An Extrinsic Look at the Riemannian Hessian

P.-A. Absil[1], Robert Mahony[2], and Jochen Trumpf[2]

[1] Department of Mathematical Engineering, ICTEAM Institute,
Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
http://sites.uclouvain.be/absil/

[2] Research School of Engineering, Australian National University

Abstract. Let f be a real-valued function on a Riemannian submanifold of a Euclidean space, and let $\bar{f}$ be a local extension of f. We show that the Riemannian Hessian of f can be conveniently obtained from the Euclidean gradient and Hessian of $\bar{f}$ by means of two manifold-specific objects: the orthogonal projector onto the tangent space and the Weingarten map. Expressions for the Weingarten map are provided on various specific submanifolds.

Keywords: Riemannian Hessian, Euclidean Hessian, Weingarten map, shape operator.

1 Introduction

This paper concerns optimization methods on Riemannian manifolds that make explicit use of second-order information. This research area is motivated by various problems in the sciences and engineering that can be formulated as optimizing a real-valued function defined on a Riemannian manifold (see, e.g., [20,16,17,12,7] for some recently considered applications), and by the well-known fact that second-order methods tend to have the edge over first-order methods in situations where an accurate solution is sought or when the Hessian gets ill conditioned (see [1] for a recent example).

The archetypical second-order optimization method is Newton's method, of which several generalizations have been proposed on manifolds. Most of them fit in the framework given in [19,5] and [2, Alg. 5]. Besides a smooth real-valued function f defined on a Riemannian manifold $\mathcal{M}$, the ingredients of the Riemannian Newton method [2, Alg. 5] are an affine connection ∇ on $\mathcal{M}$ and a retraction R on $\mathcal{M}$. Turning the Riemannian Newton method into a successful numerical algorithm relies much on choosing ∇ and R and on computing them efficiently.

A retraction R on $\mathcal{M}$ can be viewed as a tool that turns a tangent update vector into a new iterate on $\mathcal{M}$. Retractions have been given particular attention in the recent literature, in general [3] and also specifically for the important cases where $\mathcal{M}$ is the Stiefel manifold of orthonormal matrices [15,21,13] or the manifold of fixed-rank matrices [20,18].

F. Nielsen and F. Barbaresco (Eds.): GSI 2013, LNCS 8085, pp. 361–368, 2013.
© Springer-Verlag Berlin Heidelberg 2013

As for the affine connection ∇, it is instrumental in the definition of the Hessian operator of f on $\mathcal{M}$. Namely, for all $x \in \mathcal{M}$ and all z in the tangent space $\mathrm{T}_x\mathcal{M}$, one defines

$$\operatorname{Hess} f(x)[z] := \nabla_z \operatorname{grad} f \quad \in \mathrm{T}_x\mathcal{M}. \tag{1}$$

While the convergence analysis of the Riemannian Newton method in [2, §6.3] provides for using any affine connection, a natural choice for ∇ is the uniquely defined Riemannian connection, also termed Levi-Civita connection or canonical connection.

In this paper, for the case where $\mathcal{M}$ is a Riemannian submanifold of a Euclidean space $\mathcal{E}$ (examples can be found in Section 4) and where ∇ is chosen to be the Riemannian connection, we give a formula for the Hessian (1) that relies solely on four objects: (i) the classical gradient $\partial \bar{f}(x)$ of a smooth extension $\bar{f}$ of f in a neighborhood of $\mathcal{M}$ in $\mathcal{E}$, (ii) the classical Hessian $\partial^2 \bar{f}(x)$ of $\bar{f}$, (iii) the orthogonal projector $\mathcal{P}_x$ onto $\mathrm{T}_x\mathcal{M}$, (iv) the *Weingarten map* $\mathfrak{A}_x$, also called *shape operator*. (The symbol $\mathfrak{A}$ is "A" in Fraktur font.) We provide expressions for $\mathcal{P}_x$ and $\mathfrak{A}_x$ on some important Riemannian submanifolds. These expressions yield a formula for the Riemannian Hessian where f is involved only through the classical gradient and Hessian, $\partial \bar{f}(x)$ and $\partial^2 \bar{f}(x)$. These results can be exploited in various Riemannian optimization schemes, such as Newton's method or trust-region methods, where the knowledge of the Hessian is either mandatory or potentially beneficial.

