St}_{n,k}, \tag{4.11}$$
where the orthogonal complement of $\mathrm{Ver}(\mathrm{St}_{n,k})_X$ is taken with respect to $\langle\cdot,\cdot\rangle^M_X$.

Before we study the orthogonal projection onto $\mathrm{Hor}(\mathrm{St}_{n,k})$ in detail, we state the following characterization.

Lemma 4.4 *Let $X \in \mathrm{St}_{n,k}$ be fixed. Then $\mathrm{Hor}(\mathrm{St}_{n,k})_X$ defined in (4.11) is given by*

$$\mathrm{Hor}(\mathrm{St}_{n,k})_X = \{V \in T_X\mathrm{St}_{n,k} \mid X^\top M_X V = V^\top M_X X\}. \tag{4.12}$$

Proof: Let $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_X \subseteq T_X\mathrm{St}_{n,k}$. By definition, V is orthogonal to all $W \in \mathrm{Ver}(\mathrm{St}_{n,k})_X$. Using (4.10), we write $W = XA$ for some suitable $A \in \mathfrak{so}(k)$ and obtain that V satisfies

$$\langle XA, V\rangle_X^M = \mathrm{tr}\left(A^\top(X^\top M_X V)\right) = 0 \tag{4.13}$$

for all $A \in \mathfrak{so}(k)$. Hence $X^\top M_X V$ is symmetric by (4.13), i.e. $X^\top M_X V = V^\top M_X X$, as desired. □

Next, we define for fixed $X \in \mathrm{St}_{n,k}$ and $M_X \in \mathrm{SPD}(n)$ the map

$$\phi_{X,M_X} \colon \mathfrak{so}(k) \to \mathfrak{so}(k), \quad A \mapsto (X^\top M_X X)A + A(X^\top M_X X). \tag{4.14}$$

Similar to Lemma 3.4, we obtain the following result.

Lemma 4.5 *Let $X \in \mathrm{St}_{n,k}$ and $M_X \in \mathrm{SPD}(n)$. Then $\phi_{X,M_X} \colon \mathfrak{so}(k) \to \mathfrak{so}(k)$ is a linear isomorphism. For fixed $\Psi \in \mathfrak{so}(k)$, one can compute $A = \phi_{X,M_X}^{-1}(\Psi)$ by solving the Sylvester equation*

$$(X^\top M_X X)A + A(X^\top M_X X) = \Psi \tag{4.15}$$

for $A \in \mathfrak{so}(k)$.

Proof: Similar to Lemma 3.4, the assertion follows by well-known properties of the Sylvester equation. □

After this preparation, we state an explicit formula for the connection associated to $\mathrm{Hor}(\mathrm{St}_{n,k})$, i.e an explicit formula for the orthogonal projection onto $\mathrm{Ver}(\mathrm{St}_{n,k})$ with respect to $\langle\cdot,\cdot\rangle^M$.

Proposition 4.6 *Let $\mathrm{Hor}(\mathrm{St}_{n,k}) = \mathrm{Ver}(\mathrm{St}_{n,k})^\perp \subseteq T\mathrm{St}_{n,k}$ be the horizontal bundle given point-wise by (4.11) and let $\phi_{X,M_X} \colon \mathfrak{so}(k) \to \mathfrak{so}(k)$ be defined in (4.14).*

1. *For $X \in \mathrm{St}_{n,k}$ and $V \in T_X\mathrm{St}_{n,k}$, the connection $\mathcal{P} \in \Gamma^\infty\big(\mathrm{End}(T\mathrm{St}_{n,k})\big)$ associated to $\mathrm{Hor}(\mathrm{St}_{n,k})$, i.e. $\mathcal{P}\big|_{\mathrm{Ver}(\mathrm{St}_{n,k})} = \mathrm{id}_{\mathrm{Ver}(\mathrm{St}_{n,k})}$ and $\mathcal{P}\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})} = 0$, is given point-wise by*

$$\mathcal{P}\big|_X(V) = 2X\Big(\phi_{X,M_X}^{-1}\big(\mathrm{skew}\left(X^\top M_X V\right)\big)\Big). \tag{4.16}$$

2. *The connection one-form $\omega \in \Gamma^\infty(T^*\mathrm{St}_{n,k}) \otimes \mathfrak{so}(k)$ corresponding to $\mathcal{P}$ is given by*

$$\omega\big|_X(V) = 2\phi_{X,M_X}^{-1}\big(\mathrm{skew}\left(X^\top M_X V\right)\big), \quad X \in \mathrm{St}_{n,k},\ V \in T_X\mathrm{St}_{n,k}. \tag{4.17}$$

For $R \in \mathrm{O}(k)$, $X \in \mathrm{St}_{n,k}$, and $V \in T_X\mathrm{St}_{n,k}$, the equivariance property

$$\big((\triangleleft_R)^*\omega\big)\big|_X(V) = \mathrm{Ad}_{R^\top}\left(\omega\big|_X(V)\right) \tag{4.18}$$

is fulfilled. In particular, $\mathcal{P}$ is a principal connection on $\mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$.

Proof: Let $V \in T_X\mathrm{St}_{n,k}$. By Lemma 4.4, $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_X$ is satisfied iff

$$0 = X^\top M_X V - V^\top M_X X \tag{4.19}$$

holds. The next part of this proof is partially inspired by the proof of [29, Lem. 3.1]. Equation (4.10) justifies the following Ansatz. We write for $V \in T_X \mathrm{St}_{n,k}$

$$V^{\mathrm{ver}} = \mathcal{P}\big|_X(V) = XA_V, \tag{4.20}$$

where $A_V \in \mathfrak{so}(k)$ is some skew-symmetric matrix depending on V that still needs to be determined. Exploiting $T_X\mathrm{St}_{n,k} = \mathrm{Ver}(\mathrm{St}_{n,k})_X \oplus \mathrm{Hor}(\mathrm{St}_{n,k})_X$, we decompose $V = V^{\mathrm{ver}} + V^{\mathrm{hor}}$, where $V^{\mathrm{hor}} \in \mathrm{Hor}(\mathrm{St}_{n,k})_X$ is uniquely determined by

$$V^{\mathrm{hor}} \overset{(4.20)}{=} V - V^{\mathrm{ver}} = V - \mathcal{P}\big|_X(V) = V - XA_V. \tag{4.21}$$

Plugging (4.21) into (4.19) yields

$$\begin{aligned} 0 &= X^\top M_X V^{\mathrm{hor}} - (V^{\mathrm{hor}})^\top M_X X \\ &= X^\top M_X V - X^\top M_X X A_V - \big(V^\top M_X X - A_V^\top X^\top M_X X\big). \end{aligned} \tag{4.22}$$

Since $A_V^\top = -A_V$, (4.22) is equivalent to

$$(X^\top M_X X)A_V + A_V(X^\top M_X X) = X^\top M_X V - V^\top M_X X. \tag{4.23}$$

Using Lemma 4.5, the unique solution of (4.23) is given by

$$A_V = \phi_{X,M_X}^{-1}\big(X^\top M_X V - V^\top M_X X\big) = 2\phi_{X,M_X}^{-1}\big(\operatorname{skew}\big(X^\top M_X V\big)\big). \tag{4.24}$$

Plugging (4.24) into (4.20) yields

$$\mathcal{P}\big|_X(V) = X\Big(2\phi_{X,M_X}^{-1}\big(\operatorname{skew}\big(X^\top M_X V\big)\big)\Big) \tag{4.25}$$

being equivalent to (4.16).

It remains to prove Claim 2. Obviously, ω defined in (4.17) is a smooth one-form on $\mathrm{St}_{n,k}$ taking values in $\mathfrak{so}(k)$. Next, for $\Psi \in \mathfrak{so}(k)$, let $\Psi_{\mathrm{St}_{n,k}} \in \Gamma^\infty(T\mathrm{St}_{n,k})$ be the fundamental vector field associated with the action $\triangleleft\colon \mathrm{St}_{n,k} \times \mathrm{O}(k) \to \mathrm{St}_{n,k}$ defined in (4.1). Point-wise, that is for $X \in \mathrm{St}_{n,k}$,

$$\Psi_{\mathrm{St}_{n,k}}(X) = \tfrac{\mathrm{d}}{\mathrm{d}t}(X \triangleleft \exp(t\Psi))\big|_{t=0} = X\Psi. \tag{4.26}$$

By this notation, we obtain for $V \in T_X\mathrm{St}_{n,k}$, by exploiting the definition of ω in (4.17) and $\mathcal{P}$ in (4.16)

$$\big(\omega\big|_X(V)\big)_{\mathrm{St}_{n,k}}(X) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(X \triangleleft \exp\big(t\omega\big|_X(V)\big)\big)\big|_{t=0} = X\big(\omega\big|_X(V)\big) = \mathcal{P}\big|_X(V). \tag{4.27}$$

Thus $\omega \in \Gamma^\infty(T^*\mathrm{St}_{n,k}) \otimes \mathfrak{so}(k)$ is the connection one-form associated to $\mathcal{P}$, see e.g. [23, Eq. (19.1)]. To prove that ω satisfies the equivariance property, we first compute for $R \in \mathrm{O}(k)$

$$\big((\triangleleft_R)^*\omega\big)\big|_X(V) = \omega\big|_{XR}(VR) = 2\phi_{XR,M_X}^{-1}\Big(\operatorname{skew}\big(R^\top X^\top M_X V R\big)\Big), \tag{4.28}$$

where the last equality follows by Assumption (4.3), i.e. $M_X = M_{XR}$. To simplify the last equality of (4.28), we write $A_R = \phi_{XR,M_X}^{-1}\big(\operatorname{skew}\big(R^\top X^\top M_X V R\big)\big)$. By Lemma 4.5, A_R is the unique solution of the Sylvester equation

$$(R^\top X^\top M_X X R)A_R + A_R(R^\top X^\top M_X X R) = \tfrac{1}{2}R^\top\big(X^\top M_X V - V^\top M_X X\big)R \tag{4.29}$$

which is clearly equivalent to

$$(X^\top M_X X)(RA_R R^\top) + (RA_R R^\top)(X^\top M_X X) = \tfrac{1}{2}\big(X^\top M_X V - V^\top M_X X\big). \tag{4.30}$$

Denote by $A = \phi^{-1}_{X,M_X}\big(\frac{1}{2}\big(X^\top M_X V - V^\top M_X X\big)\big)$ the solution of

$$(X^\top M_X X)A + A(X^\top M_X X) = \tfrac{1}{2}\big(X^\top M_X V - V^\top M_X X\big), \tag{4.31}$$

as usual. Then $A_R = R^\top A R$ is a solution of (4.30) because of (4.31) which is unique by Lemma 4.5. Hence

$$A_R = R^\top A R = R^\top \Big(\phi^{-1}_{X,M_X}\big(\operatorname{skew}(X^\top M_X V)\big)\Big)R \tag{4.32}$$

is shown. The desired result follows by the definition of ω. □

Corollary 4.7 *Let $X \in \mathrm{St}_{n,k}$. The orthogonal projection onto $\mathrm{Hor}(\mathrm{St}_{n,k})_X$ with respect to $\langle\cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{St}_{n,k})\big)$ is given by*

$$\mathcal{P}^{\mathrm{hor}}_X \colon T_X\mathrm{St}_{n,k} \to \mathrm{Hor}(\mathrm{St}_{n,k})_X, \quad V \mapsto V - 2X\Big(\phi^{-1}_{X,M_X}\big(\operatorname{skew}(X^\top M_X V)\big)\Big). \tag{4.33}$$

Proof: Using $\mathrm{Hor}(\mathrm{St}_{n,k}) = \big(\mathrm{Ver}(\mathrm{St}_{n,k})\big)^\perp$, the assertion follows by $\mathcal{P}^{\mathrm{hor}}_X = \mathrm{id}_{T_X\mathrm{St}_{n,k}} - \mathcal{P}\big|_X$ and Proposition 4.6, Claim 1. □

Notation 4.8 Let $V \in \Gamma^\infty\big(T(\mathrm{St}_{n,k}/\mathrm{O}(k))\big)$ be a vector field. Then the horizontal lift of V with respect to the connection $\mathcal{P}$ from Proposition 4.6 is denoted by $\overline{V} \in \Gamma^\infty(T\mathrm{St}_{n,k})$. Analogously, the horizontal lift of $\check{V} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ with respect to $\mathcal{P}$ from Proposition 4.6 at $X \in \mathrm{St}_{n,k}$ is defined by $\overline{\check{V}}\big|_X = \big(\mathrm{D}\,\mathrm{pr}(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}\check{V}$.

Lemma 4.9 *Let $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ be equipped with the principal connection $\mathcal{P}$ from Proposition 4.6 associated with the horizontal bundle $\mathrm{Hor}(\mathrm{St}_{n,k}) \subseteq T\mathrm{St}_{n,k}$ defined in (4.11). Moreover, let $X \in \mathrm{St}_{n,k}$ and $R \in \mathrm{O}(k)$. Then, the following assertions are fulfilled:*

1. *The orthogonal projection $\mathcal{P}^{\mathrm{hor}} \in \Gamma^\infty\big(\mathrm{End}(T\mathrm{St}_{n,k}\big)$ from Corollary 4.7 fulfills, for all $V \in T_X\mathrm{St}_{n,k}$,*
$$\mathcal{P}^{\mathrm{hor}}\big|_{XR}(VR) = \big(\mathcal{P}^{\mathrm{hor}}\big|_X(V)\big)R. \tag{4.34}$$
2. $\mathrm{Hor}(\mathrm{St}_{n,k})_{XR} = \big(\mathrm{Hor}(\mathrm{St}_{n,k})_X\big)R.$
3. *The horizontal lift $\overline{\check{V}}\big|_X$ of $\check{V} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ at $X \in \mathrm{St}_{n,k}$ is related to the horizontal lift $\overline{\check{V}}\big|_{XR}$ at $XR \in \mathrm{St}_{n,k}$ by*
$$\overline{\check{V}}\big|_{XR} = \big(\overline{\check{V}}\big|_X\big)R. \tag{4.35}$$

Proof: First, recall from Remark 4.1 that $VR \in T_{XR}\mathrm{St}_{n,k}$ for all $V \in T_X\mathrm{St}_{n,k}$. Hence it makes sense to consider $\mathcal{P}\big|_{XR}(VR)$. Using Proposition 4.6, we can express the connection $\mathcal{P} \in \Gamma^\infty\big(\mathrm{End}(T\mathrm{St}_{n,k})\big)$ by

$$\mathcal{P}\big|_X(V) = 2X\Big(\phi^{-1}_{X,M_X}\big(\operatorname{skew}\big(X^\top M_X V\big)\big)\Big) = X\omega\big|_X(V), \tag{4.36}$$

where $\omega \in \Gamma^\infty(T^*\mathrm{St}_{n,k}) \otimes \mathfrak{so}(k)$ denotes the connection one-form associated with $\mathcal{P}$. Using the $\mathrm{O}(k)$-equivariance of ω, see Proposition 4.6, Claim 2, we obtain

$$\mathcal{P}\big|_{XR}(VR) = (XR)\big(\omega\big|_{XR}(VR)\big) = (XR)\big(R^\top \omega\big|_X(V)R\big) = \big(X\omega\big|_X(V)\big)R = \big(\mathcal{P}\big|_X(V)\big)R. \tag{4.37}$$

Since $\mathcal{P}^{\mathrm{hor}} = \mathrm{id}_{T\mathrm{St}_{n,k}} - \mathcal{P}$, Claim 1 follows from (4.37).

