Two Newton methods on the manifold of fixed-rank matrices endowed with Riemannian quotient geometries [*]

P.-A. Absil[†]   Luca Amodei[‡]   Gilles Meyer[§]

August 9, 2013

Abstract

We consider two Riemannian geometries for the manifold $\mathcal{M}(p, m \times n)$ of all $m \times n$ matrices of rank p. The geometries are induced on $\mathcal{M}(p, m \times n)$ by viewing it as the base manifold of the submersion $\pi : (M, N) \mapsto MN^{\mathrm{T}}$, selecting an adequate Riemannian metric on the total space, and turning π into a Riemannian submersion. The theory of Riemannian submersions, an important tool in Riemannian geometry, makes it possible to obtain expressions for fundamental geometric objects on $\mathcal{M}(p, m \times n)$ and to formulate the Riemannian Newton methods on $\mathcal{M}(p, m \times n)$ induced by these two geometries. The Riemannian Newton methods admit a stronger and more streamlined convergence analysis than the Euclidean counterpart, and the computational overhead due to the Riemannian geometric machinery is shown to be mild. Potential applications include low-rank matrix completion and other low-rank matrix approximation problems.

Key words.   fixed-rank matrices; manifold; differential geometry; Riemannian geometry; Riemannian submersion; Levi-Civita connection; Riemannian connection; Riemannian exponential map; geodesics; Newton's method

1   Introduction

Let m, n, and $p \leq \min\{m, n\}$ be positive integers and let $\mathcal{M}(p, m \times n)$ denote the set of all rank-p matrices of size $m \times n$,

$$\mathcal{M}(p, m \times n) = \{X \in \mathbb{R}^{m \times n} : \mathrm{rank}(X) = p\}. \tag{1}$$

Given a smooth real-valued function f defined on $\mathcal{M}(p, m \times n)$, we consider the problem

$$\min f(X) \quad \text{subject to } X \in \mathcal{M}(p, m \times n). \tag{2}$$

Problem (2) subsumes low-rank matrix approximation problems, where $f(X) \equiv \|C - X\|^2$ with $C \in \mathbb{R}^{m \times n}$ given and $\| \cdot \|$ a (semi)norm. In particular, it includes low-rank matrix completion problems, which have been the topic of much attention recently; see [KMO10, DMK11, BA11, Van13, MMBS11, DKM12] and references therein, and also Section 7. We also mention the recent e-prints [MMBS12, MAS12] that appeared after the present paper was submitted. Interestingly, low-rank matrix completion problems combine two sparsity aspects: only a few elements of C are available, and the vector of singular values of X is restricted to have only a few nonzero elements.

This paper belongs to a trend of research, see [HM94, HS95, SE10, Van13, MMS11, MMBS11], where problem (2) is tackled using differential-geometric techniques exploiting the fact that $\mathcal{M}(p, m \times n)$ is a submanifold of $\mathbb{R}^{m \times n}$. We are interested in Riemannian Newton methods (see [Smi94, ADM$^+$02, AMS08]) for problem (2), with a preference for the pure Riemannian setting [Smi94]. This setting involves defining a Riemannian metric on $\mathcal{M}(p, m \times n)$ and providing an expression for the Riemannian connection—which underlies the Riemannian Hessian—and for the Riemannian exponential. When $\mathcal{M}(p, m \times n)$ is viewed as a Riemannian submanifold of $\mathbb{R}^{m \times n}$, the necessary ingredients for computing the Riemannian Hessian are available [Van13, §2.3], but a closed-form expression of the Riemannian exponential has been elusive in that geometry.

In this paper, we follow a different approach that strongly relies on two-term factorizations of low-rank matrices. To this end, let

$$\mathbb{R}_*^{m \times p} = \{X \in \mathbb{R}^{m \times p} : \operatorname{rank}(X) = p\} \tag{3}$$

denote the set of all full-rank $m \times p$ matrices, and observe that, since the function

$$\pi : \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p} \to \mathcal{M}(p, m \times n) : (M, N) \mapsto MN^{\mathrm{T}} \tag{4}$$

is surjective, problem (2) amounts to the optimization over its domain of the function $\bar{f} = f \circ \pi$, i.e.,

$$\bar{f} : \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p} \to \mathbb{R} : (M, N) \mapsto f(MN^{\mathrm{T}}). \tag{5}$$

Pleasantly, whereas $\mathcal{M}(p, m \times n)$ is a nonlinear space, $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ is an open subset of a linear space; more precisely, $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ is the linear space $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ with a zero-measure nowhere-dense set excerpted. (Indeed, the excerpted set is an algebraic variety of codimension greater than 1.) The downside is that the minimizers of $\bar{f}$ are never isolated; indeed, for all $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$, $\bar{f} = f \circ \pi$ assumes the same value $\bar{f}(M, N)$ at all points of

$$\pi^{-1}(MN^{\mathrm{T}}) = \{(MR, NR^{-\mathrm{T}}) : R \in \mathrm{GL}(p)\}, \tag{6}$$

where

$$\mathrm{GL}(p) = \{R \in \mathbb{R}^{p \times p} : \det(R) \neq 0\}$$

denotes the general linear group of degree p. In the context of Newton-type methods, this can be a source of concern since, whereas the convergence theory of Newton's method to nondegenerate minimizers is well understood (see, e.g., [DS83, Theorem 5.2.1]), the situation becomes more intricate in the presence of non-isolated minimizers (see, e.g., [GR85]).