The paper is organized as follows. Section 2 recalls in more details the definition of the Riemannian Hessian on submanifolds of Euclidean spaces. Section 3 lays out the relation between the Riemannian Hessian and the Weingarten map. Finally, section 4 provides formulas for the Weingarten map on several specific manifolds.

An early version of Sections 2 and 3 of this paper can be found in section 6 of the technical report [4].

2 The Riemannian Hessian on Submanifolds

Let $\mathcal{M}$ be a d-dimensional Riemannian submanifold of an n-dimensional Euclidean space $\mathcal{E}$; see, e.g., [2, §3.6.1] or [9, §2.A.3] for details. Let x_0 be a point of $\mathcal{M}$, let f be a smooth real-valued function on $\mathcal{M}$ around x_0, and let $\bar{f}$ be a smooth extension of f to a neighborhood $\mathcal{U}$ of x_0 in $\mathcal{E}$.

For all $x \in \mathcal{M}$, we let $\partial \bar{f}(x)$ and $\partial^2 \bar{f}(x)$ denote the (Euclidean) gradient and (Euclidean) Hessian of $\bar{f}$ at x. In coordinates, we have

$$\partial \bar{f}(x) = \left[\partial_1 \bar{f}(x) \ \dots \ \partial_n \bar{f}(x)\right]^{\mathrm{T}}$$

and

$$[\partial^2 \bar{f}(x)]_{ij} = \partial_{ij} \bar{f}(x), \quad i, j = 1, \dots, n.$$

We also let $\mathcal{P}_x$ denote the orthogonal projector onto $\mathrm{T}_x\mathcal{M}$, defined by

$$\mathcal{P}_x : \mathrm{T}_x\mathcal{E} \simeq \mathcal{E} \ \rightarrow \ \mathrm{T}_x\mathcal{M} \ : \ \xi \mapsto \mathcal{P}_x(\xi) \tag{2}$$

with $\langle \xi - \mathcal{P}_x(\xi), \zeta \rangle = 0$ for all $\zeta \in \mathrm{T}_x\mathcal{M}$. Examples will be given in Section 4. Once an orthonormal basis is chosen for $\mathcal{E}$, $\mathcal{P}_x$ is represented as a (symmetric) matrix; hence $\mathcal{P}$ can be viewed as a matrix-valued function on $\mathcal{M}$. For any function F on $\mathcal{M}$ into a vector space, and for any $z \in \mathrm{T}_x\mathcal{M}$, we let

$$\mathrm{D}_z F = \lim_{t \to 0} F(\gamma(t)),$$

where γ is any curve on $\mathcal{M}$ with $\gamma(0) = x$ and $\gamma'(0) = z$.

We have

$$\operatorname{grad} f(x) = \mathcal{P}_x \partial \bar{f}(x), \tag{3}$$

where $\operatorname{grad} f(x)$ denotes the (Riemannian) gradient of f at x; see [2, §3.6.1] for details. Moreover, letting ∇ denote the Riemannian connection on $\mathcal{M}$, we have that $\operatorname{Hess} f(x)$, the Riemannian Hessian of f at x, is the linear transformation of $\mathrm{T}_x\mathcal{M}$ defined, for all $z \in \mathrm{T}_x\mathcal{M}$, by

$$\begin{aligned} \operatorname{Hess} f(x)[z] &= \nabla_z \operatorname{grad} f && (4)\\ &= \mathcal{P}_x \mathrm{D}_z (\operatorname{grad} f) && (5)\\ &= \mathcal{P}_x \mathrm{D}_z \left(\mathcal{P} \partial \bar{f}\right) && (6)\\ &= \mathcal{P}_x \partial^2 \bar{f}(x) z + \mathcal{P}_x \mathrm{D}_z \mathcal{P} \partial \bar{f}(x). && (7) \end{aligned}$$

Equation (4) is the definition (1). Equation (5) comes from the classical expression of the Riemannian connection on a Riemannian submanifold of a Euclidean space; see, e.g., [2, §5.3.3] or [9, §2.B.2]. Equation (6) follows from (3). Finally, (7) is an application of the product rule, observing that $\mathcal{P}$ is a matrix-valued function, $\partial \bar{f}$ a vector-valued function, and $\mathcal{P}_x \mathcal{P}_x = \mathcal{P}_x$ since $\mathcal{P}_x$ is a projector.