Next, using Claim 1, we observe

$$\mathrm{Hor}(\mathrm{St}_{n,k})_{XR} = \mathrm{im}(\mathcal{P}^{\mathrm{hor}}_{XR}) = \big(\mathrm{im}(\mathcal{P}^{\mathrm{hor}}_X)\big)R = \big(\mathrm{Hor}(\mathrm{St}_{n,k})_X\big)R \tag{4.38}$$

showing Claim 2.

It remains to prove Claim 3. Following [6, Ex. 9.25], let $\check{V} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k})$ and let $c\colon I \to \mathrm{St}_{n,k}$ be a smooth curve satisfying $c(0) = X$ and $\dot{c}(0) = \overline{\check{V}}\big|_X$. Next define $c_2\colon \mathbb{R} \to \mathrm{St}_{n,k}$ by $c_2(t) = c(t)R$. Obviously, $c_2(0) = XR$ and $\dot{c_2}(0) = \big(\overline{\check{V}}\big|_X\big)R$ holds. In addition, $\check{c}\colon \mathbb{R} \ni t \mapsto \mathrm{pr}(c(t)) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ is a smooth curve on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ fulfilling $\check{c}(t) = \mathrm{pr}(c(t)) = \mathrm{pr}\big(c(t)R\big) = \mathrm{pr}(c_2(t))$ for all $t \in \mathbb{R}$. Thus,

$$\check{V} = \mathrm{D}\,\mathrm{pr}(X)\overline{\check{V}}\big|_X = \tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{pr}(c(t))\big|_{t=0} = \tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{pr}(c_2(t))\big|_{t=0} = \mathrm{D}\,\mathrm{pr}(XR)\big(\overline{\check{V}}\big|_X\big)R. \tag{4.39}$$

Since $\big(\overline{\check{V}}\big|_X\big)R \in \mathrm{Hor}(\mathrm{St}_{n,k})_{XR}$ by Claim 2 and $\check{V} = \mathrm{D}\,\mathrm{pr}(XR)\big(\overline{\check{V}}\big|_X\big)R$ by (4.39), the desired result follows by uniqueness of horizontal lifts. □

4.2 The Grassmann Manifold as Riemannian Quotient Manifold

In this subsection, we endow $\mathrm{St}_{n,k}/\mathrm{O}(k)$ with a metric such that $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ becomes a Riemannian submersion. Moreover, Riemannian gradients and Riemannian Hessians of smooth functions on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ with respect to that metric are expressed in terms of horizontal lifts.

4.2.1 A Riemannian Metric

We start with the next lemma which defines the desired metric on $\mathrm{St}_{n,k}/\mathrm{O}(k)$.

Lemma 4.10 *Let $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ be smooth such that $M_X = M_{XR}$ holds for all $X \in \mathrm{St}_{n,k}$ and $R \in \mathrm{O}(k)$. Denote the associated metric on $\mathrm{St}_{n,k}$ by $\langle \cdot, \cdot \rangle^M$, as usual. Defining for all $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ represented by $X \in \mathrm{St}_{n,k}$ and $\check{V}, \check{W} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ point-wise*

$$\widetilde{\langle \check{V}, \check{W} \rangle}^M_{\mathrm{pr}(X)} = \langle \overline{\check{V}}\big|_X, \overline{\check{W}}\big|_X \rangle^M_X = \mathrm{tr}\,\big(\big(\overline{\check{V}}\big|_X\big)^\top M_X \overline{\check{W}}\big|_X\big) \tag{4.40}$$

yields the well-defined metric $\widetilde{\langle \cdot, \cdot \rangle}^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*(\mathrm{St}_{n,k}/\mathrm{O}(k)))\big)$ on $\mathrm{St}_{n,k}/\mathrm{O}(k)$. Moreover, the canonical projection $\mathrm{pr}\colon \big(\mathrm{St}_{n,k}, \langle \cdot, \cdot \rangle^M\big) \to \big(\mathrm{St}_{n,k}/\mathrm{O}(k), \widetilde{\langle \cdot, \cdot \rangle}^M\big)$ is a Riemannian submersion.

Proof: By Lemma 4.2, Claim 1, the action (4.1) is isometric with respect to $\langle \cdot, \cdot \rangle^M \in \Gamma^\infty(\mathrm{S}^2(T^*\mathrm{St}_{n,k}))$. Thus [6, Thm. 9.34] yields the desired result. □

4.2.2 Riemannian Gradients and Riemannian Hessians

For a smooth function $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$, we now express its Riemannian gradient and Riemannian Hessian with respect to $\widetilde{\langle \cdot,\cdot\rangle^M}$ in terms of horizontal lifts on $\mathrm{St}_{n,k}$. To this end, denote its pull-back to $\mathrm{St}_{n,k}$ via the canonical projection $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ by

$$f = \mathrm{pr}^* \check{f} = \check{f} \circ \mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathbb{R}. \tag{4.41}$$

Combining some well-known results from the literature on the so-called Riemannian quotient manifolds, see e.g. [6, Chap. 9], with the formulas from Section 3.3, the horizontal lift with respect to the connection $\mathcal{P}$ from Proposition 4.6 of the gradient $\operatorname{grad} \check{f} \in \Gamma^\infty\big(T(\mathrm{St}_{n,k}/\mathrm{O}(k))\big)$ of $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ with respect to $\widetilde{\langle \cdot,\cdot\rangle^M} \in \Gamma^\infty\big(\mathrm{S}^2(T^*(\mathrm{St}_{n,k}/\mathrm{O}(k)))\big)$ can be expressed in terms of the gradient of $f = \check{f} \circ \mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathbb{R}$ with respect to $\langle \cdot,\cdot\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{St}_{n,k})\big)$.

Before that, we state the following two lemmas concerning Riemannian gradients and Riemannian Hessians of smooth functions in the context of Riemannian submersions slightly generalizing [6, Prop. 9.38] and [6, Prop. 9.44], respectively. These lemmas will be needed in Section 4.3.2 below, as well.

We start with the following lemma adapted from [6, Prop. 9.38] and [26, Chap. 7, Lem. 34].

Lemma 4.11 *Let $\pi\colon E \to N$ be a Riemannian submersion, where the Riemannian metrics on E and N are both denoted by $\langle \cdot,\cdot\rangle$. Moreover, let $g\colon N \to \mathbb{R}$ be a smooth function. Denote by $\pi^* g = g \circ \pi\colon E \to \mathbb{R}$ the pull-back of g by π. Then, the gradient* $\operatorname{grad}(\pi^* g)$ *of $\pi^* g$ is horizontal, i.e.* $\operatorname{grad}(\pi^* g) \in \Gamma^\infty(\mathrm{Hor}(E))$. *Moreover,* $\operatorname{grad}(\pi^* g)$ *is π-related to the gradient* $\operatorname{grad} g$ *of g, i.e.*

$$\mathrm{D}\,\pi \circ \operatorname{grad}(\pi^* g) = (\operatorname{grad} g) \circ \pi. \tag{4.42}$$

In particular, the horizontal lift of $\operatorname{grad} g \in \Gamma^\infty(TN)$ *is given by* $\overline{\operatorname{grad} g} = \operatorname{grad}(\pi^* g) \in \Gamma^\infty(\mathrm{Hor}(E))$.

Proof: To show that $\operatorname{grad}(\pi^* g)$ is horizontal, let $p \in E$ and $v \in \mathrm{Ver}(E)_p$ be any arbitrary vertical vector. Then

$$\big\langle \operatorname{grad}(\pi^* g), v \big\rangle_p = \mathrm{D}(g \circ \pi)(p)v = \mathrm{D}\, g(\pi(p)) \circ \mathrm{D}\,\pi(p)v = 0, \tag{4.43}$$

where the last equality holds due to $\mathrm{D}\,\pi(p)v = 0$ since v is vertical. Thus $\operatorname{grad}(\pi^* g)$ is indeed horizontal by (4.43). To prove that $\mathrm{D}\,\pi \circ \operatorname{grad}(\pi^* g) = (\operatorname{grad} g) \circ \pi$, let $v \in \mathrm{Hor}(E)_p$. Using the chain-rule and exploiting that $\pi\colon E \to N$ is a Riemannian submersion, we compute

$$\begin{aligned}
\big\langle \mathrm{D}\,\pi(p) \operatorname{grad}(\pi^* g)(p), \mathrm{D}\,\pi(p)v \big\rangle_{\pi(p)} &= \big\langle \operatorname{grad}(\pi^* g)(p), v \big\rangle_p \\
&= \mathrm{D}(g \circ \pi)(p)v \\
&= \mathrm{D}\, g(\pi(p)) \circ \mathrm{D}\,\pi(p)v \\
&= \big\langle (\operatorname{grad} g)(\pi(p)), \mathrm{D}\,\pi(p)v \big\rangle_{\pi(p)}.
\end{aligned} \tag{4.44}$$

Comparing the left-hand side of the first equality in (4.44) with the last equality in (4.44) yields (4.42). Moreover, because $\operatorname{grad}(\pi^* g)$ is horizontal, we also obtain $\overline{\operatorname{grad} g} = \operatorname{grad}(\pi^* g)$ by (4.42). □

Next, we consider the Riemannian Hessian leading to the following lemma whose proof is adapted from [6, Prop. 9.44].

Lemma 4.12 *Let $\pi\colon E \to N$ be a Riemannian submersion and let $g\colon N \to \mathbb{R}$ be a smooth function whose pullback by π is denote by $\pi^* g = g \circ \pi\colon E \to \mathbb{R}$ as usual. Moreover, for $p \in E$ and $v \in T_{\pi(p)}N$, let $\overline{v}\big|_p = (\mathrm{D}\,\pi(p)\big|_{\mathrm{Hor}(P)_p})^{-1} v$ be the horizontal lift of v at p. Then the Hessian of g at $x = \pi(p) \in N$ fulfills*

$$\overline{\mathrm{Hess}(g)\big|_x(v)}\big|_p = \mathcal{P}_p^{\mathrm{hor}}\big(\mathrm{Hess}(\pi^* g)\big|_p(\overline{v}\big|_p\big), \tag{4.45}$$

where $\mathcal{P}^{\mathrm{hor}}\colon TP \to \mathrm{Hor}(P)$ denotes the orthogonal projection onto the horizontal bundle.

Proof: Let $V, W \in \Gamma^\infty(TN)$ be vector fields and denote their horizontal lift by $\overline{V}, \overline{W} \in \Gamma^\infty(\mathrm{Hor}(E))$, respectively. Moreover, by a slight abuse of notation, denote by ∇^{LC} the Levi-Civita covariant derivative on N and E, repsectively. Then

$$\overline{\nabla^{\mathrm{LC}}_V W} = \mathcal{P}^{\mathrm{hor}}\big(\nabla_{\overline{V}} \overline{W}\big) \tag{4.46}$$

holds by [26, Chap. 7, Lem. 45]. Recalling that for $x \in N$ and $v \in T_x N$, the Riemannian Hessian of g is given by $\mathrm{Hess}(g)\big|_x(v) = \nabla^{\mathrm{LC}}_v \operatorname{grad} g\big|_x$, see e.g. [12, Prop. 8.1], one obtains by (4.46)

$$\begin{aligned}
\overline{\mathrm{Hess}(g)\big|_x(v)}\big|_p &= \overline{\nabla^{\mathrm{LC}}_v \operatorname{grad} g\big|_x}\big|_p \\
&\overset{(4.46)}{=} \mathcal{P}_p^{\mathrm{hor}}\big(\nabla^{\mathrm{LC}}_{\overline{v}\big|_p} \overline{\operatorname{grad} g}\big|_p\big) \\
&= \mathcal{P}_p^{\mathrm{hor}}\big(\nabla^{\mathrm{LC}}_{\overline{v}\big|_p} \operatorname{grad}(\pi^* g)\big|_p\big) \\
&= \mathcal{P}_p^{\mathrm{hor}}\big(\mathrm{Hess}(\pi^* g)\big|_p(\overline{v}\big|_p)\big),
\end{aligned} \tag{4.47}$$

where the third equality follows by Lemma 4.11. This yields the desired result. □

We now apply Lemma 4.11 as well as Lemma 4.12 to the specific Riemannian submersion $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$, i.e. in the situation where $\pi = \mathrm{pr}$, $E = \mathrm{St}_{n,k}$ with metric $\langle \cdot, \cdot \rangle^M$ and $N = \mathrm{St}_{n,k}/\mathrm{O}(k)$ with metric $\widetilde{\langle \cdot, \cdot \rangle^M}$ as well as $g = \check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$.

Proposition 4.13 *Let $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ and $f = \check{f} \circ \mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathbb{R}$. Denote the gradient of f with respect to $\langle \cdot, \cdot \rangle^M$ by $\operatorname{grad} f \in \Gamma^\infty(T\mathrm{St}_{n,k})$, as usual. Moreover, let $F\colon U \to \mathbb{R}$ be some smooth extension of f, where $U \subseteq \mathbb{R}^{n \times k}$ is open. Then the horizontal lift $\overline{\operatorname{grad} \check{f}} \in \Gamma^\infty\big(\mathrm{Hor}(\mathrm{St}_{n,k})\big)$ with respect to the connection $\mathcal{P}$ from Proposition 4.6, of the gradient $\operatorname{grad} \check{f} \in \Gamma^\infty\big(T(\mathrm{St}_{n,k}/\mathrm{O}(k)\big)$ with respect to $\widetilde{\langle \cdot, \cdot \rangle^M}$ satisfies, for all $X \in \mathrm{St}_{n,k}$,*

$$\overline{\operatorname{grad} \check{f}(\mathrm{pr}(X))}\big|_X = \operatorname{grad} f(X) = P_X\big(M_X^{-1} \nabla F(X)\big). \tag{4.48}$$

Proof: Because the canonical projection $\mathrm{pr}\colon \big(\mathrm{St}_{n,k}, \langle \cdot, \cdot \rangle^M\big) \to \big(\mathrm{St}_{n,k}/\mathrm{O}(k), \widetilde{\langle \cdot, \cdot \rangle^M}\big)$ is a Riemannian submersion, the first equality of (4.48) follows by Lemma 4.11. The second equality of (4.48) holds due to Lemma 3.6 . □

Next we consider the Riemannian Hessian of $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ with respect to $\widetilde{\langle \cdot, \cdot \rangle^M}$.