The proposed remedy to this downside consists in elaborating a Riemannian Newton method that evolves conceptually on $\mathcal{M}(p, m \times n)$—avoiding the structural degeneracy in $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$—while still being formulated in $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$. This is made possible by endowing $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ and $\mathcal{M}(p, m \times n)$ with Riemannian metrics that turn π into a Riemannian submersion. The theory of Riemannian submersions [O'N66, O'N83] then provides a way of representing the Riemannian connection and the Riemannian exponential of $\mathcal{M}(p, m \times n)$ in terms of the same objects of $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$.

It should be pointed out that the local quadratic convergence of the Riemannian Newton method is retained if the Riemannian connection is replaced by any affine connection and the Riemannian exponential is replaced by any first-order approximation, termed retraction; see [AMS08, §6.3]. The preference for the pure Riemannian setting is thus mainly motivated by the mathematical elegance of a method fully determined by the sole Riemannian metric.

Some of the material of this paper is inspired from the PhD thesis [Mey11] and the talk [ADY09].

The paper is organized as follows. In the short sections 2 and 3, we show that π is a submersion and we recall some fundamentals of Riemannian submersions. A first, natural but unsuccessful attempt at turning π into a Riemannian submersion is presented in Section 4. Two ways of achieving success are then presented in sections 5 and 6. In Section 5, the strategy consists of introducing a non-Euclidean Riemannian metric on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$, whereas in Section 6, the plan of action is to restrict $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ by imposing orthonormality of one of the factors. We obtain closed-form expressions for the Riemannian connection (in both cases) and for the Riemannian exponential (in the latter case). Numerical experiments are conducted in Section 7, and conclusions are drawn in Section 8.

2 $\mathcal{M}(p, m \times n)$ as a quotient manifold

The set $\mathcal{M}(p, m \times n)$ of rank-p matrices of size $m \times n$ is known to be an embedded submanifold of dimension $p(m + n - p)$ of $\mathbb{R}^{m \times n}$, connected whenever $\max\{m, n\} > 1$; see [HM94, Ch. 5, Prop. 1.14]. Hence π (4) is a smooth surjective map between two manifolds.

We show that π is a submersion, i.e., that the differential of π is everywhere surjective. Observe that the tangent space to $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ at (M, N) is given by

$$\mathrm{T}_{(M,N)} \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p} = \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p};$$

this comes from the fact that $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ is an open submanifold of the Euclidean space $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ [AMS08, §3.5.1]. For all $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ and all $(\dot{M}, \dot{N}) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$, we have $\mathrm{D}\pi(M, N)[(\dot{M}, \dot{N})] = \dot{M}N^{\mathrm{T}} + M\dot{N}^{\mathrm{T}}$. Working in a coordinate system where $M = \begin{bmatrix} I & 0 \end{bmatrix}^{\mathrm{T}}$ and $N = \begin{bmatrix} I & 0 \end{bmatrix}^{\mathrm{T}}$, one readily sees that the dimension of the range of the map $(\dot{M}, \dot{N}) \mapsto \mathrm{D}\pi(M, N)[(\dot{M}, \dot{N})]$ is equal to $p(m + n - p)$, the dimension of the codomain of π. Hence π is a submersion.

As a consequence, by the submersion theorem [AMS08, Proposition 3.3.3], the fibers $\pi^{-1}(MN^{\mathrm{T}})$ are p^2-dimensional submanifolds of $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$. Moreover, by [AMR88, Proposition 3.5.23], the equivalence relation $\sim$ on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$, defined by $(M_a, N_a) \sim (M_b, N_b)$ if and only if $\pi(M_a, N_a) = \pi(M_b, N_b)$, is regular and $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}/\sim$ is a quotient manifold diffeomorphic to $\mathcal{M}(p, m \times n)$.

3   Riemannian submersion: principles

Turning π into a Riemannian submersion amounts to endowing its domain $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ with a Riemannian metric $\bar{g}$ that satisfies a certain invariance condition, described next.

By definition, the vertical space $\mathcal{V}_{(M,N)}$ at a point $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ is the tangent space to the fiber $\pi^{-1}(MN^{\mathrm{T}})$ (6). We obtain

$$\mathcal{V}_{(M,N)} = \{(M\dot{R}, -N\dot{R}^{\mathrm{T}}) : \dot{R} \in \mathbb{R}^{p \times p}\}. \tag{7}$$

Let $\bar{g}$ be a Riemannian metric on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$. Then one defines the *horizontal space* $\mathcal{H}_{(M,N)}$ at (M, N) to be the orthogonal complement of $\mathcal{V}_{(M,N)}$ in $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ relative to $\bar{g}_{(M,N)}$, i.e.,

$$\mathcal{H}_{(M,N)} = \{(\dot{M}, \dot{N}) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} : \bar{g}_{(M,N)}((\dot{M}, \dot{N}), (M\dot{R}, -N\dot{R}^{\mathrm{T}})) = 0, \forall \dot{R} \in \mathbb{R}^{p \times p}\}. \tag{8}$$