Expression (7) features the four ingredients alluded to in the introduction, namely $\partial \bar{f}(x)$, $\partial^2 \bar{f}(x)$, $\mathcal{P}_x$, $\mathcal{P}_x \mathrm{D}_z \mathcal{P}$. The rest of this paper is devoted to establishing the relation of $\mathcal{P}_x \mathrm{D}_z \mathcal{P}$ with the Weingarten map and to working out formulas for $\mathcal{P}_x \mathrm{D}_z \mathcal{P}$ on various specific Riemannian submanifolds.

3 The Riemannian Hessian and the Weingarten Map

We are thus concerned with $\mathcal{P}_x \mathrm{D}_z \mathcal{P}$, where $z \in \mathrm{T}_x\mathcal{M}$. In this section, we establish a relation (8) between $\mathcal{P}_x \mathrm{D}_z \mathcal{P}$ and the Weingarten map, defined next. This relation does not seem to have been previously pointed out in the literature, but it is present in the technical report [4].

Definition 1 (Weingarten map). *The* Weingarten map *of the submanifold $\mathcal{M}$ at x is the operator $\mathfrak{A}_x$ that takes as arguments a tangent vector $z \in \mathrm{T}_x\mathcal{M}$ and a normal vector $v \in \mathrm{T}_x^\perp\mathcal{M}$ and returns the tangent vector*

$$\mathfrak{A}_x(z, v) = -\mathcal{P}_x \mathrm{D}_z V,$$

where V is any local extention of v to a normal vector field on $\mathcal{M}$.

It is known [6, Prop. II.2.1] that $\mathcal{P}_x \mathrm{D}_z V$ does not depend on the choice of the extension V, and this makes the above definition valid. The next result confirms this fact and gives an alternate expression of $\mathfrak{A}_x(z, v)$. Let

$$\mathcal{P}_x^\perp = I - \mathcal{P}_x$$

denote the orthogonal projector onto the normal space to $\mathcal{M}$ at x. It is useful to keep in mind that, in our convention, D applies only to the expression that directly follows: $\mathrm{D}_z FG = (\mathrm{D}_z F)G \neq \mathrm{D}_z(FG)$.

Theorem 1. *The Weingarten map $\mathfrak{A}_x$ satisfies*

$$\mathfrak{A}_x(z, \mathcal{P}_x^\perp u) = \mathcal{P}_x \mathrm{D}_z \mathcal{P} u = \mathcal{P}_x \mathrm{D}_z \mathcal{P} \mathcal{P}_x^\perp u, \tag{8}$$

for all $x \in \mathcal{M}$, $z \in \mathrm{T}_x\mathcal{M}$, and $u \in \mathrm{T}_x\mathcal{E} \simeq \mathcal{E}$.

Proof. We first show that

$$\mathcal{P}_x \mathrm{D}_z \mathcal{P} = \mathcal{P}_x \mathrm{D}_z \mathcal{P} \mathcal{P}_x^\perp, \tag{9}$$

which takes care of the second equality in (8). Since $\mathcal{P}\mathcal{P}^\perp = 0$, we have $0 = \mathrm{D}_z \mathcal{P} \mathcal{P}_x^\perp + \mathcal{P}_x \mathrm{D}_z \mathcal{P}^\perp = \mathrm{D}_z \mathcal{P}\mathcal{P}_x^\perp - \mathcal{P}_x \mathrm{D}_z \mathcal{P}$. It follows that $\mathcal{P}_x \, \mathrm{D}_z \mathcal{P} \, \mathcal{P}_x = 0$. Hence, since $\mathcal{P}_x + \mathcal{P}_x^\perp = I$, we have $\mathcal{P}_x \mathrm{D}_z \mathcal{P} = \mathcal{P}_x \mathrm{D}_z \mathcal{P}(\mathcal{P}_x + \mathcal{P}_x^\perp) = \mathcal{P}_x \, \mathrm{D}_z \mathcal{P} \, \mathcal{P}_x + \mathcal{P}_x \, \mathrm{D}_z \mathcal{P} \, \mathcal{P}_x^\perp = \mathcal{P}_x \, \mathrm{D}_z \mathcal{P} \, \mathcal{P}_x^\perp$, and the claim (9) is proven.