Proposition 4.14 *Let $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ and $f = \mathrm{pr}^* \check{f} = \check{f} \circ \mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathbb{R}$. The Riemannian Hessian $\mathrm{Hess}(\check{f}) \in \Gamma^\infty\big(\mathrm{End}(T(\mathrm{St}_{n,k}/\mathrm{O}(k)))\big)$ with respect to $\widetilde{\langle \cdot, \cdot \rangle^M}$ satisfies for all $X \in \mathrm{St}_{n,k}$ and $\check{V} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$*

$$\overline{\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X)}(\check{V})}\big|_X = \mathcal{P}_X^{\mathrm{hor}}\Big(\mathrm{Hess}(f)\big|_X\big(\overline{\check{V}}\big|_X\big)\Big), \tag{4.49}$$

where $\mathrm{Hess}(f) \in \Gamma^\infty\big(\mathrm{End}(T\mathrm{St}_{n,k})\big)$ *denotes the Riemannian Hessian of* f *with respect to* $\langle \cdot, \cdot \rangle^M$ *from Theorem 3.9 and* $\mathcal{P}_X^{\mathrm{hor}} \colon T_X \mathrm{St}_{n,k} \to \mathrm{Hor}(\mathrm{St}_{n,k})_X$ *denotes the orthogonal projection onto the horizontal bundle given by Corollary 4.7.*

PROOF: This is a consequence of Lemma 4.12. □

The Hessian of $\check{f}$ at a critical point $\mathrm{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ is of our main interest.

Corollary 4.15 *Let* $\mathrm{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ *be a critical point of* $\check{f} \colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ *which is represented by* $X_* \in \mathrm{St}_{n,k}$ *and let* $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}$*. Using the notation from Proposition 4.14*

$$
\begin{aligned}
&\overline{\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}(\mathrm{D}\,\mathrm{pr}(X_*)V)}\big|_{X_*} \\
&= \mathcal{P}_{X_*}^{\mathrm{hor}}\Big(P_{X_*}\Big(M_{X_*}^{-1}\,\mathrm{D}(\nabla F)(X_*)V - M_{X_*}^{-1} V X_*^\top \nabla F(X_*)\Big)\Big)
\end{aligned}
\tag{4.50}
$$

is satisfied.

PROOF: Let $\mathrm{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ be a critical point of $\check{f}$. Then all $X \in \mathrm{St}_{n,k}$ fulfilling $\mathrm{pr}(X) = \mathrm{pr}(X_*)$ are critical points of $f = \check{f} \circ \mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathbb{R}$. Moreover, let $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}$. Then $\check{V} = \mathrm{D}\,\mathrm{pr}(X_*)V$ fulfills $V = \overline{\check{V}}\big|_{X_*}$ because $\mathrm{D}\,\mathrm{pr}(X_*)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}} \to T_{\mathrm{pr}(X_*)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ is a linear isomorphism. Hence the desired result follows by Proposition 4.14 combined with Corollary 3.10. □

4.3 The Grassmann Manifold realized by Projection Matrices

In the sequel, we identify the Grassmann manifold with the following embedded submanifold [14, Thm. 2.1]

$$
\mathrm{Gr}_{n,k} = \{P \in \mathbb{R}^{n\times n} \mid P = P^2 = P^\top,\ \mathrm{tr}(P) = k\} \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}. \tag{4.51}
$$

Moreover, see e.g. [14, Thm. 2.1], the tangent space of $\mathrm{Gr}_{n,k}$ at $P \in \mathrm{Gr}_{n,k}$ can be parameterized by

$$
T_P \mathrm{Gr}_{n,k} = \{[P, \Omega] \mid \Omega \in \mathfrak{so}(n)\} \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}. \tag{4.52}
$$

Define

$$
\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}, \quad X \mapsto XX^\top. \tag{4.53}
$$

Clearly, π is surjective. Its tangent map at $X \in \mathrm{St}_{n,k}$ reads

$$
\mathrm{D}\,\pi(X) \colon T_X \mathrm{St}_{n,k} \to T_{XX^\top}\mathrm{Gr}_{n,k}, \quad V \mapsto VX^\top + XV^\top. \tag{4.54}
$$

Using (4.54), one can show that π is indeed a surjective submersion, see e.g. [3, Sec. 2.4]. Moreover, by [3, Sec. 2.4], the vertical bundle $\ker(\mathrm{D}\,\pi) \subseteq T\mathrm{St}_{n,k}$ associated to $\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ coincides with the vertical bundle $\mathrm{Ver}(\mathrm{St}_{n,k})$ associated to the surjective submersion $\mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$. Thus, both vertical bundles are denoted by $\mathrm{Ver}(\mathrm{St}_{n,k}) = \ker(\mathrm{D}\,\pi) = \ker(\mathrm{D}\,\mathrm{pr})$ in the sequel. Note that $\mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ and $\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ are related by the following commutative diagram

$$
\begin{array}{ccc}
 & \mathrm{St}_{n,k} & \\
 {}^{\mathrm{pr}}\swarrow & & \searrow^{\pi} \\
\mathrm{St}_{n,k}/\mathrm{O}(k) & \xrightarrow{\quad \phi \quad} & \mathrm{Gr}_{n,k}
\end{array}
\tag{4.55}
$$

where

$$\phi\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathrm{Gr}_{n,k}, \quad \mathrm{pr}(X) \mapsto XX^\top \tag{4.56}$$

is a diffeomorphism, see Lemma 4.19 below. Since (4.53) is a surjective submersion, the restriction of $\mathrm{D}\,\pi(X)$ to $\mathrm{Hor}(\mathrm{St}_{n,k})_X$ characterized in Lemma 4.4 yields the linear isomorphism

$$\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\colon \mathrm{Hor}(\mathrm{St}_{n,k})_X \to T_{XX^\top}\mathrm{Gr}_{n,k}, \quad V \mapsto VX^\top + XV^\top. \tag{4.57}$$

An explicit expression for the inverse of (4.57) is of interest.

Lemma 4.16 *Let $X \in \mathrm{St}_{n,k}$. The inverse $\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}\colon T_{XX^\top}\mathrm{Gr}_{n,k} \to \mathrm{Hor}(\mathrm{St}_{n,k})_X$ of* (4.57)*, is given by*

$$\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z) = ZX - 2X\Big(\phi_{X,M_X}^{-1}\big(\mathrm{skew}(X^\top M_X ZX)\big)\Big), \tag{4.58}$$

where $Z \in T_{XX^\top}\mathrm{Gr}_{n,k}$ and $\phi_{X,M_X}\colon \mathfrak{so}(k) \to \mathfrak{so}(k)$ denotes the map from Lemma 4.5, as usual.

Proof: Let $Z \in T_{XX^\top}\mathrm{Gr}_{n,k}$ and write $V = \big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z) \in \mathrm{Hor}(\mathrm{St}_{n,k})_X$. Then $V \in T_X\mathrm{St}_{n,k}$ is uniquely characterized by the following two conditions

$$Z = \mathrm{D}\,\pi(X)V = VX^\top + XV^\top, \tag{4.59}$$

$$X^\top M_X V = V^\top M_X X, \tag{4.60}$$

where (4.60) ensures that $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_X$ due to Lemma 4.4. Multiplying (4.59) by X from the right, we get

$$V = ZX - XV^\top X = ZX - XA_Z, \tag{4.61}$$

where $A_Z = V^\top X \in \mathrm{so}(k)$ is some skew-symmetric matrix. Plugging (4.61) into (4.60) leads to

$$X^\top M_X(ZX - XA_Z) = (X^\top Z^\top - A_Z^\top X^\top)M_X X \tag{4.62}$$

Using $A_Z^\top = -A_Z$, (4.62) is equivalent to

$$(X^\top M_X X)A_Z + A_Z(X^\top M_X X) = X^\top M_X ZX - X^\top Z^\top M_X X. \tag{4.63}$$

By Lemma 4.5, the solution of (4.63) is unique. Moreover, using again Lemma 4.5, $A_Z = \phi_{X,M_X}^{-1}\big(2\,\mathrm{skew}(X^\top M_X ZX)\big)$. Plugging A_Z into (4.61) yields (4.58) as desired.

Note that for $Z \in T_{XX^\top}\mathrm{Gr}_{n,k}$, $V = ZX - XA_Z$ is indeed an element in $T_X\mathrm{St}_{n,k}$. To verify this, using (4.52), we write $Z = [XX^\top, \Omega]$ for some suitable $\Omega \in \mathrm{so}(n)$ implying $X^\top ZX = X^\top(XX^\top\Omega - \Omega XX^\top)X = 0$. Thus we obtain

$$X^\top V = X^\top(ZX - XA_Z) = A_Z = -A_Z^\top = V^\top X \tag{4.64}$$

showing $V \in T_X\mathrm{St}_{n,k}$. In addition, it is straightforward to verify that V satisfies indeed (4.59) and (4.60). This yields the desired result. □

The inverse of (4.57) satisfies the following equivariance property that will by useful in the sequel.

Lemma 4.17 *Let $X \in \mathrm{St}_{n,k}$, $R \in \mathrm{O}(k)$ and $Z \in T_{XX^\top}\mathrm{Gr}_{n,k}$ and assume that $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ satisfies Assumption 4.3, i.e. $M_X = M_{XR}$. Then*

$$\big(\mathrm{D}\,\pi(XR)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_{XR}}\big)^{-1}(Z) = \Big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z)\Big)R \tag{4.65}$$

holds.

Proof: We write $V = \big(\mathrm{D}\,\pi(XR)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_{XR}}\big)^{-1}(Z)$. By Lemma 4.16, $V \in \mathrm{Hor}(\mathrm{St}_{n,k})_{XR}$ is given by

$$V = Z(XR) - 2(XR)\Big(\phi^{-1}_{XR,M_{XR}}\big(\operatorname{skew}((XR)^\top M_X Z(XR))\big)\Big). \tag{4.66}$$

Analogously to the proof of Proposition 4.6, Claim 2, one shows that

$$\phi^{-1}_{XR,M_{XR}}\big(\operatorname{skew}((XR)^\top M_X Z(XR)\big) = R^\top\Big(\phi^{-1}_{X,M_X}\big(\operatorname{skew}(X^\top M_X Z X\big)\Big)R. \tag{4.67}$$

By plugging (4.67) into (4.66), we obtain

$$V = \Big(ZX - 2X\Big(\phi^{-1}_{X,M_X}\big(\operatorname{skew}(X^\top M_X Z X\big)\Big)\Big)R. \tag{4.68}$$

This yields the desired result. □

4.3.1 A Riemannian Metric

We now define a Riemannian metric on $\mathrm{Gr}_{n,k}$ such that $\pi\colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ becomes a Riemannian submersion.

Lemma 4.18 *Let $X \in \mathrm{St}_{n,k}$ and $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ be a smooth map satisfying Assumption 4.3. Moreover, let $Z_1, Z_2 \in T_{XX^\top}\mathrm{Gr}_{n,k}$ and write for $i \in \{1,2\}$ $A_{Z_i} = 2\phi^{-1}_{X,M_X}\big(\operatorname{skew}(X^\top M_X Z_i X)\big)$. Using this notation, we define point-wise*

$$\begin{aligned}
&\langle\!\langle Z_1, Z_2\rangle\!\rangle^M_{XX^\top} \\
&= \operatorname{tr}\Big(\big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z_1)\big)^\top M_X\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z_2)\big)\Big) \\
&= \operatorname{tr}\Big(\big(Z_1X - XA_{Z_1}\big)^\top M_X\big(Z_2 - XA_{Z_2}\big)\Big).
\end{aligned} \tag{4.69}$$

Then the following assertions are fulfilled:

1. *$\langle\!\langle\cdot,\cdot\rangle\!\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{Gr}_{n,k})\big)$ given point-wise by (4.69) is a well-defined Riemannian metric on $\mathrm{Gr}_{n,k}$.*
2. *If $\mathrm{St}_{n,k}$ is equipped with the metric $\langle\cdot,\cdot\rangle^M$ and $\mathrm{Gr}_{n,k}$ is equipped with the metric $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$, the map $\pi\colon \mathrm{St}_{n,k} \ni X \mapsto XX^\top \in \mathrm{Gr}_{n,k}$ is a Riemannian submersion.*

Proof: Using Lemma 4.17 and Assumption 4.3, one can directly verify that $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$ is well-defined. Indeed, let $R \in \mathrm{O}(k)$. Calculating

$$\begin{aligned}
&\langle\!\langle Z_1, Z_2\rangle\!\rangle^M_{(XR)(XR)^\top} \\
&= \operatorname{tr}\Big(\big(\big(\mathrm{D}\,\pi(XR)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_{XR}}\big)^{-1}(Z_1)\big)^\top M_{XR}\big(\mathrm{D}\,\pi(XR)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_{XR}}\big)^{-1}(Z_2)\big)\Big) \\
&\overset{(4.65)}{=} \operatorname{tr}\Big(R^\top\Big(\big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z_1)\big)^\top M_X\big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z_2)\big)R\Big) \\
&= \langle\!\langle Z_1, Z_2\rangle\!\rangle_{XX^\top}
\end{aligned} \tag{4.70}$$

shows that $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$ is well-defined. Hence $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*\mathrm{Gr}_{n,k})\big)$ holds and $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$ is positive definite by construction. This proves Claim 1.

Now, Claim 2 is an immediate consequence of the definition of $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$. □

Next we relate $\mathrm{Gr}_{n,k}$ equipped with $\langle\!\langle \cdot, \cdot \rangle\!\rangle^M$ to $\mathrm{St}_{n,k}/\mathrm{O}(k)$ equipped with $\widetilde{\langle \cdot, \cdot \rangle^M}$.

Lemma 4.19 *The map* $\phi \colon \mathrm{St}_{n,k}/\mathrm{O}(k) \ni \mathrm{pr}(X) \mapsto XX^\top \in \mathrm{Gr}_{n,k}$ *from* (4.56) *is a diffeomorphism. Moreover, it is an isometry if* $\mathrm{Gr}_{n,k}$ *is equipped with* $\langle\!\langle \cdot, \cdot \rangle\!\rangle^M$ *and* $\mathrm{St}_{n,k}/\mathrm{O}(k)$ *is endowed with* $\widetilde{\langle \cdot, \cdot \rangle^M}$.

Proof: We first prove that $\phi \colon \mathrm{St}_{n,k}/\mathrm{O}(k) \ni \mathrm{pr}(X) \mapsto XX^\top \in \mathrm{Gr}_{n,k}$, is a diffeomorphism. To this end, we recall that the diagram (4.55) is commutative, i.e $\pi = \phi \circ \mathrm{pr}$ holds. This yields that ϕ is well-defined and hence smooth by [21, Thm. 4.29] since $\mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ is a surjective submersion and $\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ is smooth. Moreover, ϕ is clearly bijective. Exploiting again $\pi = \phi \circ \mathrm{pr}$, we obtain for $X \in \mathrm{St}_{n,k}$ by the chain-rule

$$\mathrm{D}\,\phi(\mathrm{pr}(X)) \circ \mathrm{D}\,\mathrm{pr}(X) = \mathrm{D}\,\pi(X). \tag{4.71}$$

Restricting (4.71) to $\mathrm{Hor}(\mathrm{St}_{n,k})_X$ given by Lemma 4.4, we conclude

$$\mathrm{D}\,\phi(\mathrm{pr}(X)) = \mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})} \circ \big(\mathrm{D}\,\mathrm{pr}(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})}\big)^{-1}. \tag{4.72}$$

Hence, the linear map $\mathrm{D}\,\phi(\mathrm{pr}(X))$ is a linear isomorphism because it is a composition of linear isomorphisms. Thus $\phi \colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathrm{Gr}_{n,k}$ is local diffeomorphism around each $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$. Since, in addition, ϕ is bijective, ϕ is in fact a diffeomorphism.