Next, given a tangent vector $\dot{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$, there is one and only one

$$\dot{X}_{\uparrow(M,N)} \in \mathcal{H}_{(M,N)} \quad \text{such that} \quad \mathrm{D}\pi(M,N)[\dot{X}_{\uparrow(M,N)}] = \dot{X}_{MN^{\mathrm{T}}}, \tag{9}$$

where $\mathrm{D}\pi(X)[\dot{X}]$ denotes the differential of π at X applied to $\dot{X}$. This $\dot{X}_{\uparrow(M,N)}$ is termed the *horizontal lift* of $\dot{X}_{MN^{\mathrm{T}}}$ at (M, N). If (and only if), for all $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$, all $\dot{X}_{MN^{\mathrm{T}}}, \check{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$, and all $R \in \mathrm{GL}(p)$, it holds that

$$\bar{g}_{(M,N)}(\dot{X}_{\uparrow(M,N)}, \check{X}_{\uparrow(M,N)}) = \bar{g}_{(MR,NR^{-\mathrm{T}})}(\dot{X}_{\uparrow(MR,NR^{-\mathrm{T}})}, \check{X}_{\uparrow(MR,NR^{-\mathrm{T}})}), \tag{10}$$

then there is a (unique) Riemannian metric g on $\mathcal{M}(p, m \times n)$ consistently defined by

$$g_{MN^{\mathrm{T}}}(\dot{X}_{MN^{\mathrm{T}}}, \check{X}_{MN^{\mathrm{T}}}) = \bar{g}_{(M,N)}(\dot{X}_{\uparrow(M,N)}, \check{X}_{\uparrow(M,N)}).$$

The submersion $\pi : (\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g}) \to (\mathcal{M}(p, m \times n), g)$ is then termed a *Riemannian submersion*, and $(\mathcal{M}(p, m \times n), g)$ is termed a *Riemannian quotient manifold* of $(\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g})$. (We will sometimes omit the Riemannian metrics in the notation when they are clear from the context or undefined.)

In summary, in order to turn π into a Riemannian submersion, we "just" have to choose a Riemannian metric $\bar{g}$ of $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ that satisfies the invariance condition (10).

4 $\mathcal{M}(p, m \times n)$ as a non-Riemannian quotient manifold

In this section, we consider on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ the Euclidean metric $\bar{g}$, defined by

$$\bar{g}_{(M,N)}\left((\dot{M}, \dot{N}), (\check{M}, \check{N})\right) := \mathrm{trace}(\dot{M}^{\mathrm{T}}\check{M}) + \mathrm{trace}(\dot{N}^{\mathrm{T}}\check{N}), \tag{11}$$

and we show that the invariance condition (10) does not hold. Hence $\pi : (\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g}) \to \mathcal{M}(p, m \times n)$ cannot be turned into a Riemannian submersion.

The horizontal space (8) is

$$\mathcal{H}_{(M,N)} = \{(\dot{M}, \dot{N}) : \mathrm{trace}(\dot{M}^{\mathrm{T}}M\dot{R}) + \mathrm{trace}(-\dot{N}^{\mathrm{T}}N\dot{R}^{\mathrm{T}}) = 0, \forall \dot{R} \in \mathbb{R}^{p \times p}\}.$$

Using the identities $\mathrm{trace}(A) = \mathrm{trace}(A^{\mathrm{T}})$ and $\mathrm{trace}(AB) = \mathrm{trace}(BA)$, we obtain the identity $\mathrm{trace}(\dot{M}^{\mathrm{T}}M\dot{R}) + \mathrm{trace}(-\dot{N}^{\mathrm{T}}N\dot{R}^{\mathrm{T}}) = \mathrm{trace}\left((\dot{R}^{\mathrm{T}}(M^{\mathrm{T}}\dot{M} - \dot{N}^{\mathrm{T}}N)\right)$. It follows that the following propositions are equivalent:

1. $(\dot{M}, \dot{N}) \in \mathcal{H}_{(M,N)}$,

2. $M^{\mathrm{T}}\dot{M} = \dot{N}^{\mathrm{T}}N$, $\dot{M} \in \mathbb{R}_*^{m \times p}$, $\dot{N} \in \mathbb{R}_*^{n \times p}$,

3. $\exists L_M \in \mathbb{R}^{(m-p) \times p}$, $L_N \in \mathbb{R}^{(n-p) \times p}$, $S \in \mathbb{R}^{p \times p} : \begin{cases} \dot{M} = M_\perp L_M + M(M^{\mathrm{T}}M)^{-1}S \\ \dot{N} = N_\perp L_N + N(N^{\mathrm{T}}N)^{-1}S^{\mathrm{T}}, \end{cases}$

where $M_\perp$ denotes an orthonormal $m \times (m - p)$ matrix such that $M^{\mathrm{T}}M_\perp = 0$, and likewise for $N_\perp$.

Let $X = MN^{\mathrm{T}}$ and let $\dot{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$. We seek an expression for the *horizontal lift* $\dot{X}_{\uparrow(M,N)} = (\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})$ of $\dot{X}_{MN^{\mathrm{T}}}$ at (M, N), defined by (9). By a reasoning similar to the one detailed in Section 5.3 below, we obtain

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = (\dot{X}_{MN^{\mathrm{T}}}N - MK)(N^{\mathrm{T}}N)^{-1} \quad \text{and} \quad \dot{X}_{\uparrow\mathrm{N}(M,N)} = (\dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}}M - NK^{\mathrm{T}})(M^{\mathrm{T}}M)^{-1},$$

where K solves the Sylvester equation

$$M^{\mathrm{T}}MK + KN^{\mathrm{T}}N = M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N.$$

One sees by inspection, or by a numerical check, that the invariance condition (10) does not hold, and this concludes the argument.