For the first equality in (9), we have, for all extension U of u,

$$-\mathcal{P}_x \mathrm{D}_z(\mathcal{P}^\perp U) = -\mathcal{P}_x \mathrm{D}_z \mathcal{P}^\perp U - \mathcal{P}_x \mathcal{P}_x^\perp \mathrm{D}_z U = -\mathcal{P}_x \mathrm{D}_z \mathcal{P}^\perp U = \mathcal{P}_x \mathrm{D}_z \mathcal{P} U.$$

This concludes the proof.

A consequence of Theorem 1 for the Riemannian Hessian expression (7) is that $\mathcal{P}_x \mathrm{D}_z \mathcal{P} \partial \bar{f}(x) = \mathcal{P}_x \mathrm{D}_z \mathcal{P} \mathcal{P}_x^\perp \partial \bar{f}(x) = \mathfrak{A}_x(z, \mathcal{P}_x^\perp \partial \bar{f}(x))$. Observe in particular that $\mathcal{P}_x \, \mathrm{D}_z \mathcal{P} \, \partial \bar{f}(x)$ depends on $\partial \bar{f}(x)$ only through is normal component $\mathcal{P}_x^\perp \partial \bar{f}(x)$. In summary we have obtained the expression

$$\operatorname{Hess} f(x)[z] = \mathcal{P}_x \partial^2 \bar{f}(x) z + \mathfrak{A}_x(z, \mathcal{P}_x^\perp \partial \bar{f}). \tag{10}$$

4 Projector and Weingarten Map on Specific Manifolds

We now present formulas for the projector $\mathcal{P}$ and the Weingarten map $\mathfrak{A}$ on various specific manifolds. All the formulas provided for $\mathcal{P}$ and most—but apparently not all—of those provided for $\mathfrak{A}$ can be found in the literature.

4.1 The Stiefel Manifold

The *Stiefel manifold* of orthonormal p-frames in $\mathbb{R}^n$, denoted by $\mathrm{St}(p, n)$, is the submanifold of the Euclidean space $\mathbb{R}^{n \times p}$ defined by

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^\mathrm{T} X = I_p\},$$

where I_p stands for the identity matrix of size p. We point out that the Riemannian metric obtained on $\mathrm{St}(p,n)$ by making it a Riemannian submanifold of $\mathbb{R}^{n\times p}$ is different from the canonical metric mentioned in [8, §2.3.1]. The orthogonal projector $\mathcal{P}_X$ onto $\mathrm{T}_X\mathrm{St}(p,n)$ is given by

$$\begin{aligned}\mathcal{P}_X U &= (I - XX^{\mathrm{T}})U + X\frac{1}{2}(X^{\mathrm{T}}U - U^{\mathrm{T}}X)\\ &= U - X\frac{1}{2}(X^{\mathrm{T}}U + U^{\mathrm{T}}X);\end{aligned}$$

see, e.g., [2, §3.6.1].

Let $Z \in \mathrm{T}_X\mathcal{M}$ and $V \in \mathrm{T}_X^{\perp}\mathcal{M}$. Hence $V = XS$ with $S = S^{\mathrm{T}}$ and $Z = X_{\perp}K + X\Omega$ where $\Omega = -\Omega^{\mathrm{T}}$, K is an arbitary $(n-p)\times p$ matrix, and $X_{\perp}$ is an orthonormal $n\times(n-p)$ matrix such that $X^{\mathrm{T}}X_{\perp} = 0$; see [2, §3.6.1] for details. We have

$$\mathcal{P}_X \mathrm{D}_Z \mathcal{P} V = \mathcal{P}_X\left(V - Z\frac{1}{2}(X^{\mathrm{T}}V + V^{\mathrm{T}}X) - X\frac{1}{2}(Z^{\mathrm{T}}V + V^{\mathrm{T}}Z)\right).$$