It remains to show that ϕ is even an isometry. Indeed, exploiting that $\mathrm{pr} \colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ and $\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ are Riemannian submersions, the linear map $\mathrm{D}\,\phi(\mathrm{pr}(X))$ is not only a linear isomorphism but even a linear isometry because it is a composition of linear isometries by (4.72). This yields the desired result. □

4.3.2 Riemannian Gradients and Riemannian Hessians

We now derive formulas for gradients and Hessians of smooth functions on $\mathrm{Gr}_{n,k}$ with respect to the metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle^M$. To this end, we apply Lemma 4.11 as well as Lemma 4.12 to the Riemannian submersion $\pi \colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$, where $\mathrm{St}_{n,k}$ and $\mathrm{Gr}_{n,k}$ are equipped with $\langle \cdot, \cdot \rangle^M$ and $\langle\!\langle \cdot, \cdot \rangle\!\rangle^M$, respectively. Before that, we state the following auxiliary result.

Lemma 4.20 *Let* $g \colon \mathrm{Gr}_{n,k} \to \mathbb{R}$ *be smooth and let* $G \colon U \to \mathbb{R}$ *be some smooth extension of* g *to an open subset* $U \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}$ *containing* $\mathrm{Gr}_{n,k}$ *whose gradient at* $P \in U$ *with respect to the Frobenius scalar product is denoted by* $\nabla G(P)$*, as usual. Moreover, define* $\overline{\pi} \colon \mathbb{R}^{n\times k} \ni X \mapsto XX^\top \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ *and set* $\overline{U} = \overline{\pi}^{-1}(U)$*. Then*

$$F \colon \overline{U} \to \mathbb{R}, \quad X \mapsto (G \circ \overline{\pi})(X) = G(XX^\top) \tag{4.73}$$

is a smooth extension of the pull-back $f = \pi^* g = g \circ \pi \colon \mathrm{St}_{n,k} \to \mathbb{R}$ *to the open subset* $\overline{U} \subseteq \mathbb{R}^{n\times k}$ *containing* $\mathrm{St}_{n,k}$*. The gradient of* F *at* $X \in \overline{U}$ *with respect to the Frobenius scalar product reads as*

$$\nabla F(X) = 2\nabla G(XX^\top)X. \tag{4.74}$$

Proof: Obviously, $\overline{\pi} \colon \mathbb{R}^{n\times k} \ni X \mapsto XX^\top \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ is a smooth, and, in particular, continuous map. Thus $\overline{U} = \overline{\pi}^{-1}(U) \subseteq \mathbb{R}^{n\times k}$ is an open subset of $\mathbb{R}^{n\times k}$ being the preimage of the open $U \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}$ under the continuous map $\overline{\pi}$. In addition, $\overline{\pi}\big|_{\mathrm{St}_{n,k}} = \pi$ is clearly satisfied. Thus $F = G \circ \overline{\pi} \colon \overline{U} \to \mathbb{R}$ is a smooth extension of $f = \pi^* g = g \circ \pi \colon \mathrm{St}_{n,k} \to \mathbb{R}$.

Next, we compute the tangent map of F at $X \in \overline{U}$ evaluated at $V \in T_X\overline{U} \cong \mathbb{R}^{n\times k}$. It is given by

$$\begin{aligned} \mathrm{D}\, F(X)V &= \mathrm{D}\, G(\overline{\pi}(X)) \circ \mathrm{D}\, \overline{\pi}(X)V \\ &= \operatorname{tr}\Big(\big(\nabla G(XX^\top)\big)^\top \big(VX^\top + XV^\top\big)\Big) \\ &= 2 \operatorname{tr}\Big(\big(\nabla G(XX^\top)X\big)^\top V\Big), \end{aligned} \tag{4.75}$$

where we exploited $\nabla G(XX^\top) = \big(\nabla G(XX^\top)\big)^\top$ to obtain the last equality. Consequently, the gradient of F at $X \in \overline{U}$ with respect to the Frobenius scalar product is given by (4.74) as desired. □

After this preparation, we are in the position to determine the gradient of a smooth $g\colon \mathrm{Gr}_{n,k} \to \mathbb{R}$ with respect to $\langle\!\langle \cdot,\cdot \rangle\!\rangle^M$. To this end, Lemma 4.11 is used in the situation $E = \mathrm{St}_{n,k}$ with metric $\langle \cdot,\cdot \rangle^M$ and $N = \mathrm{Gr}_{n,k}$ with metric $\langle\!\langle \cdot,\cdot \rangle\!\rangle^M$. Combined with Lemma 4.20, this leads to the following proposition.

Proposition 4.21 *Let $g\colon \mathrm{Gr}_{n,k} \to \mathbb{R}$ be smooth and let $G\colon U \to \mathbb{R}$ be some smooth extension to an open $U \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}$. Moreover, let $X \in \mathrm{St}_{n,k}$ and denote by $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ the orthogonal projection with respect to $\langle \cdot,\cdot \rangle^M$ from Lemma 3.5. Then, the gradient of $\pi^* g = g \circ \pi\colon \mathrm{St}_{n,k} \to \mathbb{R}$ at $X \in \mathrm{St}_{n,k}$ with respect to $\langle \cdot,\cdot \rangle^M$ reads as*

$$\operatorname{grad}(\pi^* g)(X) = 2P_X\big(M_X^{-1}\nabla G(XX^\top)X\big) \tag{4.76}$$

and the gradient of g with respect to $\langle\!\langle \cdot,\cdot \rangle\!\rangle^M$ at $XX^\top \in \mathrm{Gr}_{n,k}$ is given by

$$\begin{aligned} &\operatorname{grad} g(XX^\top) \\ &= 2\Big(\Big(P_X\big(M_X^{-1}\nabla G(XX^\top)X\big)\Big)X^\top + X\Big(P_X\big(M_X^{-1}\nabla G(XX^\top)X\big)\Big)^\top\Big). \end{aligned} \tag{4.77}$$

Proof: As in Lemma 4.20, consider the smooth map $\overline{\pi}\colon \mathbb{R}^{n\times k} \ni X \mapsto XX^\top \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$. Then $F = G \circ \overline{\pi}\colon \overline{U} \to \mathbb{R}$ is a smooth extension of $f = \pi^* g$ to the open subset $\overline{U} = \overline{\pi}^{-1}(U) \subseteq \mathbb{R}^{n\times k}$ whose gradient at $X \in \overline{U}$ reads as $\nabla F(X) = 2\nabla G(XX^\top)X$. Because $\pi\colon \mathrm{St}_{n,k} \to \mathrm{Gr}_{n,k}$ is a Riemannian submersion, according to Lemma 4.11, the gradient of g at $XX^\top \in \mathrm{Gr}_{n,k}$ is given by

$$\operatorname{grad} g(XX^\top) = \operatorname{grad} g(\pi(X)) \overset{(4.42)}{=} \mathrm{D}\,\pi(X)\operatorname{grad}(\pi^* g)(X) = \mathrm{D}\,\pi(X)\operatorname{grad} f(X). \tag{4.78}$$

Next, using $\nabla F(X) = 2\nabla G(XX^\top)X$ and Lemma 3.6, we obtain for the gradient of $f = \pi^* g\colon \mathrm{St}_{n,k} \to \mathbb{R}$

$$\operatorname{grad} f(X) = P_X\big(M_X^{-1}\nabla F(X)\big) = 2P_X\big(M_X^{-1}\nabla G(XX^\top)X\big), \tag{4.79}$$

showing (4.76). Plugging (4.79) into (4.78) and using the formula for $\mathrm{D}\,\pi(X)$ from (4.54) yields (4.77) as desired. □

Next, we derive a formula for the Riemannian Hessian of $g\colon \mathrm{Gr}_{n,k} \to \mathbb{R}$ with respect to $\langle\!\langle \cdot,\cdot \rangle\!\rangle^M$ by using Lemma 4.20 and Theorem 3.9. In addition, Lemma 4.12 is applied, as well, where $E = \mathrm{St}_{n,k}$ is equipped with $\langle \cdot,\cdot \rangle^M$ and $N = \mathrm{Gr}_{n,k}$ is endowed with $\langle\!\langle \cdot,\cdot \rangle\!\rangle^M$.

Proposition 4.22 *Let $g\colon \mathrm{Gr}_{n,k} \to \mathbb{R}$ be smooth and let $G\colon U \to \mathbb{R}$ be some smooth extension to an open subset $U \subseteq \mathbb{R}^{n\times n}_{\mathrm{sym}}$ containing $\mathrm{Gr}_{n,k}$. Moreover, let $X \in \mathrm{St}_{n,k}$ and $P_X\colon \mathbb{R}^{n\times k} \to T_X\mathrm{St}_{n,k}$ be the orthogonal projection with respect to $\langle\cdot,\cdot\rangle^M$ from Lemma 3.5. Then, the Riemannian Hessian of g with respect to $\langle\!\langle\cdot,\cdot\rangle\!\rangle^M$ at $XX^\top \in \mathrm{Gr}_{n,k}$ evaluated at $Z \in T_{XX^\top}\mathrm{Gr}_{n,k}$ reads as*

$$\mathrm{Hess}(g)\big|_{XX^\top}(Z) = \mathrm{D}\,\pi(X) \circ \mathrm{Hess}(\pi^*g)\big|_X \circ \big(\mathrm{D}\,\pi\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z), \tag{4.80}$$

where $\big(\mathrm{D}\,\pi\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}\colon T_{XX^\top}\mathrm{Gr}_{n,k} \to \mathrm{Hor}(\mathrm{St}_{n,k})_X$ is given explicitly by Lemma 4.16. Moreover, $\mathrm{Hess}(\pi^*g)\big|_X\colon T_X\mathrm{St}_{n,k} \to T_X\mathrm{St}_{n,k}$ is the Riemannian Hessian with respect to $\langle\cdot,\cdot\rangle^M$ of the pull-back $\pi^*g = g\circ\pi\colon \mathrm{St}_{n,k} \to \mathbb{R}$ which is given by

$$\begin{aligned}
&\mathrm{Hess}(\pi^*g)\big|_X(V)\\
&= 2P_X\bigg(M_X^{-1}\Big(\mathrm{D}(\nabla G)(XX^\top)(V - XX^\top V) + \nabla G(XX^\top)V\Big)\\
&\qquad - M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}\nabla G(XX^\top)X\bigg)\\
&\quad + P_X\bigg(\Big(M_X^{-1}\big(\mathrm{D}\,M(X)V\big)M_X^{-1}X - M_X^{-1}V\Big)\Big(2X^\top\nabla G(XX^\top)X - X^\top M_X\,\mathrm{grad}(\pi^*g)(X)\Big)\bigg)\\
&\quad + P_X\Big(\Gamma_X(V, \mathrm{grad}(\pi^*g)(X))\Big)
\end{aligned} \tag{4.81}$$

where $\mathrm{grad}(\pi^*g)(X)$ is given by (4.76) from Proposition 4.21 and $\Gamma_X(V, \mathrm{grad}(\pi^*f)(X))$ is given by (3.28).

Proof: Because of $\pi(X) = XX^\top$, the Riemannian Hessian of g satisfies by Lemma 4.12,

$$\begin{aligned}
&\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}\Big(\mathrm{Hess}(g)\big|_{XX^\top}(Z)\Big)\\
&= \mathcal{P}^{\mathrm{hor}}_X\Big(\mathrm{Hess}(\pi^*g)\big|_X\Big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z)\Big).
\end{aligned} \tag{4.82}$$

Using $\mathrm{D}\,\pi(X)V = \mathrm{D}\,\pi(X)\big(\mathcal{P}_X(V) + \mathcal{P}^{\mathrm{hor}}_X(V)\big) = \mathrm{D}\,\pi(X)\mathcal{P}^{\mathrm{hor}}_X(V)$ because of $\mathrm{im}\,\mathcal{P}_X = \ker(\mathrm{D}\,\pi(X))$, applying $\mathrm{D}\,\pi(X)$ to both sides of (4.82) yields

$$\mathrm{Hess}(g)\big|_{XX^\top}(Z) = \mathrm{D}\,\pi(X)\Big(\mathrm{Hess}(\pi^*g)\big|_X\Big(\big(\mathrm{D}\,\pi(X)\big|_{\mathrm{Hor}(\mathrm{St}_{n,k})_X}\big)^{-1}(Z)\Big)\Big) \tag{4.83}$$

showing (4.80). It remains to prove (4.81). To this end, we rely on Lemma 4.20. Define again $\overline{\pi}\colon \mathbb{R}^{n\times k} \ni X \mapsto XX^\top \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$. Then $F = G\circ\overline{\pi}\colon \overline{U} \to \mathbb{R}$ is a smooth extension of $f = \pi^*g$ to the open subset $\overline{U} = \overline{\pi}^{-1}(U) \subseteq \mathbb{R}^{n\times k}$. Moreover, the gradient of F at $X \in \overline{U}$ with respect to the Frobenius scalar product reads as

$$\nabla F(X) = 2\nabla G(XX^\top)X = 2\nabla G(\pi(X))X \tag{4.84}$$

according to Lemma 4.20. Using the chain-rule, we obtain for the tangent map of (4.84) at $X \in \mathrm{St}_{n,k}$ evaluated at $V \in T_X\mathrm{St}_{n,k}$

$$\begin{aligned}
\mathrm{D}(\nabla F(X))V &= 2\big(\mathrm{D}(\nabla G)(XX^\top)(VX^\top + XV^\top)\big)X + 2\nabla G(XX^\top)V\\
&= 2\Big(\mathrm{D}(\nabla G)(XX^\top)(V - XX^\top V) + \nabla G(XX^\top)V\Big).
\end{aligned} \tag{4.85}$$

Plugging (4.84) and (4.85) into (3.32) from Theorem 3.9 yields the desired result. □

5 An Application to Optimization

In this section, we illustrate an application of the submersion metric on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ investigated in Section 4 in the context of Riemannian optimization. To be more precise, we consider the so-called generalized Rayleigh quotient associated with a symmetric matrix $A \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$, defined in (5.1) below. This cost induces a smooth function on the Grassmann manifold whose critical points are closely related to invariant subspaces of A. Indeed, as already pointed out in the introduction, some of the algorithms for computing invariant subspaces of a symmetric matrix that have been developed rely on optimizing the generalized Rayleigh quotient on the Grassmann manifold by means of a Riemannian optimization method. An important ingredient for such a method is the choice of a Riemannian metric which, in general, influences its performance. This fact gives rise to the notion of Riemannian preconditioning, see e.g [24, 29].