We will now consider two approaches to obtain the invariance condition (10). The first one, followed in Section 5, is to modify the Riemannian metric $\bar{g}$ without altering the total space $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$. The other one, investigated in Section 6, consists of restricting the total space without modifying the expression (11) of the Riemannian metric $\bar{g}$.

5 $\mathcal{M}(p, m \times n)$ as a Riemannian quotient manifold of $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$

In this section, we proceed as in Section 4, but now with a different Riemannian metric $\bar{g}$, defined in (12) below. As we will see, the rationale laid out in Section 4 now leads to the conclusion that $\pi : (\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g}) \to \mathcal{M}(p, m \times n)$, with $\bar{g}$ given by (12) instead of (11), *can* be turned into a Riemannian submersion. This endows $\mathcal{M}(p, m \times n)$ with a Riemannian metric, g. We then work out formulas for the Riemannian gradient and Hessian of f on the Riemannian manifold $(\mathcal{M}(p, m \times n), g)$, and we state the corresponding Newton method.

5.1 Riemannian metric in total space

Inspired from the case of the Grassmann manifold viewed as a Riemannian quotient manifold of $\mathbb{R}_*^{n \times p}$ [AMS08, Example 3.6.4], we consider the Riemannian metric $\bar{g}$ on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ defined by

$$\bar{g}_{(M,N)}\left((\dot{M}, \dot{N}), (\check{M}, \check{N})\right) := \mathrm{trace}\left((M^{\mathrm{T}}M)^{-1}\dot{M}^{\mathrm{T}}\check{M} + (N^{\mathrm{T}}N)^{-1}\dot{N}^{\mathrm{T}}\check{N}\right). \qquad (12)$$

We now proceed to show that it satisfies the invariance condition (10).

5.2 Horizontal space

The elements $(\dot{M}, \dot{N})$ of the horizontal space $\mathcal{H}_{(M,N)}$ (8) are readily found to be characterized by

$$M^{\mathrm{T}}\dot{M}(M^{\mathrm{T}}M)^{-1} = (N^{\mathrm{T}}N)^{-1}\dot{N}^{\mathrm{T}}N. \tag{13}$$

In other words,

$$\mathcal{H}_{(M,N)} = \{(\dot{M}, \dot{N}) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p} : N^{\mathrm{T}}N M^{\mathrm{T}}\dot{M} = \dot{N}^{\mathrm{T}}N M^{\mathrm{T}}M\}. \tag{14}$$

5.3 Horizontal lift

Let $X = MN^{\mathrm{T}}$ and let $\dot{X}_{MN^{\mathrm{T}}}$ belong to $\mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$. We seek an expression for the *horizontal lift* $\dot{X}_{\uparrow(M,N)} = (\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})$ defined in (9). In view of (13), we find that the horizontality condition $(\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)}) \in \mathcal{H}_{(M,N)}$ is equivalent to

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = M_{\perp}L_{\mathrm{M}} + M(M^{\mathrm{T}}M)^{-1}K(M^{\mathrm{T}}M) \tag{15a}$$

$$\dot{X}_{\uparrow\mathrm{N}(M,N)} = N_{\perp}L_{\mathrm{N}} + N(N^{\mathrm{T}}N)^{-1}K^{\mathrm{T}}(N^{\mathrm{T}}N), \tag{15b}$$

where $L_{\mathrm{M}} \in \mathbb{R}^{(m-p) \times p}$, $L_{\mathrm{N}} \in \mathbb{R}^{(n-p) \times p}$ and $K \in \mathbb{R}^{p \times p}$. Since $\mathrm{D}\pi(M, N)[\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)}] \equiv M\dot{X}_{\uparrow\mathrm{N}(M,N)}^{\mathrm{T}} + \dot{X}_{\uparrow\mathrm{M}(M,N)}N^{\mathrm{T}}$, the definition (9) implies that

$$\dot{X}_{MN^{\mathrm{T}}} = M\dot{X}_{\uparrow\mathrm{N}(M,N)}^{\mathrm{T}} + \dot{X}_{\uparrow\mathrm{M}(M,N)}N^{\mathrm{T}}. \tag{16}$$

Replacing (15) in (16) yields

$$\dot{X}_{MN^{\mathrm{T}}} = ML_{\mathrm{N}}^{\mathrm{T}}N_{\perp}^{\mathrm{T}} + M(N^{\mathrm{T}}N)K(N^{\mathrm{T}}N)^{-1}N^{\mathrm{T}} + M_{\perp}L_{\mathrm{M}}N^{\mathrm{T}} + M(M^{\mathrm{T}}M)^{-1}K(M^{\mathrm{T}}M)N^{\mathrm{T}}. \tag{17}$$