Since V and $X\frac{1}{2}(Z^{\mathrm{T}}V + V^{\mathrm{T}}Z)$ belong to the normal space $\mathrm{T}_X^{\perp}\mathrm{St}(p,n)$, and since $\frac{1}{2}(X^{\mathrm{T}}V + V^{\mathrm{T}}X) = S$, we are left with

$$\begin{aligned}\mathcal{P}_X \mathrm{D}_Z \mathcal{P} V &= -\mathcal{P}_X ZS\\ &= -ZS + X\frac{1}{2}(X^{\mathrm{T}}ZS + SZ^{\mathrm{T}}X)\\ &= -ZS + \frac{1}{2}X\Omega S - \frac{1}{2}XS\Omega\\ &= -ZX^{\mathrm{T}}V - \frac{1}{2}XZ^{\mathrm{T}}V - \frac{1}{2}VX^{\mathrm{T}}Z\\ &= -ZX^{\mathrm{T}}V - X\frac{1}{2}(Z^{\mathrm{T}}V + V^{\mathrm{T}}Z).\end{aligned}$$

In summary, for all $Z \in \mathrm{T}_X\mathcal{M}$ and $V \in \mathrm{T}_X^{\perp}\mathcal{M}$, we have

$$\mathfrak{A}_X(Z,V) = -ZX^{\mathrm{T}}V - X\frac{1}{2}(Z^{\mathrm{T}}V + V^{\mathrm{T}}Z).$$

An equivalent formula can be found in [11, §4.1].

4.2 The Sphere

The unit sphere S^{n-1} is the Stiefel manifold $\mathrm{St}(p,n)$ with $p = 1$. The orthogonal projector $\mathcal{P}_x$ onto the tangent space reduces to

$$\mathcal{P}_x u = (I - xx^{\mathrm{T}})u = u - xx^{\mathrm{T}}u,$$

and the Weingarten map reduces to

$$\mathfrak{A}_x(z,v) = -zx^{\mathrm{T}}v.$$

4.3 The Orthogonal Group

The orthogonal group $\mathrm{O}(n)$ is the Stiefel manifold $\mathrm{St}(p,n)$ with $p = n$. The orthogonal projector $\mathcal{P}_X$ onto the tangent space reduces to

$$\mathcal{P}_X U = X\frac{1}{2}(X^{\mathrm{T}}U - U^{\mathrm{T}}X),$$

and the Weingarten map reduces to

$$\mathfrak{A}_X(Z,V) = -X\frac{1}{2}(V^{\mathrm{T}}Z - Z^{\mathrm{T}}V).$$

4.4 The Grassmann Manifold

Let $\mathrm{Gr}_{m,n}$ denote the Grassmann manifold of m-dimensional subspaces of $\mathbb{R}^n$, viewed as the set of rank-m orthogonal projectors in R^n, i.e.,

$$\mathrm{Gr}_{m,n} = \{X \in \mathbb{R}^{n\times n} : X^{\mathrm{T}} = X, X^2 = X, \mathrm{tr}X = n\}.$$

Then, from [10, Prop. 2.1], we have that $\mathcal{P}_X = \mathrm{ad}_X^2$ with $\mathrm{ad}_X A := [X,A] := XA - AX$ and $\mathrm{ad}_X^2 := \mathrm{ad}_X \circ \mathrm{ad}_X$. It follows that, for all $Z \in \mathrm{T}_X\mathrm{Gr}_{m,n}$ and all $V \in \mathrm{T}_X^{\perp}\mathrm{Gr}_{m,n}$, it holds that

$$\begin{aligned}\mathcal{P}_X\,\mathrm{D}_Z\mathcal{P}\,V &= \mathrm{ad}_X^2\,(\mathrm{ad}_Z\mathrm{ad}_X V + \mathrm{ad}_X\mathrm{ad}_Z V)\\ &= \mathrm{ad}_X^2\mathrm{ad}_Z\mathrm{ad}_X V + \mathrm{ad}_X\mathrm{ad}_Z V\\ &= \mathrm{ad}_X\mathrm{ad}_Z V\\ &= -\mathrm{ad}_X\mathrm{ad}_V Z,\end{aligned}$$

where $\mathrm{ad}_A B := [A,B] := AB - BA$. One recovers from (10) the Hessian formula of [10, (2.109)].