In the remainder of this section, we consider Riemannian submersion metrics on the quotient manifold $\mathrm{St}_{n,k}/\mathrm{O}(k)$ from Section 4 adapted to the generalized Rayleigh quotient. Based on the analysis of the Hessian at a critical point in Section 5.1 and using heuristic arguments, we propose a construction of a submersion metric that yields a Riemannian preconditioning scheme for the generalized Rayleigh quotient. Afterwards, this metric is used when a geometric conjugate gradient (CG) algorithm is adapted from [1, Alg. 13] to minimize the generalized Rayleigh quotient.

5.1 The Generalized Rayleigh Quotient

Let $A = A^\top \in \mathbb{R}^{n\times n}$. For simplicity, we assume in addition that A is positive definite. The generalized Rayleigh quotient, see e.g. [15, Sec. 1.3], is defined by

$$f\colon \mathrm{St}_{n,k} \to \mathbb{R}, \quad X \mapsto \tfrac{1}{2}\operatorname{tr}\big(X^\top A X\big), \tag{5.1}$$

where the factor $\frac{1}{2}$ is introduced for convenience. Obviously, the map

$$F\colon \mathbb{R}^{n\times k} \to \mathbb{R}, \quad X \mapsto \tfrac{1}{2}\operatorname{tr}(X^\top A X). \tag{5.2}$$

is a smooth extension of (5.1). Clearly, one has

$$\nabla F(X) = AX \quad \text{and} \quad \mathrm{D}(\nabla F)(X)V = AV \tag{5.3}$$

By Corollary 3.8 combined with (5.3), the gradient of the generalized Rayleigh quotient at $X \in \mathrm{St}_{n,k}$ with respect to the Euclidean metric is given by the well-known formula $\operatorname{grad} f(X) = AX - XX^\top AX$. Consequently, $X_* \in \mathrm{St}_{n,k}$ is a critical point of (5.1) iff

$$AX_* = X_*(X_*^\top A X_*) \tag{5.4}$$

is satisfied. In particular, X_* spans an invariant subspace of A. Furthermore, see e.g. [21, Thm. 4.29] or [1, Prop. 3.4.5], by the invariance $f\big(X \triangleleft R\big) = f(XR) = f(X)$, where $\triangleleft\colon \mathrm{St}_{n,k} \times \mathrm{O}(k) \to \mathrm{St}_{n,k}$ is given by (4.1), the generalized Rayleigh quotient induces the smooth map

$$\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}, \quad \mathrm{pr}(X) \mapsto \operatorname{tr}(X^\top A X), \tag{5.5}$$

where $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ is represented by $X \in \mathrm{St}_{n,k}$. In particular, $f = \check{f}\circ \mathrm{pr}$ is satisfied.

Using the formulas listed above, an explicit expression for the Riemannian gradient and Riemannian Hessian of the generalized Rayleigh quotient (5.1) with respect to $\langle\cdot,\cdot\rangle^M$ can be obtained by using the results of Section 3.3. In addition, relying on Section 4.2.2, the Riemannian Hessian of $\check{f}$ defined in (5.5) with respect to $\widetilde{\langle\cdot,\cdot\rangle}^M$ can be expressed in terms of horizontal lifts on $\mathrm{St}_{n,k}$.

5.2 The Hessian of the Generalized Rayleigh Quotient at a Critical Point

By [1, Thm. 4.5.6], the speed of local linear convergence of the accelerated line search algorithm [1, Alg. 1] to a non-degenerated local minimum depends on the condition number of the Riemannian Hessian of the cost at the local minimum. Moreover, in the context of Riemannian preconditioning, see e.g. [24, 25, 29], one tries to select a Riemannian metric such that the Riemannian Hessian of the cost at a critical point is well-conditioned.

Motivated by that fact, we now consider an estimate for the minimal and maximal eigenvalue of $\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}$. They are denoted by $\lambda_{\min}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big)$ and $\lambda_{\max}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big)$, respectively. Similarly, the minimal and maximal eigenvalue of a symmetric matrix $B = B^\top \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ are denoted by $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$, respectively. Using this notation, we obtain the following proposition whose proof generalizes ideas from [25, Thm. 6.1].

Proposition 5.1 *Let $\mathrm{St}_{n,k}/\mathrm{O}(k)$ be endowed with the Riemannian metric $\widetilde{\langle\cdot,\cdot\rangle^M}$ defined by an arbitrary smooth map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ fulfilling $M_{XR} = M_X$ for all $X \in \mathrm{St}_{n,k}$ and $R \in \mathrm{O}(k)$. Let $\operatorname{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ be a critical point of (5.5) represented by $X_* \in \mathrm{St}_{n,k}$, i.e. $AX_* = X_*(X_*^\top AX_*)$. Then,*

$$\lambda_{\min}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big) \geq \lambda_{\min}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top AX_*)\otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big) \tag{5.6}$$

and

$$\lambda_{\max}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big) \leq \lambda_{\max}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top AX_*)\otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big) \tag{5.7}$$

is satisfied, where $\otimes$ denotes the Kronecker product.

Proof: Let $V \in \operatorname{Hor}\big(\mathrm{St}_{n,k}\big)_{X_*}$. Since $\operatorname{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ is a critical point of $\check{f}$, $X_* \in \mathrm{St}_{n,k}$ is a critical point of f. Thus we obtain, by using (5.3) and Corollary 4.15,

$$\overline{\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}(\mathrm{D}\operatorname{pr}(X_*)V)} = \mathcal{P}^{\mathrm{hor}}_{X_*}\Big(P_{X_*}\Big(M_{X_*}^{-1}AV - M_{X_*}^{-1}V(X_*^\top AX_*)\Big)\Big). \tag{5.8}$$

Since $\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}$ is self-adjoint with respect to $\widetilde{\langle\cdot,\cdot\rangle^M}_{\operatorname{pr}(X_*)}$, we compute $\lambda_{\min}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big)$ by minimizing the associated Rayleigh quotient, i.e.

$$\begin{aligned}\lambda_{\min}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big) &= \min_{\check{V}\in T_{\operatorname{pr}(X_*)}\mathrm{Gr}_{n,k}} \frac{\widetilde{\big\langle \check{V}, \operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}(\check{V})\big\rangle^M}_{\operatorname{pr}(X_*)}}{\widetilde{\langle \check{V},\check{V}\rangle^M}_{\operatorname{pr}(X_*)}} \\ &= \min_{\check{V}\in T_{\operatorname{pr}(X_*)}\mathrm{Gr}_{n,k}} \frac{\big\langle \overline{\check{V}}\big|_{X_*}, \overline{\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}(\mathrm{D}\operatorname{pr}(X_*)\overline{\check{V}}\big|_{X_*})}\big|_{X_*}\big\rangle^M_{X_*}}{\big\langle \overline{\check{V}}\big|_{X_*}, \overline{\check{V}}\big|_{X_*}\big\rangle^M_{X_*}},\end{aligned} \tag{5.9}$$

where the second equality holds since $\mathrm{D}\operatorname{pr}(X_*)\big|_{\operatorname{Hor}(\mathrm{St}_{n,k})_{X_*}} \colon \operatorname{Hor}(\mathrm{St}_{n,k})_{X_*} \to T_{\operatorname{pr}(X_*)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ is an isometry. Using

$$\big\langle V, \mathcal{P}^{\mathrm{hor}}_{X_*}\big(P_{X_*}(W)\big)\big\rangle^M_{X_*} = \big\langle V, W\big\rangle^M_{X_*}, \tag{5.10}$$

for all $V \in \operatorname{Hor}(\mathrm{St}_{n,k})_{X_*}$ and $W \in T_{X_*}\mathbb{R}^{n\times k} \cong \mathbb{R}^{n\times k}$, as well as $\mathrm{D}\operatorname{pr}(X_*)(\operatorname{Hor}(\mathrm{St}_{n,k})_{X_*}) = T_{\operatorname{pr}(X_*)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$, we obtain by plugging (5.8) into (5.9)

$$\begin{aligned}\lambda_{\min}\big(\operatorname{Hess}(\check{f})\big|_{\operatorname{pr}(X_*)}\big) &= \min_{0\neq V\in\operatorname{Hor}(\mathrm{St}_{n,k})_{X_*}} \frac{\operatorname{tr}\Big(V^\top M_{X_*}\Big(M_{X_*}^{-1}AV - M_{X_*}^{-1}V(X_*^\top AX_*)\Big)\Big)}{\operatorname{tr}(V^\top M_{X_*}V)} \\ &= \min_{0\neq V\in\operatorname{Hor}(\mathrm{St}_{n,k})_{X_*}} \frac{\operatorname{tr}\Big(V^\top\Big(AV - V(X_*^\top AX_*)\Big)\Big)}{\operatorname{tr}(V^\top M_{X_*}V)}.\end{aligned} \tag{5.11}$$

To simplify (5.11), we use the so-called vec operator

$$\mathrm{vec}\colon \mathbb{R}^{n\times k} \to \mathbb{R}^{nk}, \tag{5.12}$$

see e.g. [4, Sec. 7.1]. Moreover, using properties of the Kronecker product, see e.g. [4, Sec. 7.1], we obtain

$$\mathrm{vec}(AV) = \mathrm{vec}(AVI_k) = (I_k^\top \otimes A)\mathrm{vec}(V) = (I_k \otimes A)\mathrm{vec}(V) \tag{5.13}$$

and

$$\mathrm{vec}(V(X_*^\top AX_*)) = \mathrm{vec}(I_n V(X_*^\top AX_*)) = ((X_*^\top AX_*) \otimes I_n)\mathrm{vec}(V) \tag{5.14}$$

as well as

$$\mathrm{vec}(M_{X_*}V) = \mathrm{vec}(M_{X_*}VI_k) = (I_k \otimes M_{X_*})\mathrm{vec}(V). \tag{5.15}$$

Using $\mathrm{tr}(V^\top W) = \mathrm{vec}(V)^\top \mathrm{vec}(W)$ for all $V, W \in \mathbb{R}^{n\times k}$ and plugging (5.13), (5.14), and (5.15) into (5.11) yields

$$\begin{aligned}
\lambda_{\min}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big) &= \min_{0\neq V\in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}} \frac{\mathrm{vec}(V)^\top \mathrm{vec}\big(AV - V(X_*^\top AX_*)\big)}{\mathrm{vec}(V)^\top \mathrm{vec}(M_{X_*}V)} \\
&= \min_{0\neq V\in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}} \frac{\mathrm{vec}(V)^\top \big(I_k\otimes A - (X_*^\top AX_*)\otimes I_n\big)\mathrm{vec}(V)}{\mathrm{vec}(V)^\top (I_k \otimes M_{X_*})\mathrm{vec}(V)} \\
&\geq \min_{0\neq V\in \mathbb{R}^{n\times k}} \frac{\mathrm{vec}(V)^\top \big(I_k\otimes A - (X_*^\top AX_*)\otimes I_n\big)\mathrm{vec}(V)}{\mathrm{vec}(V)^\top (I_k \otimes M_{X_*})\mathrm{vec}(V)} \\
&\overset{v=\mathrm{vec}(V)}{=} \min_{0\neq v\in \mathbb{R}^{nk}} \frac{v^\top \big(I_k\otimes A - (X_*^\top AX_*)\otimes I_n\big)v)}{v^\top (I_k \otimes M_{X_*})v} \\
&= \min_{0\neq v\in \mathbb{R}^{nk}} \frac{v^\top (I_k\otimes M_{X_*})^{-1/2}\big(I_k\otimes A - (X_*^\top AX_*)\otimes I_n)\big)(I_k\otimes M_{X_*})^{-1/2}v}{v^\top v} \\
&= \lambda_{\min}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top AX_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big)
\end{aligned} \tag{5.16}$$

showing (5.6). An analogous argument shows

$$\begin{aligned}
\lambda_{\max}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big) &= \max_{0\neq V\in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_*}} \frac{\big\langle V, \overline{\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}(\mathrm{D}\,\mathrm{pr}(X_*)V)}\big|_{X_*}\big\rangle^M_{X_*}}{\langle V, V\rangle^M_{X_*}} \\
&\leq \lambda_{\max}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top AX_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big)
\end{aligned} \tag{5.17}$$

as desired. □

If $\mathrm{pr}(X_*)$ is a local non-degenerated minimum of $\check{f}$, we can also estimate the condition number of $\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}$ which is denoted by $\kappa\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big)$ and defined as

$$\kappa\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big) = \frac{\big|\lambda_{\max}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big)\big|}{\big|\lambda_{\min}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big)\big|}. \tag{5.18}$$

In the sequel, the condition number of a symmetric invertible matrix $B \in \mathbb{R}^{nk\times nk}$ is denoted by $\kappa(B)$, as well.

Corollary 5.2 *Let* $\mathrm{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ *be a local non-degenerated minimum of the generalized Rayleigh quotient* $\check{f}$. *Then*

$$\kappa\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big) \geq \kappa\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big) \tag{5.19}$$

is satisfied.

Proof: Because $\mathrm{pr}(X_*)$ is a non-degenerated local minimum of $\check{f}$, all eigenvalues of $\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}$ are positive. Thus, by using the estimates from Proposition 5.1, we obtain

$$\begin{aligned}
\kappa\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big) &= \frac{\big|\lambda_{\max}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big)\big|}{\big|\lambda_{\min}\big(\mathrm{Hess}(\check{f})\big|_{\mathrm{pr}(X_*)}\big)\big|} \\
&\geq \frac{\big|\lambda_{\max}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big)\big|}{\big|\lambda_{\min}\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big)\big|} \\
&= \kappa\big((I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2})\big)
\end{aligned} \tag{5.20}$$

as desired. □

5.3 A Specific Riemannian Submersion Metric

Based on Corollary 5.2, we now construct of a metric adapted to the generalized Rayleigh quotient which is strongly motivated by the preconditioning scheme proposed in [25, Sec. 6]. Inspired by this reference, our aim is to define the Riemannian metric $\langle \cdot, \cdot \rangle^M$ such that, heuristically, M_{X_*} fulfills

$$(I_k \otimes M_{X_*}^{-1/2})\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes M_{X_*}^{-1/2}) \approx I_{nk} \tag{5.21}$$

approximately. In fact, if (5.21) was satisfied exactly, we would obtain $\kappa\big(\mathrm{Hess}(f)\big|_{\mathrm{pr}(X_*)}\big) = 1$ by Corollary 5.2. However, in general, we cannot expect to find some $M_{X_*} \in \mathrm{SPD}(n)$ such that (5.21) is satisfied. Hence we propose the following heuristic construction. Clearly, if (5.21) was an equality, it would be equivalent to

$$I_k \otimes A - (X_*^\top A X_*) \otimes I_n = I_k \otimes M_{X_*}. \tag{5.22}$$

In case that a solution of (5.22) does not exist, we consider the associated least-square problem.