Multiplying (17) on the left by $(M^{\mathrm{T}}M)^{-1}M^{\mathrm{T}}$ and on the right by $N_{\perp}$ yields

$$L_{\mathrm{N}}^{\mathrm{T}} = (M^{\mathrm{T}}M)^{-1}M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N_{\perp}, \tag{18a}$$

multiplying (17) on the left by $M_{\perp}^{T}$ and on the right by $N(N^{\mathrm{T}}N)^{-1}$ yields

$$L_{\mathrm{M}} = M_{\perp}^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N(N^{\mathrm{T}}N)^{-1}, \tag{18b}$$

and multiplying (17) on the left by M^{T} and on the right by N yields

$$M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N = M^{\mathrm{T}}MN^{\mathrm{T}}NK + KM^{\mathrm{T}}MN^{\mathrm{T}}N. \tag{18c}$$

Replacing (18) into (15) yields

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = M_{\perp}M_{\perp}^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N(N^{\mathrm{T}}N)^{-1} + M(M^{\mathrm{T}}M)^{-1}KM^{\mathrm{T}}M \tag{19a}$$

$$\dot{X}_{\uparrow\mathrm{N}(M,N)} = N_{\perp}N_{\perp}^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}}M(M^{\mathrm{T}}M)^{-1} + N(N^{\mathrm{T}}N)^{-1}K^{\mathrm{T}}N^{\mathrm{T}}N. \tag{19b}$$

We can further exploit the identities $M_{\perp}M_{\perp}^{\mathrm{T}} = I - M(M^{\mathrm{T}}M)^{-1}M^{\mathrm{T}}$, and likewise for N, as well as (18c), to rewrite (19) as

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = (\dot{X}_{MN^{\mathrm{T}}}N - MN^{\mathrm{T}}NK)(N^{\mathrm{T}}N)^{-1} \tag{20a}$$

$$\dot{X}_{\uparrow\mathrm{N}(M,N)} = (\dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}}M - NM^{\mathrm{T}}MK^{\mathrm{T}})(M^{\mathrm{T}}M)^{-1}. \tag{20b}$$

This result is formalized as follows:

Proposition 5.1 *Consider the submersion π (4) and the horizontal distribution (14). Let $(M, N) \in \mathbb{R}_{*}^{m \times p} \times \mathbb{R}_{*}^{n \times p}$ and let $\dot{X}_{MN^{\mathrm{T}}}$ be in $\mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$. Then the horizontal lift of $\dot{X}_{MN^{\mathrm{T}}}$ at (M, N) is $\dot{X}_{\uparrow(M,N)} = (\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})$ given by (20), where K is the solution of the Sylvester equation (18c).*

5.4 Constitutive equation of horizontal lifts

A horizontal lift $\dot{X}_{\uparrow(M,N)}$ fully specifies $\dot{X}_{MN^{\mathrm{T}}} = \mathrm{D}\pi(M,N)[\dot{X}_{\uparrow(M,N)}] \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$ as well as its horizontal lift at any other point of the fiber $\pi^{-1}(MN^{\mathrm{T}})$ (6). Let us obtain an expression for $\dot{X}_{\uparrow(MR,NR^{-\mathrm{T}})}$ in terms of $\dot{X}_{\uparrow(M,N)}$. The expression (20) of horizontal lifts yields after routine manipulations

$$\dot{X}_{\uparrow\mathrm{M}(MR,NR^{-\mathrm{T}})} = \dot{X}_{\uparrow\mathrm{M}(M,N)}R, \quad \dot{X}_{\uparrow\mathrm{N}(MR,NR^{-\mathrm{T}})} = \dot{X}_{\uparrow\mathrm{N}(M,N)}R^{-\mathrm{T}}. \tag{21}$$

We have obtained:

Proposition 5.2 *Consider the submersion π (4) and the horizontal distribution (14). Then a vector field $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p} \ni (M, N) \mapsto \dot{X}_{\uparrow(M,N)} \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ is a horizontal lift if and only if (21) holds for all $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ and all $R \in \mathrm{GL}(p)$.*

5.5 Riemannian submersion

Routine manipulations using (21) yield that $\bar{g}$ (12) satisfies the invariance condition (10). (As an aside, we point out that this result can also be obtained from the fact that we have a principal fiber bundle $(\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \mathcal{M}(p, m \times n), \pi, \mathrm{GL}(p))$ and that the Riemannian metric $\bar{g}$ is invariant by the action of $\mathrm{GL}(p)$ on $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$.) Hence there is a (unique) Riemannian metric g on $\mathcal{M}(p, m \times n)$ that makes

$$\pi : (\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g}) \to (\mathcal{M}(p, m \times n), g) : (M, N) \mapsto MN^{\mathrm{T}} \tag{22}$$

a Riemannian submersion. The Riemannian metric g is consistently defined by

$$g_{MN^{\mathrm{T}}}(\dot{X}_{MN^{\mathrm{T}}}, \check{X}_{MN^{\mathrm{T}}}) := \bar{g}_{(M,N)}(\dot{X}_{\uparrow(M,N)}, \check{X}_{\uparrow(M,N)}). \tag{23}$$

5.6 Horizontal projection

We will need an expression for the projection $P_{(M,N)}^{\mathrm{h}}(\dot{M}, \dot{N})$ of $(\dot{M}, \dot{N}) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ onto the horizontal space (14) along the vertical space (7).