4.5 The Fixed-Rank Manifol

Let $\mathcal{M}_p(m\times n)$ denote the set of all $m\times n$ matrices of rank p. This is a submanifold of $\mathbb{R}^{m\times n}$ of dimension $(m+n-p)p$; see [14, Example 8.14]. Let $X \in \mathcal{M}_p(m\times n)$ and, without loss of generality, let $X = U\Sigma V^T$ with $U \in \mathrm{St}(p,m)$ and $V \in \mathrm{St}(p,n)$. The projector $\mathcal{P}_X$ onto $\mathrm{T}_X\mathcal{M}_p(m\times n)$ is given by [20, §2.1]

$$\mathcal{P}_X W = \mathrm{P}_U W \mathrm{P}_V + \mathrm{P}_U^{\perp} W \mathrm{P}_V + \mathrm{P}_U W \mathrm{P}_V^{\perp} = W\mathrm{P}_V + \mathrm{P}_U W - \mathrm{P}_U W \mathrm{P}_V,$$

where $\mathrm{P}_U := UU^{\mathrm{T}}$ and $\mathrm{P}_U^{\perp} := I - \mathrm{P}_U$.

We now turn to the Weingarten map. Let $Z \in \mathrm{T}_X\mathcal{M}_p(m\times n)$. Let $\dot U \in \mathrm{T}_U\mathrm{St}(p,m)$, $\dot\Sigma$ diagonal, and $\dot V \in \mathrm{T}_V\mathrm{St}(p,n)$ be such that $Z = \mathrm{D}_{\dot U,\dot\Sigma,\dot V}(U\Sigma V^{\mathrm{T}}) = \dot U\Sigma V^{\mathrm{T}} + U\dot\Sigma V^{\mathrm{T}} + U\Sigma\dot V^{\mathrm{T}}$. We also let $\dot{\mathrm{P}}_U = \mathrm{D}_{\dot U}\mathrm{P}_U = \dot U U^{\mathrm{T}} + U\dot U^{\mathrm{T}}$, and likewise with $\dot{\mathrm{P}}_V$. Let $W \in \mathrm{T}_X^{\perp}\mathcal{M}_p(m\times n)$. We have

$$\begin{aligned}\mathcal{P}_X\,\mathrm{D}_Z\mathcal{P}\,W &= \mathcal{P}_X\left(W\dot{\mathrm{P}}_V + \dot{\mathrm{P}}_U W - \dot{\mathrm{P}}_U W \mathrm{P}_V - \mathrm{P}_U W \dot{\mathrm{P}}_V\right)\\ &= \mathcal{P}_X\left(\mathrm{P}_U^{\perp} W \dot{\mathrm{P}}_V + \dot{P}_U W \mathrm{P}_V^{\perp}\right)\\ &= \mathrm{P}_U^{\perp} W \dot{\mathrm{P}}_V \mathrm{P}_V + \mathrm{P}_U \dot{\mathrm{P}}_U W \mathrm{P}_V^{\perp}.\end{aligned}$$

Since $W \in \mathrm{T}_X^{\perp}\mathcal{M}_p(m \times n)$, we have $W = U_\perp L_W V_\perp^{\mathrm{T}}$ with L_W arbitary; this follows from the expression of $\mathrm{T}_X\mathcal{M}_p(m \times n)$ in [20, §2.1]. Hence $U^T W = 0$, $\mathrm{P}_U^{\perp} W = W$, $WV = 0$, $W\mathrm{P}_V^{\perp} = W$. Using these equations, one obtains

$$\mathrm{P}_U^{\perp} W \dot{\mathrm{P}}_V \mathrm{P}_V = W(\dot{V}V^{\mathrm{T}} + V\dot{V}^{\mathrm{T}})\mathrm{P}_V = W\dot{V}^{\mathrm{T}}V^{\mathrm{T}}\mathrm{P}_V = W\dot{V}^{\mathrm{T}}V.$$

Likewise, we obtain

$$\mathrm{P}_U \dot{\mathrm{P}}_U W \mathrm{P}_V^{\perp} = U\dot{U}^{\mathrm{T}}W.$$

In summary, we have

$$\mathcal{P}_X\, \mathrm{D}_Z \mathcal{P}\, W = W\dot{V}^{\mathrm{T}}V + U\dot{U}^{\mathrm{T}}W.$$