Lemma 5.3 *Using the notation introduced above, and writing* $M = M_{X_*}$ *for short, the unique solution of the optimization problem*

$$\min_{M \in \mathbb{R}^{n\times n}_{\mathrm{sym}}} \big\| I_k \otimes M - \big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)\big\|_F^2 \tag{5.23}$$

is given by

$$M_* = A - \tfrac{1}{k}\,\mathrm{tr}(X_*^\top A X_*) I_n. \tag{5.24}$$

Proof: Set $B = I_k \otimes A - (X_*^\top A X_*) \otimes I_n$ and define $\ell \colon \mathbb{R}^{n\times n}_{\mathrm{sym}} \to \mathbb{R}$ by

$$\ell(M) = \|I_k \otimes M - B\|_F^2 = \mathrm{tr}\,\big(I_k \otimes (M^2) - 2(I_k \otimes M)B + B^2\big), \tag{5.25}$$

where we exploited that M and B are both symmetric. The tangent map of ℓ at $M \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ evaluated at $V \in T_M\mathbb{R}^{n\times n}_{\mathrm{sym}} \cong \mathbb{R}^{n\times n}_{\mathrm{sym}}$ is given by

$$\begin{aligned} \mathrm{D}\,\ell(M)V &= \operatorname{tr}\Big(I_k \otimes (2MV) - 2(I_k \otimes V)B\Big) \\ &= \operatorname{tr}\Big(I_k \otimes (2MV) - 2\operatorname{tr}\Big(\big(I_k \otimes A - (X_*^\top A X_*) \otimes I_n\big)(I_k \otimes V)\big)\Big) \\ &= 2\operatorname{tr}(I_k)\operatorname{tr}(MV) - 2\operatorname{tr}(I_k)\operatorname{tr}(AV) + 2\operatorname{tr}(X_*^\top A X_*)\operatorname{tr}(I_n V) \\ &= 2\operatorname{tr}\big(\big(kM - kA + \operatorname{tr}(X_*^\top A X_*)I_n\big)V\big), \end{aligned} \tag{5.26}$$

where we exploited well-known properties of the Kronecker product, see e.g. [4, Sec. 7.1]. Because of (5.26), the gradient of ℓ with respect to the Frobenius scalar product is given by

$$\nabla\ell(M) = 2\big(kM - kA + \operatorname{tr}(X_*^\top A X_*)I_n\big), \quad M \in \mathbb{R}^{n\times n}_{\mathrm{sym}}. \tag{5.27}$$

In particular, $M_* \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ is a critical point of ℓ iff $\nabla\ell(M_*) = 0$, i.e. M_* given by (5.24) is a critical point of ℓ. It remains to show that M_* is the unique minimum of ℓ. To this end, compute for $V, W \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ the second derivative of ℓ at $M \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$. We have

$$\mathrm{D}^2\,\ell(M)(V,W) = \tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{D}\,\ell(M+tW)V\big|_{t=0} \overset{(5.26)}{=} 2\operatorname{tr}\big((kWV)\big) = 2k\operatorname{tr}\big(V^\top W\big). \tag{5.28}$$

Hence $M_* \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ given by (5.24) is indeed the unique solution of the minimization problem because the map $\mathrm{D}^2\,\ell(M) \in \mathrm{S}^2\big((\mathbb{R}^{n\times n}_{\mathrm{sym}})^*\big)$ is positive definite for all $M \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ by (5.28). □

Inspired by [25, Sec. 6] and using Lemma 5.3, our aim is now to construct a metric $\widetilde{\langle \cdot,\cdot \rangle}^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*(\mathrm{St}_{n,k}/\mathrm{O}(k)))\big)$ by defining a map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ satisfying Assumption 4.3 such that $M_{X_*} \approx A - \frac{1}{k}\operatorname{tr}(X_*^\top A X_*)I_n$ holds at a non-degenerated local minimum $\mathrm{pr}(X_*) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ represented by $X_* \in \mathrm{St}_{n,k}$. As in [25, Sec. 6], let $\widehat{M} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ be some symmetric matrix approximating A, see Remark 5.5, Item 1 below for some specific choices, and define

$$\mathrm{St}_{n,k} \to \mathrm{SPD}(n), \quad X \mapsto M_X = \widehat{M} - \chi\big(\tfrac{1}{k}\operatorname{tr}(X^\top A X)\big)I_n, \tag{5.29}$$

where $\chi\colon \mathbb{R} \to \mathbb{R}$ is a smooth map which fulfills $\chi(x) < \lambda_{\min}(\widehat{M})$ for all $x \in \mathbb{R}$ and approximates the map $x \mapsto \min\big(x, \lambda_{\min}(\widehat{M})\big)$ smoothly. A specific construction for such a map χ is given in Remark 5.5, Item 2 below.

The next lemma shows that the construction proposed above leads indeed to a well-defined metric on $\mathrm{St}_{n,k}/\mathrm{O}(k)$.

Lemma 5.4 *Let $\widehat{M} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ be some symmetric matrix and let $\chi\colon \mathbb{R} \to \mathbb{R}$ be a smooth function fulfilling $\chi(x) < \lambda_{\min}(\widehat{M})$ for all $x \in \mathbb{R}$. Then, $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ given by (5.29) is well-defined and smooth. Moreover, it satisfies Assumption 4.3. In particular, it defines the submersion metric $\widetilde{\langle \cdot,\cdot \rangle}^M \in \Gamma^\infty\big(\mathrm{S}^2(T^*(\mathrm{St}_{n,k}/\mathrm{O}(k)))\big)$ on $\mathrm{St}_{n,k}/\mathrm{O}(k)$.*

Proof: Obviously, (5.29) yields a smooth map $\mathrm{St}_{n,k} \to \mathbb{R}^{n\times n}_{\mathrm{sym}}$. Moreover, by the assumption $\chi(x) < \lambda_{\min}(\widehat{M})$ for all $x \in \mathbb{R}$, we obtain for each $X \in \mathrm{St}_{n,k}$

$$\lambda_{\min}(M_X) = \lambda_{\min}(\widehat{M}) - \chi\big(\tfrac{1}{k}\operatorname{tr}(X^\top A X)\big) > 0 \tag{5.30}$$

proving that $M_X \in \mathrm{SPD}(n)$. Moreover, let $R \in \mathrm{O}(k)$. Then, the invariance property follows by $M_{XR} = \widehat{M} - \chi\big(\frac{1}{k}\operatorname{tr}((XR)^\top A(XR))\big)I_n = \widehat{M} - \chi\big(\frac{1}{k}\operatorname{tr}(X^\top A X)\big)I_n = M_X$ as desired. □

As Lemma 5.4 reveals, the map (5.29) yields a well-defined metric on $\mathrm{St}_{n,k}/\mathrm{O}(k)$. The next remark comments on some specific choices for $\widehat{M}$ and χ involved in the definition of (5.29).

Remark 5.5 1. By Lemma 5.4, in principle, $\widehat{M}$ can be an arbitrary symmetric matrix. However, in view of the discussion prior to (5.29), we propose to construct $\widehat{M} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ by using some preconditioning scheme for A, see e.g. [31], such that for all $a \in \mathbb{R}$ with $a < \lambda_{\min}(M)$ and $B \in \mathbb{R}^{n\times k}$, matrix products of the form $(\widehat{M} + aI_n)^{-1}B$ and $(\widehat{M} + aI_n)B$ can be computed efficiently.

2. We define $\chi = \chi_\lambda \colon \mathbb{R} \to \mathbb{R}$ motivated by the generalized soft-plus function, see [32, Eq. (1)]. For a fixed $\lambda = \lambda_{\min}(\widehat{M}) - \epsilon \in \mathbb{R}$, and some $s > 0$ and $\epsilon \geq 0$, we set

$$\chi_\lambda \colon \mathbb{R} \to \mathbb{R}, \quad x \mapsto \chi_\lambda(x) = -\tfrac{1}{s}\log(1 + \mathrm{e}^{-sx}) + \lambda. \tag{5.31}$$

In particular, because of $\mathrm{e}^x > 0$ for all $x \in \mathbb{R}$ and $\log(1 + x) > 0$ for all $x > 0$, the inequality $\chi_\lambda(x) < \lambda = \lambda_{\min}(\widehat{M}) - \epsilon \leq \lambda_{\min}(\widehat{M})$ is satisfied for all $x \in \mathbb{R}$, i.e. $\lambda_{\min}(\widehat{M}) - \chi_\lambda(x) > \epsilon \geq 0$.

Remark 5.6 In principle, Corollary 5.2 can be used to estimate the condition number of the Riemannian Hessian of the generalized Rayleigh quotient with respect to the metric $\widetilde{\langle\cdot,\cdot\rangle^M}$ on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ associated to (5.29). However, a detailed investigation in this direction is out of the scope of this paper.

5.4 A Geometric Conjugate Gradient Algorithm

In this section, we apply the geometric CG method from [1, Alg. 13] to the generalized Rayleigh quotient $f \colon \mathrm{St}_{n,k} \ni X \mapsto \frac{1}{2}\operatorname{tr}(X^\top AX) \in \mathbb{R}$ such that a geometric CG method on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ for minimizing $\check{f} \colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$ is induced. Here, the Riemannian metric $\langle\cdot,\cdot\rangle^M$ on $\mathrm{St}_{n,k}$ is defined via the smooth map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ from (5.29), where $\chi = \chi_\lambda \colon \mathbb{R} \to \mathbb{R}$ is given by Remark 5.5, Item 2. Moreover, we consider the following choices for $\widehat{M}$:

1. $\widehat{M} = A$;
2. $\widehat{M} = \operatorname{diag}(A)$ (known as Jaccobi preconditioning of A, see e.g. [31, Sec. 3.1]).

In addition, we consider the metric induced by the Euclidean metric on $\mathrm{St}_{n,k}$ which is given by $\langle\cdot,\cdot\rangle^M$ when $M_X = I_n$ for all $X \in \mathrm{St}_{n,k}$. Note that these choices of $\langle\cdot,\cdot\rangle^M$ on $\mathrm{St}_{n,k}$ induce the submersion metrics $\widetilde{\langle\cdot,\cdot\rangle^M}$ on $\mathrm{St}_{n,k}/\mathrm{O}(k)$. In the sequel, we consider specific differential geometric quantities required to apply [1, Alg. 13] to $\mathrm{St}_{n,k}/\mathrm{O}(k)$ endowed with the metric $\widetilde{\langle\cdot,\cdot\rangle^M}$. In more detail, a specific retraction and a specific vector transport on $\mathrm{St}_{n,k}$ are studied which induce a retraction and a vector transport on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ considered as a Riemannian quotient manifold of $(\mathrm{St}_{n,k}, \langle\cdot,\cdot\rangle^M)$.

Remark 5.7 We point out that the main focus of this subsection is on providing the differential geometric ingredients required for a geometric CG method on $(\mathrm{St}_{n,k}/\mathrm{O}(k), \widetilde{\langle\cdot,\cdot\rangle^M})$. Our specific choices lead to Algorithm 1 below. Studying the convergence properties of Algorithm 1 is out of the scope of this text. Nevertheless, a numerical experiment is performed in Section 5.4.5.

5.4.1 Retractions and Vector Transport

To perform a geometric CG algorithm as proposed in [1, Alg. 13], a retraction as well as an associated vector transport has to be chosen. These choices are specified in the sequel. For the retraction, we use

$$\mathrm{R}\colon T\mathrm{St}_{n,k} \to \mathrm{St}_{n,k}, \quad (X, V) \mapsto (X + V)(I_k + V^\top V)^{-1/2} \tag{5.32}$$

based on the polar decomposition, see e.g. [1, Ex. 4.1.3]. A calculation, see also [6, Eq.(9.9)], shows that for $X \in \mathrm{St}_{n,k}$, $V \in T_X\mathrm{St}_{n,k}$ and $R \in \mathrm{O}(k)$

$$\mathrm{R}_{XR}(VR) = (X + V)R\big(R^\top (I_k + V^\top V)^{-1/2} R\big) = \big(\mathrm{R}_X(V)\big)R \tag{5.33}$$

is satisfied. This property leads to the following lemma.

Lemma 5.8 *The map*

$$\check{\mathrm{R}}\colon T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \to \mathrm{St}_{n,k}/\mathrm{O}(k) \tag{5.34}$$

defined for $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ *represented by* $X \in \mathrm{St}_{n,k}$ *and* $\check{V} \in T_{\mathrm{pr}(X)}\mathrm{St}_{n,k}/\mathrm{O}(k)$ *by*

$$\check{\mathrm{R}}_{\mathrm{pr}(X)}(\check{V}) = \mathrm{pr}\left(\mathrm{R}_X\big(\overline{\check{V}}\big|_X\big)\right) \tag{5.35}$$

is a well-defined retraction on $\mathrm{St}_{n,k}/\mathrm{O}(k)$*.*

PROOF: Lemma 4.9, Claim 3 combined with [6, Thm. 9.32] yields the desired result. □

Next, in Proposition 5.10 below, we construct a vector-transport on $\mathrm{St}_{n,k}/\mathrm{O}(k)$ associated with the retraction $\check{\mathrm{R}}\colon T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ defined in (5.35) whose definition is inspired by [1, Sec. 8.1.4]. We refer to [1, Sec. 8.1] for the notion of vector transports. As a preparation, we need the following lemma.

Lemma 5.9 *Assume that* $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ *defining the metric* $\langle \cdot, \cdot \rangle^M$ *on* $\mathrm{St}_{n,k}$ *satisfies Assumption 4.3, i.e* $M_{XR} = M_X$ *for all* $X \in \mathrm{St}_{n,k}$ *and* $R \in \mathrm{O}(k)$*. Then, for* $X \in \mathrm{St}_{n,k}$*,* $V \in \mathbb{R}^{n\times k}$ *and* $R \in \mathrm{O}(k)$*, the orthogonal projection from Lemma 3.5 fulfills*

$$P_{XR}(VR) = \big(P_X(V)\big)R. \tag{5.36}$$

PROOF: Recall from Lemma 3.5 that $P_{XR}\colon \mathbb{R}^{n\times k} \to T_{XR}\mathrm{St}_{n,k}$ evaluated at $VR \in \mathbb{R}^{n\times k}$ is given by

$$P_{XR}(VR) = VR - 2M_{XR}^{-1}XR\Big(\varphi^{-1}_{XR,M_{XR}}\big(\mathrm{sym}(R^\top X^\top VR)\big)\Big). \tag{5.37}$$

Exploiting the assumption $M_{XR} = M_X$, one shows analogously to the proof of Proposition 4.6, Claim 2

$$\varphi^{-1}_{XR,M_{XR}}\big(\mathrm{sym}(R^\top X^\top VR)\big) = R^\top \Big(\varphi^{-1}_{X,M_X}\big(\mathrm{sym}(X^\top V)\big)\Big)R. \tag{5.38}$$

Plugging (5.38) into (5.37) yields the desired result. □

Proposition 5.10 *Let* $\mathrm{R}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}$ *be the retraction from* (5.32) *and define*

$$\begin{aligned} \tau\colon T\mathrm{St}_{n,k} \oplus T\mathrm{St}_{n,k} &\to \mathrm{Hor}(\mathrm{St}_{n,k}) \\ (V_X, W_X) &\mapsto \tau\big|_X(V_X, W_X) = \big(\mathcal{P}^{\mathrm{hor}}_{\mathrm{R}_X(V_X)} \circ P_{\mathrm{R}_X(V_X)}\big)(W_X), \end{aligned} \tag{5.39}$$

where the subscript X *in* V_X, W_X *indicates that* $V_X, W_X \in T_X\mathrm{St}_{n,k}$*. If this is clear from the context, we write* $V = V_X$ *and* $W = W_X$*, as usual. Using this notation, the following assertions are fulfilled:*

1. *Let* $X \in \mathrm{St}_{n,k}$*,* $V, W \in T_X\mathrm{St}_{n,k}$ *and* $R \in \mathrm{O}(k)$*. Then*

$$\tau\big|_{XR}(VR, WR) = \big(\tau\big|_X(V, W)\big)R \tag{5.40}$$

is fulfilled.