Since the projection is along the vertical space, we have

$$P_{(M,N)}^{\mathrm{h}}(\dot{M}, \dot{N}) = (\dot{M} + M\dot{R}, \dot{N} - N\dot{R}^{\mathrm{T}}) \tag{24}$$

for some $\dot{R} \in \mathbb{R}^{p \times p}$. It remains to obtain $\dot{R}$ by imposing horizontality of (24). Since horizontal vectors are characterized by (13), we find that (24) is horizontal if and only if

$$M^{\mathrm{T}}(\dot{M} + M\dot{R})(M^{\mathrm{T}}M)^{-1} = (N^{\mathrm{T}}N)^{-1}(\dot{N}^{\mathrm{T}} - \dot{R}N^{\mathrm{T}})N,$$

that is,

$$M^{\mathrm{T}}M\dot{R}(M^{\mathrm{T}}M)^{-1} + (N^{\mathrm{T}}N)^{-1}\dot{R}N^{\mathrm{T}}N = -M^{\mathrm{T}}\dot{M}(M^{\mathrm{T}}M)^{-1} + (N^{\mathrm{T}}N)^{-1}\dot{N}^{\mathrm{T}}N,$$

which can be rewritten as the Sylvester equation

$$N^{\mathrm{T}}NM^{\mathrm{T}}M\dot{R} + \dot{R}N^{\mathrm{T}}NM^{\mathrm{T}}M = -N^{\mathrm{T}}NM^{\mathrm{T}}\dot{M} + \dot{N}^{\mathrm{T}}NM^{\mathrm{T}}M. \tag{25}$$

In summary:

Proposition 5.3 *The projection $P_{(M,N)}^{\mathrm{h}}(\dot{M}, \dot{N})$ of $(\dot{M}, \dot{N}) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ onto the horizontal space (14) along the vertical space (7) is given by (24) where $\dot{R}$ is the solution of the Sylvester equation (25).*

5.7 Riemannian connection on the total space

Since the chosen Riemannian metric $\bar{g}$ (12) on the total space $\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}$ is not the Euclidean metric (11), it can be expected that the Riemannian connection on $(\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}, \bar{g})$ is not the plain differential. We show that this is indeed the case and we provide a formula for the Riemannian connection $\bar{\nabla}$ on $(\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}, \bar{g})$. The motivation for obtaining this formula is that the Riemannian Newton equation on $(\mathcal{M}(p, m\times n), g)$ requires the Riemannian connection on $(\mathcal{M}(p, m\times n), g)$, which is readily obtained from $\bar{\nabla}$ as we will see in Section 5.8. The general theory of Riemannian connections (also called Levi-Civita connections) can be found in [AMS08, §5.3] or in any Riemannian geometry textbook such as [dC92].

The development relies on Koszul's formula

$$2g(\nabla_\chi \eta, \xi) = \partial_\chi g(\eta, \xi) + \partial_\eta g(\chi, \xi) - \partial_\xi g(\chi, \eta) + g([\chi, \eta], \xi) - g([\chi, \xi], \eta) - g([\eta, \xi], \chi), \quad (26)$$

that characterizes the Riemannian connection ∇ on a manifold endowed with a Riemannian metric g, where χ, η, ζ are tangent vector fields, $\partial \chi$ denotes the derivative along χ, and $[\chi, \eta]$ denotes the Lie bracket of χ and η.

After lengthy but routine calculations, we obtain the following expression for the Riemannian connection $\bar{\nabla}$ on $(\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}, \bar{g})$, where $\mathrm{sym}(A) := \frac{1}{2}(A + A^{\mathrm{T}})$ denotes the symmetric part of A:

$$\left(\bar{\nabla}_{\dot{X}} \dot{Y}\right)_{\mathrm{M}} = \partial_{\dot{X}} \dot{Y}_{\mathrm{M}} - \dot{Y}_{\mathrm{M}}(M^{\mathrm{T}}M)^{-1}\mathrm{sym}(\dot{X}_{\mathrm{M}}^{\mathrm{T}}M) - \dot{X}_{\mathrm{M}}(M^{\mathrm{T}}M)^{-1}\mathrm{sym}(\dot{Y}_{\mathrm{M}}^{\mathrm{T}}M)$$
$$+ M(M^{\mathrm{T}}M)^{-1}\mathrm{sym}(\dot{X}_{\mathrm{M}}^{\mathrm{T}}\dot{Y}_{\mathrm{M}}) \quad (27\mathrm{a})$$

and

$$\left(\bar{\nabla}_{\dot{X}} \dot{Y}\right)_{\mathrm{N}} = \partial_{\dot{X}} \dot{Y}_{\mathrm{N}} - \dot{Y}_{\mathrm{N}}(N^{\mathrm{T}}N)^{-1}\mathrm{sym}(\dot{X}_{\mathrm{N}}^{\mathrm{T}}N) - \dot{X}_{\mathrm{N}}(N^{\mathrm{T}}N)^{-1}\mathrm{sym}(\dot{Y}_{\mathrm{N}}^{\mathrm{T}}N)$$
$$+ N(N^{\mathrm{T}}N)^{-1}\mathrm{sym}(\dot{X}_{\mathrm{N}}^{\mathrm{T}}\dot{Y}_{\mathrm{N}}), \quad (27\mathrm{b})$$

for all $(M, N) \in \mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}$, all $\dot{X} \in \mathrm{T}_{(M,N)}\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}$ and all tangent vector fields $\dot{Y}$ on $\mathbb{R}_*^{m\times p} \times \mathbb{R}_*^{n\times p}$.