We now seek an alternate expression where only X, Z, and W appear. To this end, observe that the pseudo-inverse of X is given by $X^+ = V\Sigma^{-1}U^{\mathrm{T}}$. Then, recalling that $WV = 0$, we find that

$$\begin{aligned} WZ^{\mathrm{T}}(X^+)^{\mathrm{T}} &= W(\dot{V}\Sigma U^{\mathrm{T}} + V(\dot{\Sigma}U^{\mathrm{T}} + \Sigma\dot{U}^{\mathrm{T}}))U\Sigma^{-1}V^{\mathrm{T}} \\ &= W\dot{V}\Sigma U^{\mathrm{T}}U\Sigma^{-1}V^{\mathrm{T}} \\ &= W\dot{V}V^{\mathrm{T}}. \end{aligned}$$

Similarly, we obtain that

$$(X^+)^{\mathrm{T}}Z^{\mathrm{T}}W = U\dot{U}^{\mathrm{T}}W.$$

In conclusion, we have

$$\mathcal{P}_X\, \mathrm{D}_Z \mathcal{P}\, W = WZ^{\mathrm{T}}(X^+)^{\mathrm{T}} + (X^+)^{\mathrm{T}}Z^{\mathrm{T}}W.$$

It is interesting to note that this expession, combined with (10), provides an expression that allows to recover the Hessian formula found in [20, §2.3].

Acknowledgment. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

This work was financially supported by the Belgian FRFC (Fonds de la Recherche Fondamentale Collective).

References

1. Absil, P.A., Boumal, N.: Low-rank matrix completion via trust-regions on the Grassmann manifold. Tech. Rep. UCL-INMA-2012.07, UCLouvain (2012)
2. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008), `http://sites.uclouvain.be/absil/amsbook/`

3. Absil, P.A., Malick, J.: Projection-like retractions on matrix manifolds. SIAM J. Optim. 22(1), 135–158 (2012)
4. Absil, P.A., Trumpf, J., Mahony, R., Andrews, B.: All roads lead to Newton: Feasible second-order methods for equality-constrained optimization. Tech. Rep. UCL-INMA-2009.024, UCLouvain (2009)
5. Adler, R.L., Dedieu, J.P., Margulies, J.Y., Martens, M., Shub, M.: Newton's method on Riemannian manifolds and a geometric model for the human spine. IMA J. Numer. Anal. 22(3), 359–390 (2002)
6. Chavel, I.: Riemannian geometry, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 98. Cambridge University Press, Cambridge (2006); a modern introduction
7. 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), 210–234 (2012)
8. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. 20(2), 303–353 (1998)
9. Gallot, S., Hulin, D., Lafontaine, J.: Riemannian geometry, 3rd edn. Universitext. Springer, Berlin (2004)
10. Helmke, U., Hüper, K., Trumpf, J.: Newton's method on Grassmann manifolds, arXiv:0709.2205v2 (September 2007)
11. Hendriks, H., Landsman, Z.: Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. J. Multivariate Anal. 67(2), 227–243 (1998)
12. Jeuris, B., Vandebril, R., Vandereycken, B.: A survey and comparison of contemporary algorithms for computing the matrix geometric mean. Electron. Trans. Numer. Anal. 39, 379–402 (2012)
13. Jiang, B., Dai, Y.H.: A framework of constraint preserving update schemes for optimization on stiefel manifold, arXiv:1301.0172 (2013)
14. Lee, J.M.: Introduction to smooth manifolds. Graduate Texts in Mathematics, vol. 218. Springer, New York (2003)
15. Nishimori, Y., Akaho, S.: Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. Neurocomputing 67, 106–135 (2005)
16. Ring, W., Wirth, B.: Optimization methods on Riemannian manifolds and their application to shape space. SIAM J. Optim. 22(2), 596–627 (2012)
17. Schulz, V.: A Riemannian view on shape optimization, arXiv:1203.1493 (2012)
18. Shalit, U., Weinshall, D., Chechik, G.: Online learning in the embedded manifold of low-rank matrices. J. of Mach. Learn. Res. 13, 429–458 (2013)
19. Shub, M.: Some remarks on dynamical systems and numerical analysis. In: Lara-Carrero, L., Lewowicz, J. (eds.) Proc. VII ELAM, Equinoccio, U. Simón Bolívar, Caracas, pp. 69–92 (1986)
20. Vandereycken, B.: Low-rank matrix completion by Riemannian optimization (2011) (submitted)
21. Wen, Z., Yin, W.: A feasible method for optimization with orthogonality constraints. Math. Program (2012), doi:10.1007/s10107-012-0584-1