2. *Define*

$$\check{\tau}\colon T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \oplus T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \to T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \tag{5.41}$$

point-wise for tangent vectors $\check{V}, \check{W} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$ *at* $\mathrm{pr}(X) \in (\mathrm{St}_{n,k}/\mathrm{O}(k))$ *represented by* $X \in \mathrm{St}_{n,k}$ *via*

$$\check{\tau}\big|_{\mathrm{pr}(X)}\big(\check{V}, \check{W}\big) = \mathrm{D}\,\mathrm{pr}(X)\tau\big|_X\big(\overline{\check{V}}\big|_X, \overline{\check{W}}\big|_X\big). \tag{5.42}$$

Then (5.41) *is a well-defined vector transport on* $\mathrm{St}_{n,k}/\mathrm{O}(k)$*. Moreover, for* $\check{V}, \check{W} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$*, the horizontal lift of* $\check{\tau}\big(\check{V}, \check{W}\big)$ *at* $X \in \mathrm{St}_{n,k}$ *is given by*

$$\overline{\check{\tau}(\check{V}, \check{W})}\big|_X = \Big(\mathcal{P}^{\mathrm{hor}}_{\mathrm{R}_X(\overline{\check{V}}|_X)} \circ P_{\mathrm{R}_X(\overline{\check{V}}|_X)}\Big)\big(\overline{\check{W}}\big|_X\big). \tag{5.43}$$

Proof: We start with Claim 1. Using (5.33), i.e. $\mathrm{R}_{XR}(VR) = \big(\mathrm{R}_X(V)\big)R$, we compute

$$\begin{aligned}
\tau\big|_{XR}(VR, WR) &= \big(\mathcal{P}^{\mathrm{hor}}_{\mathrm{R}_{XR}(VR)} \circ P_{\mathrm{R}_{XR}(VR)}\big)(WR) \\
&= \Big(\mathcal{P}^{\mathrm{hor}}_{(\mathrm{R}_X(V))R}\Big(P_{(\mathrm{R}_X(V))R}(WR)\Big)\Big) \\
&= \Big(\mathcal{P}^{\mathrm{hor}}_{(\mathrm{R}_X(V))R}\Big(\big(P_{(\mathrm{R}_X(V))}(W)\big)R\Big)\Big) \\
&= \Big(\mathcal{P}^{\mathrm{hor}}_{(\mathrm{R}_X(V))}\Big(P_{(\mathrm{R}_X(V))}(W)\Big)\Big)R,
\end{aligned} \tag{5.44}$$

where the third equality follows from Lemma 5.9 and the last equality is fulfilled because of Lemma 4.9, Claim 1.

It remains to prove Claim 2. To prove that $\check{\tau}$ is well-defined, we compute for $X \in \mathrm{St}_{n,k}$, $R \in \mathrm{O}(k)$, and $\check{V}, \check{W} \in T_{\mathrm{pr}(X)}(\mathrm{St}_{n,k}/\mathrm{O}(k))$

$$\begin{aligned}
\check{\tau}\big|_{\mathrm{pr}(XR)}\big(\check{V}, \check{W}\big) &= \mathrm{D}\,\mathrm{pr}(XR)\tau\big|_{XR}\big(\overline{\check{V}}\big|_{XR}, \overline{\check{W}}\big|_{XR}\big) \\
&= \mathrm{D}\,\mathrm{pr}(XR)\tau\big|_{XR}\Big((\overline{\check{V}}\big|_X)R, (\overline{\check{W}}\big|_X)R\Big) \\
&= \mathrm{D}\,\mathrm{pr}(XR)\Big(\Big(\tau\big|_X\big(\overline{\check{V}}\big|_X, \overline{\check{W}}\big|_X\big)\Big)R\Big) \\
&= \mathrm{D}\,\mathrm{pr}(X)\tau\big|_X\big(\overline{\check{V}}\big|_X, \overline{\check{W}}\big|_X\big) \\
&= \check{\tau}\big|_{\mathrm{pr}(X)}\big(\check{V}, \check{W}\big),
\end{aligned} \tag{5.45}$$

where the second equality holds by Lemma 4.9, Claim 3 and the third equality relies on Claim 1. Equation (5.45) shows that $\check{\tau}$ is indeed well-defined, i.e. independent of $X \in \mathrm{St}_{n,k}$ representing $\mathrm{pr}(X) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$.

Inspired by the proof of [6, Thm. 9.32], we now show that $\check{\tau}$ is smooth. Let $\mathrm{pr}(X_0) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ and let $s\colon U \to \mathrm{St}_{n,k}$ be a local section of $\mathrm{pr}\colon \mathrm{St}_{n,k} \to \mathrm{St}_{n,k}/\mathrm{O}(k)$ defined on some open $U \subseteq \mathrm{St}_{n,k}/\mathrm{O}(k)$ with $\mathrm{pr}(X_0) \in U$. Using $\mathrm{pr} \circ s = \mathrm{id}_U$, we obtain for $\mathrm{pr}(X) \in U$ and $\check{V}, \check{V} \in T_{\mathrm{pr}(X)}U$

$$\check{\tau}\big|_{\mathrm{pr}(X)}(\check{V}, \check{W}) = \mathrm{D}\,\mathrm{pr}\,\big(s(\mathrm{pr}(X))\big)\tau\big|_{s(\mathrm{pr}(X))}\big(\overline{\check{V}}\big|_{s(\mathrm{pr}(X))}, \overline{\check{W}}\big|_{s(\mathrm{pr}(X))}\big) \tag{5.46}$$

proving that $\check{\tau}$ restricted to U is smooth because it is a composition of smooth maps. Repeating this argument for each $\mathrm{pr}(X_0) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$ shows that $\check{\tau}\colon T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \oplus T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \to T(\mathrm{St}_{n,k}/\mathrm{O}(k))$ is in fact a smooth map.

Now, it is straightforward to verify that $\check{\tau}$ is indeed a vector transport consistent with the retraction $\check{\mathrm{R}}\colon T(\mathrm{St}_{n,k}/\mathrm{O}(k)) \to \mathrm{St}_{n,k}/\mathrm{O}(k)$. This yields the desired result. □

5.4.2 The Algorithm

Next, we present Algorithm 1. It is obtained by applying [1, Alg. 13] to the generalized Rayleigh quotient on $\mathrm{St}_{n,k}$. The Steps 2a– 2d are explained in more detail below.

Algorithm 1 Geometric CG method applied to the generalized Rayleigh Quotient

Goal: Find a minimizer of $\check{f}\colon \mathrm{St}_{n,k}/\mathrm{O}(k) \to \mathbb{R}$.
Require: Riemannian metric defined by a smooth map $\mathrm{St}_{n,k} \ni X \mapsto M_X \in \mathrm{SPD}(n)$ fulfilling Assumption 4.3.
Input: Initial point $X_0 \in \mathrm{St}_{n,k}$ representing $\mathrm{pr}(X_0) \in \mathrm{St}_{n,k}/\mathrm{O}(k)$.

1. Set $\eta_0 = -\operatorname{grad} f(X_0)$.
2. **for** $k = 0, 1, \ldots$ **do**
   (a) Compute a step size $\alpha_k \in \mathbb{R}$ and set $X_{k+1} = \mathrm{R}_{X_k}(\alpha_k \eta_k)$.
   (b) Compute $\beta_{k+1} \in \mathbb{R}$ and set
   $$\eta_{k+1} = -\operatorname{grad} f(X_{k+1}) + \beta_{k+1} \mathcal{P}^{\mathrm{hor}}_{X_{k+1}}\big(P_{X_{k+1}}(\eta_k)\big). \tag{5.47}$$
   (c) If restart criterion is satisfied: Set $\eta_{k+1} = -\operatorname{grad} f(X_{k+1})$.
   (d) If stopping criterion is satisfied: Break.
3. **end for**

Output: Iterates $\{X_k\}$ in $\mathrm{St}_{n,k}$ representing iterates $\{\mathrm{pr}(X_k)\}$ in $\mathrm{St}_{n,k}/\mathrm{O}(k)$.

Notation 5.11 From now on, besides the k appearing in $\mathrm{St}_{n,k}$ and $\mathrm{O}(k)$, we also denote by k the index of the iterates of Algorithm 1. However, the meaning of k will always be clear from the context.

In Algorithm 1, $\operatorname{grad} f(X)$ with respect to $\langle\cdot,\cdot\rangle^M$ is evaluated at $X \in \mathrm{St}_{n,k}$ by using Lemma 3.6. Moreover, $\mathcal{P}^{\mathrm{hor}}_X$ and P_X are evaluated by using Corollary 4.7 and Lemma 3.5, respectively. Step 2a and Step 2b are detailed in Section 5.4.3 and Section 5.4.4, respectively, below. For the specific choices of Step 2c and Step 2d used in the numerical experiment, we refer to Section 5.4.5.

5.4.3 Step-Size Computation

Motivated by [18], see also [8,16], the step size α_k in Algorithm 1, Step 2a is computed by a one-dimensional Newton-step. Given the search direction $\eta_k \in \mathrm{Hor}(\mathrm{St}_{n,k})_k$, we set

$$\alpha_k = -\frac{\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \mathrm{R}_{X_k})(t\eta_k)\big|_{t=0}}{\left|\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f \circ \mathrm{R}_{X_k})(t\eta_k)\big|_{t=0}\right|}. \tag{5.48}$$

Here, the absolute value in the denominator of (5.48) appears to improve the global convergence as in [18, Eq. (31)].

To evaluate (5.48), we rely on an expression for $f \circ \mathrm{R}_{X_k}(t\eta_k)$ obtained in [2, Sec. 5.2]. In more detail, we write $X = X_k$ and $\eta = \eta_k \in T_{X_k}\mathrm{St}_{n,k}$ for short. Then, following [2, Sec. 5.2 and Sec. 3.2], let

$$\eta^\top \eta = S D_\beta S^\top \tag{5.49}$$

be a diagonalization of $\eta^\top \eta \in \mathbb{R}^{k\times k}_{\mathrm{sym}}$ with $D_\beta = \mathrm{diag}(\beta_1, \dots, \beta_k) \in \mathbb{R}^{k\times k}$ and $S \in \mathrm{O}(k)$. Next, set $\eta_S = \eta S$ and $X_S = XS$. Define the diagonal matrices

$$D_\alpha = \mathrm{diag}(X_S^\top A X_S), \quad D_\gamma = \mathrm{diag}(\eta_S^\top A \eta_S), \quad D_\zeta = \mathrm{diag}(\eta_S^\top A X_S). \tag{5.50}$$

Using these definitions, and denoting for $i \in \{1, \dots, k\}$ the diagonal entries by $\alpha_i = (D_\alpha)_{ii}$, $\beta_i = (D_\beta)_{ii}$, $\gamma_i = (D_\gamma)_{ii}$, and $\zeta_i = (D_\zeta)_{ii}$, respectively, one obtains for $t \in \mathbb{R}$

$$(f \circ \mathrm{R}_X)(t\eta) = \tfrac{1}{2} \operatorname{tr}\Big(\big(I_k + t^2 D_\beta\big)^{-1}\big(D_\alpha + 2tD_\zeta + t^2 D_\gamma\big)\Big) = \tfrac{1}{2}\sum_{i=1}^{k} \tfrac{\alpha_i + 2\zeta_i t + \gamma_i t^2}{1+\beta_i t^2} \tag{5.51}$$

very similar to [2, Eq. (5.3) and Eq. (5.5)], see also [2, Eq. (5.6)] for its derivative with respect to t. Using (5.51), a straightforward computation yields

$$\tfrac{\mathrm{d}}{\mathrm{d}t}(f \circ \mathrm{R}_X)(t\eta)\big|_{t=0} = \operatorname{tr}(D_\zeta) \quad \text{and} \quad \tfrac{\mathrm{d}^2}{\mathrm{d}t^2}(f \circ \mathrm{R}_X)(t\eta)\big|_{t=0} = \operatorname{tr}(D_\gamma - D_\alpha D_\beta). \tag{5.52}$$

Thus, plugging (5.52) into (5.48), the step-size α_k in Algorithm 1, Step 2a is computed by

$$\alpha_k = -\frac{\operatorname{tr}(D_\zeta)}{\left|\operatorname{tr}\big(D_\gamma - D_\alpha D_\beta\big)\right|}. \tag{5.53}$$

Moreover, the retraction from (5.32) needs to be evaluated at $\alpha_k\eta_k$ in Algorithm 1, Step 2a. As in [2, Eq. (5.2)], using the diagonalization of $\eta_k^\top \eta_k = S D_\beta S^\top$ from (5.49), this is done via the formula

$$X_{k+1} = \mathrm{R}_{X_k}(\alpha_k \eta_k) = (X_k + \alpha_k \eta_k) S\big(I_k + \alpha_k^2 D_\beta\big)^{-1/2} S^\top. \tag{5.54}$$

5.4.4 Computation of β_{k+1}

In Step 2b of Algorithm 1, the value of β_{k+1} is computed using the Polak-Ribiere formula, see e.g. [1, Eq. (8.29)]. In our specific case, this leads to

$$\begin{aligned}\beta_{k+1} &= \frac{\big\langle \operatorname{grad} f(X_{k+1}), \operatorname{grad} f(X_{k+1}) - \mathcal{P}^{\mathrm{hor}}_{X_{k+1}}(P_{X_{k+1}}(\operatorname{grad} f(X_k)))\big\rangle^M_{X_{k+1}}}{\big\langle \operatorname{grad} f(X_k), \operatorname{grad} f(X_k)\big\rangle^M_{X_k}} \\ &= \frac{\operatorname{tr}\Big(\operatorname{grad} f(X_{k+1})^\top M_{X_{k+1}}\big(\operatorname{grad} f(X_{k+1}) - \operatorname{grad} f(X_{X_k})\big)\Big)}{\operatorname{tr}\Big(\operatorname{grad} f(X_k)^\top M_{X_k} \operatorname{grad} f(X_k)\Big)},\end{aligned} \tag{5.55}$$

where, similar to [2, Sec. 5.1], the second equality follows from the definition of the metric $\langle\cdot,\cdot\rangle^M$ and exploits that $\operatorname{grad} f(X_{k+1}) \in \mathrm{Hor}(\mathrm{St}_{n,k})_{X_{k+1}}$ is satisfied.