5.8 Riemannian connection on the quotient space

Let ∇ denote the Riemannian connection on the quotient space $\mathcal{M}(p, m\times n)$ endowed with the Riemannian metric g (23). A classical result in the theory of Riemannian submersions (see [O'N66, Lemma 1] or [AMS08, §5.3.4]) states that

$$(\nabla_{\dot{X}_{MN^{\mathrm{T}}}} \dot{Y})_{\uparrow(M,N)} = P_{(M,N)}^{\mathrm{h}}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}} \dot{Y}_\uparrow),$$

for all $\dot{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m\times n)$ and all tangent vector fields $\dot{Y}$ on $\mathcal{M}(p, m\times n)$. That is, the horizontal lift of the Riemannian connection of the quotient space is given by the horizontal projection (24) of the Riemannian connection (27) of the total space.

5.9 Riemannian Newton equation

For a real-valued function f on a Riemannian manifold $\mathcal{M}$ with Riemannian metric g, we let $\mathrm{grad}\, f(x)$ denote the gradient of f at $x \in \mathcal{M}$—defined as the unique tangent vector to $\mathcal{M}$ at

x that satisfies $g_x(\text{grad } f(x), \xi_x) = Df(x)[\xi_x]$ for all $\xi_x \in T_x\mathcal{M}$—and the plain Riemannian Newton equation is given by

$$\nabla_{\eta_x}\text{grad } f = -\text{grad } f(x)$$

for the unknown $\eta_x \in T_x\mathcal{M}$, where ∇ stands for the Riemannian connection; see, e.g., [AMS08, §6.2].

We now turn to the manifold $\mathcal{M}(p, m \times n)$ endowed with the Riemannian metric g (23) and we obtain an expression of the Riemannian Newton equation by means of its horizontal lift through the Riemannian submersion π (22). First, on the total space $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ endowed with the Riemannian metric $\bar{g}$ (12), we readily obtain the following expression for the gradient of $\bar{f}$ (5):

$$\text{grad } \bar{f}(M, N) = (\partial_M \bar{f}(M, N)M^T M, \partial_N \bar{f}(M, N)N^T N),$$

where $\partial_M \bar{f}(M, N)$ denotes the Euclidean (i.e., classical) gradient of $\bar{f}$ with respect to its first argument, i.e., $(\partial_M \bar{f}(M, N))_{i,j} = \frac{d}{dt}\bar{f}(M + te_i e_j^T, N)|_{t=0}$, and likewise for $\partial_N \bar{f}(M, N)$ with the second argument. Then the horizontal lift of the Newton equation at a point (M, N) of the total space $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$, for the unknown $\dot{X}_{\uparrow(M,N)}$ in the horizontal space $\mathcal{H}_{(M,N)}$ (14), is

$$P^h_{(M,N)}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}}\text{grad } \bar{f}) = -\text{grad } \bar{f}(M, N), \tag{28}$$

where P^h is the horizontal projection given in Section 5.6 and $\bar{\nabla}$ is the Riemannian connection on $(\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}, \bar{g})$ given in Section 5.7. To obtain (28), we have used the fact (see [AMS08, (3.39)]) that $(\text{grad } f)_{\uparrow(M,N)} = \text{grad } \bar{f}(M, N)$, where the left-hand side denotes the horizontal lift of $\text{grad } f(MN^T)$ at (M, N).

Intimidating as it may be in view of the expressions of P^h and $\bar{\nabla}$, the Newton equation (28) is nevertheless merely a linear system of equations. Indeed, $\dot{X}_{\uparrow(M,N)} \mapsto P^h_{(M,N)}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}}\text{grad } \bar{f})$ is a linear transformation of the horizontal space $\mathcal{H}_{(M,N)}$. Thus (28) can be solved using "matrix-free" linear solvers such as GMRES. Moreover, in addition to computing the Euclidean gradient of $\bar{f}$ and the Euclidean derivative of the Euclidean gradient of f along $\dot{X}_{\uparrow(M,N)}$, computing $P^h_{(M,N)}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}}\text{grad } \bar{f})$ requires only $O(p^2(m + n + p))$ flops.

5.10 Newton's method

In order to spell out on $(\mathcal{M}(p, m \times n), g)$ the Riemannian Newton method as defined in [AMS08, §6.2], the last missing ingredient is a retraction R that turns the Newton vector $\dot{X}_{MN^T}$ into an updated iterate $R_{MN^T}\dot{X}_{MN^T}$ in $\mathcal{M}(p, m \times n)$. The general definition of a retraction can be found in [AMS08, §4.1].

The quintessential retraction on a Riemannian manifold is the Riemannian exponential; see [AMS08, §5.4]. However, computing the Riemannian exponential amounts to solving the differential equation $\nabla_{\dot{X}}\dot{X} = 0$, which may not admit a closed-form solution. In the case of $(\mathcal{M}(p, m \times n), g)$, we are not aware of such a closed-form solution, and this makes the exponential retraction impractical.