5.4.5 Numerical Experiments

In this section we present some numerical experiments using GNU Octave, see [9]. The result of Algorithm 1 for minimizing the generalized Rayleigh quotient with respect to different Riemannian metrics as defined at the beginning of Section 5.4 above, are shown in Figure 1. Although our implementation is not optimized, we make the following comments:

1. Algorithm 1 needs only on multiplication of A by an $(n \times k)$-matrix in each iteration although both matrices, AX_k (for evaluating $\operatorname{grad} f(X_k)$) and $A\eta_k$, see (5.50), are required in each iteration. Indeed, the observation from [2, Eq. (6.3)] can applied to Algorithm 1, as well. Given AX_k, one can compute AX_{k+1} by using the right-hand side of the following equation which follows by multiplying (5.54) by A from the left

$$AX_{k+1} = (AX_k + \alpha_k A\eta_k) S \big(I_k + \alpha_k^2 D_\beta\big)^{-1/2} S^\top. \tag{5.56}$$

2. In each iteration, Algorithm 1 needs only two matrix multiplications of $M_{X_k}^{-1}$ by an $(n \times k)$-matrix, namely $M_{X_k}^{-1} X_k$ and $M_{X_k}(AX_k)$. Indeed, recall that $P_X \colon \mathbb{R}^{n\times k} \to T_X \mathrm{St}_{n,k}$ from Lemma 3.5 is given by

$$P_X \colon \mathbb{R}^{n\times k} \to T_X \mathrm{St}_{n,k}, \quad V \mapsto V - 2M_X^{-1} X \Big(\varphi_{X,M_X}^{-1}\big(\operatorname{sym}(X^\top V)\big)\Big) \tag{5.57}$$

and for $f \colon \mathrm{St}_{n,k} \ni \mapsto \frac{1}{2}\operatorname{tr}(X^\top AX) \in \mathbb{R}$, the gradient given by Lemma 3.6 reads as

$$\operatorname{grad} f(X) = M_X^{-1}AX - 2M_X^{-1} X \Big(\varphi_{X,M_X}^{-1}\big(\operatorname{sym}(X^\top M_X^{-1}AX)\big)\Big). \tag{5.58}$$

By Lemma 3.4, $S = \varphi_{X,M_X}^{-1}(\operatorname{sym}(X^\top V))$ in (5.57) and $S = \varphi_{X,M_X}^{-1}(\operatorname{sym}(X^\top M_X^{-1}AX))$ in (5.58) are the unique solutions of $(X^\top M_X^{-1} X) S + S(X^\top M_X^{-1} X) = T$, where $T = \operatorname{sym}(X^\top V)$ and $T = \operatorname{sym}(X^\top M_X^{-1}AX)$, respectively.

3. Algorithm 1 needs only one matrix multiplication of M_X by an $(n \times k)$-matrix in each iteration. Indeed, using for $X \in \mathrm{St}_{n,k}$ and $V \in T_X\mathrm{St}_{n,k}$ the equality $\operatorname{skew}(X^\top M_X V) = -\operatorname{skew}(V^\top M_X X)$, the projection onto the horizontal bundle obtained in Corollary 4.7 can be rewritten as

$$\mathcal{P}_X^{\mathrm{hor}} \colon T_X\mathrm{St}_{n,k} \to \operatorname{Hor}(\mathrm{St}_{n,k})_X, \quad V \mapsto V + 2X\Big(\phi_{X,M_X}^{-1}\big(\operatorname{skew}(V^\top M_X X)\big)\Big), \tag{5.59}$$

where $A = \phi_{X,M_X}^{-1}\big(\operatorname{skew}(V^\top M_X X)\big) \in \mathfrak{so}(k)$ is the unique solution of the Sylvester equation $(X^\top M_X X)A + A(X^\top M_X X) = \operatorname{skew}(V^\top M_X X)$ according to Lemma 4.5.

4. We also point out that the computation of β_k detailed in Section 5.4.4 needs no further multiplication of the matrices A, M_X or M_X^{-1}. In fact, slightly rewriting (5.55), we obtain

$$\beta_k = \frac{\operatorname{tr}\Big(\big(M_{X_{k+1}} \operatorname{grad} f(X_{k+1})\big)^\top \big(\operatorname{grad} f(X_{k+1}) - \operatorname{grad} f(X_{X_k})\big)\Big)}{\operatorname{tr}\Big(\big(M_{X_k} \operatorname{grad} f(X_k)\big)^\top M \operatorname{grad} f(X_k)\Big)}. \tag{5.60}$$

Next, by multiplying $\operatorname{grad} f(X_k)$ given by (5.58) from the left by M_{X_k}, we obtain

$$M_{X_k} \operatorname{grad} f(X_k) = AX_k - 2X_k\Big(\varphi_{X_k,M_{X_k}}^{-1}\big(\operatorname{sym}(X_k^\top M_{X_k}^{-1} AX_k)\big)\Big). \tag{5.61}$$

Clearly, (5.61) is still satisfied if every index k is replaced by $k+1$. Thus, plugging (5.61) into (5.60) proves the claim concerning the computation of β_{k+1}.

After these comments, we now present the results of a numerical experiment, where the restart criterion, Algorithm 1, Step 2c, is chosen to be satisfied whenever $\dim(\mathrm{Gr}_{n,k}) = n(n-k)$ divides the number of iterations. Moreover, the function $\chi_\lambda \colon \mathbb{R} \to \mathbb{R}$ given in Remark 5.5, Item 2, is used with the parameters $s = 10$ and $\lambda = \lambda_{\min}(\widehat{M}) - \epsilon$ with $\epsilon = 10^{-4}$.

Concerning stopping criterion, Algorithm 1, Step 2d, the algorithm is stopped after 350 iterations or before, if $\| \operatorname{grad} f(X_{k+1}) \|_{X_{k+1}} < 10^{-5}$ is satisfied, where $\| \cdot \|_{X_{k+1}}$ is induced by the scalar product $\langle \cdot, \cdot \rangle^{M}_{X_{k+1}}$.

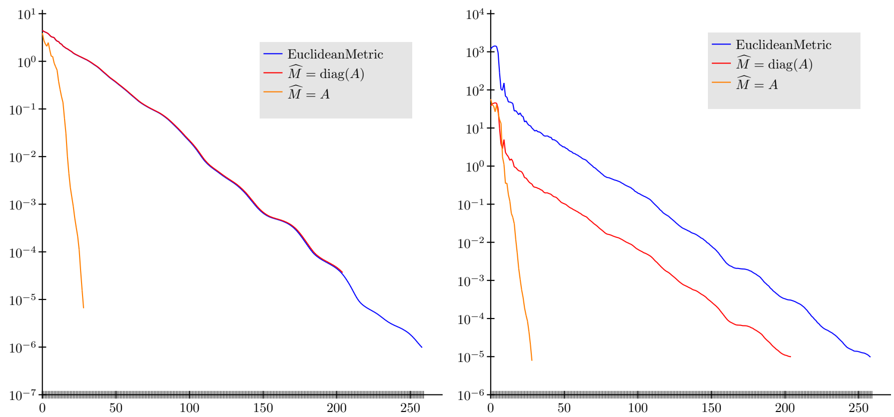

Figure 1: Distance (left-hand side) to the known solution and Riemannian norm of the Riemannian gradients (right-hand side) of the iterates produced by Algorithm 1 using different metrics on $\mathrm{St}_{1000,10}/\mathrm{O}(10)$. The matrix $A \in \mathbb{R}^{1000 \times 1000}$ used for the definition of the generalized Rayleigh quotient is similar to $D = \mathrm{diag}(1, \ldots, 1000) \in \mathbb{R}^{1000 \times 1000}$. In more detail, $A = QDQ^{\top}$, where Q is the Q-factor of the QR-decomposition of a randomly generated matrix.

Acknowledgments

The first author has been supported by the German Federal Ministry of Education and Research (BMBF-Projekt, funding numbers: 05M20WWA and 05M20WBA, Verbundprojekt 05M2020 - DyCA). The work of the second author has been supported by Fundação para a Ciência e Tecnologia (FCT) under the project UIDB/00048/2020 (`https://doi.org/10.54499/UIDB/00048/2020`). The last author has been supported by the German Federal Ministry of Education and Research (BMBF-Projekt 05M20WWA: Verbundprojekt 05M2020 - DyCA). Part of this research was accomplished while the first author visited the Department of Mathematics, University of Coimbra.

References

[1] Absil, P.-A., Mahony, R., Sepulchre, R. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, Princeton, NJ, 2008.

[2] Alimisis, F., Saad, Y., Vandereycken, B. *Gradient-Type Subspace Iteration Methods for the Symmetric Eigenvalue Problem.* SIAM J. Matrix Anal. Appl., **45**(4) (2024), 2360–2386. `https://doi.org/10.1137/23M1590792`.

[3] Bendokat, T., Zimmermann, R., Absil, P.-A. *A Grassmann manifold handbook: basic geometry and computational aspects.* Adv. Comput. Math., **50**(1) (2024), Paper No. 6, 1–51. `https://doi.org/10.1007/s10444-023-10090-8`.

[4] Bernstein, D. S. *Matrix mathematics.* Princeton University Press, Princeton, NJ, second edition, 2009. `https://doi.org/10.1515/9781400833344`.

[5] Bloch, A. M., Crouch, P. E., Sanyal, A. K. *A variational problem on Stiefel manifolds*. Nonlinearity, **19**(10) (2006), 2247–2276. https:/doi.org/10.1088/0951-7715/19/10/002.

[6] Boumal, N. *An introduction to optimization on smooth manifolds*. To appear with Cambridge University Press, 2022.

[7] Conway, J. H., Hardin, R. H., Sloane, N. J. A. *Packing lines, planes, etc.: packings in Grassmannian spaces*. Experiment. Math., **5**(2) (1996), 139–159.

[8] Curtef, O., Dirr, G., Helmke, U. *Riemannian optimization on tensor products of Grassmann manifolds: applications to generalized Rayleigh-quotients*. SIAM J. Matrix Anal. Appl., **33**(1) (2012), 210–234. https:/doi.org/10.1137/100792032.

[9] Eaton, J. W., Bateman, D., Hauberg, S., Wehbring, R. *GNU Octave version 8.4.0 manual: a high-level interactive language for numerical computations*, 2023. https://www.gnu.org/software/octave/doc/v8.4.0/.

[10] Edelman, A., Arias, T. A., Smith, S. T. *The Geometry of Algorithms with Orthogonality Constraints*. SIAM J. Matrix Anal. Appl., **20**(2) (1998), 303–353. https:/doi.org/10.1137/S0895479895290954.

[11] Fujii, K. *Introduction to Grassmann manifolds and quantum computation*. J. Appl. Math., **2**(8) (2002), 371–405. https:/doi.org/10.1155/S1110757X02110163.

[12] Gallier, J., Quaintance, J. *Differential Geometry and Lie Groups—A Second Course*, volume 13 of *Geometry and Computing*. Springer, Cham, 2020.

[13] Harandi, M., Hartley, R., Salzmann, M., Trumpf, J. *Dictionary learning on Grassmann manifolds*. In: *Algorithmic advances in Riemannian geometry and applications*, Adv. Comput. Vis. Pattern Recognit., pages 145–172. Springer, Cham, 2016.

[14] Helmke, U., Hüper, K., Trumpf, J. *Newton's method on Graßmann manifolds*, 2007. https://arxiv.org/abs/0709.2205.

[15] Helmke, U., Moore, J. B. *Optimization and Dynamical Systems*. Communications and Control Engineering Series. Springer, London, 1994.

[16] Hüper, K., Kleinsteuber, M., Shen, H. *Averaging complex subspaces via a Karcher mean approach*. Signal Processing, **93**(2) (2013), 459–467. https:/doi.org/https://doi.org/10.1016/j.sigpro.2012.08.020.

[17] Hüper, K., Markina, I., Silva Leite, F. *A Lagrangian approach to extremal curves on Stiefel manifolds*. J. Geom. Mech., **13**(1) (2021), 55–72. https:/doi.org/10.3934/jgm.2020031.

[18] Kleinsteuber, M., Hüper, K. *An Intrinsic CG Algorithm for Computing Dominant Subspaces*. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 4, pages IV–1405–IV–1408. 2007. https:/doi.org/10.1109/ICASSP.2007.367342.

[19] Knyazev, A. V. *Toward the Optimal Preconditioned Eigensolver: Locally Optimal Block Preconditioned Conjugate Gradient Method*. SIAM Journal on Scientific Computing, **23**(2) (2001), 517–541. https:/doi.org/10.1137/S1064827500366124.

[20] Lancaster, P., Rodman, L. *Algebraic Riccati Equations*. Oxford Science Publications. Oxford University Press, New York, 1995.

[21] Lee, J. M. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[22] Lee, J. M. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2018.

[23] Michor, P. W. *Topics in Differential Geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[24] Mishra, B., Sepulchre, R. *Riemannian preconditioning*. SIAM Journal on Optimization, **26**(1) (2016), 635–660. `https:/doi.org/10.1137/140970860`.

[25] Mor, U., Shustin, B., Avron, H. *Solving trust region subproblems using Riemannian optimization*. Numer. Math., **154**(1-2) (2023), 1–33. `https:/doi.org/10.1007/s00211-023-01360-0`.

[26] O'Neill, B. *Semi-Riemannian Geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press, New York, 1983.

[27] Parlett, B. N. *The Symmetric Eigenvalue Problem*, volume 20 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. `https:/doi.org/10.1137/1.9781611971163`.

[28] Shao, N., Chen, W. *Riemannian acceleration with preconditioning for symmetric eigenvalue problems*. Numer. Math., **157**(1) (2025), 307–354. `https:/doi.org/10.1007/s00211-025-01451-0`.

[29] Shustin, B., Avron, H. *Riemannian optimization with a preconditioning scheme on the generalized Stiefel manifold*. J. Comput. Appl. Math., **423** (2023), Paper No. 114953, 26. `https:/doi.org/10.1016/j.cam.2022.114953`.

[30] Turaga, P., Veeraraghavan, A., Chellappa, R. *Statistical analysis on Stiefel and Grassmann Manifolds with applications in Computer Vision*. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. 2008. `https:/doi.org/10.1109/CVPR.2008.4587733`.

[31] Wathen, A. J. *Preconditioning*. Acta Numer., **24** (2015), 329–376. `https:/doi.org/10.1017/S0962492915000021`.

[32] Wiemann, P. F. V., Kneib, T., Hambuckers, J. *Using the softplus function to construct alternative link functions in generalized linear models and beyond*. Statist. Papers, **65**(5) (2024), 3155–3180. `https:/doi.org/10.1007/s00362-023-01509-x`.