Fortunately, other retractions are readily available. A retraction on $\mathcal{M}(p, m \times n)$ is given by

$$R_{MN^T}(\dot{X}_{MN^T}) := (M + \dot{X}_{\uparrow M(M,N)})(N + \dot{X}_{\uparrow N(M,N)})^T, \tag{29}$$

where $\dot{X}_{\uparrow M(M,N)}$ and $\dot{X}_{\uparrow N(M,N)}$ are horizontal lifts as defined in Proposition 5.1. It is readily checked that the definition is consistent, i.e., it depends on MN^T and not on the specific choices of (M, N) in the fiber (6).

With all these elements in place, we can describe Newton's method as follows.

Theorem 5.4 (Riemannian Newton on $\mathcal{M}(p, m \times n)$ with Riemannian metric (23))
Let f be a real-valued function on the Riemannian manifold $\mathcal{M}(p, m \times n)$ (1), endowed with the Riemannian metric g (23), with the associated Riemannian connection ∇, and with the retraction (29). Then the Riemannian Newton method [AMS08, Algorithm 5] for f maps $MN^{\mathrm{T}} \in \mathcal{M}(p, m \times n)$ to $(M + \dot{X}_{\mathrm{M}})(N + \dot{X}_{\mathrm{N}})^{\mathrm{T}}$, where $(\dot{X}_{\mathrm{M}}, \dot{X}_{\mathrm{N}})$ is the solution $\dot{X}_{\uparrow(M,N)}$ of the Newton equation (28).

Note that, in practice, it is not necessary to form MN^{T}. Given an initial point $M_0 N_0^{\mathrm{T}}$, one can instead generate a sequence $\{(M_k, N_k)\}$ in $\mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ by applying the iteration map $(M, N) \mapsto (M + \dot{X}_{\mathrm{M}}, N + \dot{X}_{\mathrm{N}})$. The Newton sequence on $\mathcal{M}(p, m \times n)$ is then $\{M_k N_k^{\mathrm{T}}\}$, and it depends on $M_0 N_0^{\mathrm{T}}$ but not on the particular M_0 and N_0.

The following convergence result follows directly from the general convergence analysis of the Riemannian Newton method [AMS08, Theorem 6.3.2]. A critical point of $f :$ $\mathcal{M}(p, m \times n) \to \mathbb{R}$ is a point X_* where grad $f(X_*) = 0$. It is termed *nondegenerate* if the Hessian $\mathrm{T}_{X_*}\mathcal{M}(p, m \times n) \ni \dot{X} \mapsto \nabla_{\dot{X}}\mathrm{grad}\, f \in \mathrm{T}_{X_*}\mathcal{M}(p, m \times n)$ is invertible. These definitions do not depend on the Riemannian metric nor on the affine connection ∇.

Theorem 5.5 (quadratic convergence) *Let X_* be a nondegenerate critical point of f. Then there exists a neighborhood $\mathcal{U}$ of X_* in $\mathcal{M}(p, m \times n)$ such that, for all initial iterates $X_0 \in \mathcal{U}$, the iteration described in Theorem 5.4 generates an infinite sequence $\{X_k\}$ converging superlinearly (at least quadratically) to X_*.*

6 $\mathcal{M}(p, m \times n)$ as a Riemannian quotient manifold with an orthonormal factor

We now follow the second plan of action mentioned at the end of Section 1. Bear in mind that the meaning of much of the notation introduced above will be superseded by new definitions below.

6.1 A smaller total space

Let
$$\mathrm{St}(p, m) = \{M \in \mathbb{R}^{m \times p} : M^{\mathrm{T}}M = I_p\}, \tag{30}$$
denote the *Stiefel manifold* of orthonormal $m \times p$ matrices. For all $X \in \mathcal{M}(p, m \times n)$, there exists (M, N) with M orthonormal such that $X = MN^{\mathrm{T}}$. To see this, take $(M, N) \in \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p}$ such that $X = MN^{\mathrm{T}}$, let $M = QR$ be a QR decomposition of M, where R is invertible since M has full rank, and observe that $X = MR^{-1}(NR^{\mathrm{T}})^{\mathrm{T}} = Q(NR^{\mathrm{T}})^{\mathrm{T}}$. Hence
$$\pi : \mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p} \to \mathcal{M}(p, m \times n) : (M, N) \mapsto MN^{\mathrm{T}} \tag{31}$$
is a smooth surjective map between two manifolds.

As in Section 2, but now with the restricted total space $\mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p}$, we show that π (31) is a submersion. The tangent space at M to $\mathrm{St}(p, m)$ is given by (see [AMS08, Example 3.5.2])
$$\mathrm{T}_M \mathrm{St}(p, m) = \{\dot{M} \in \mathbb{R}^{m \times p} : M^{\mathrm{T}}\dot{M} + \dot{M}^{\mathrm{T}}M = 0\}$$
$$= \{M\Omega + M_\perp W : \Omega = -\Omega^{\mathrm{T}} \in \mathbb{R}^{p \times p}, W \in \mathbb{R}^{(m-p) \times p}\},$$

and we have
$$\mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}\right)=\left(\mathrm{T}_M\mathrm{St}(p,m)\right)\times\mathbb{R}^{n\times p}.$$

For all $(M,N)\in\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}$ and all $(\dot{M},\dot{N})\in\mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}\right)$, we have $\mathrm{D}\pi(M,N)[(\dot{M},\dot{N})]=\dot{M}N^{\mathrm{T}}+M\dot{N}^{\mathrm{T}}$. Here again, we can work in a coordinate system where $M=\begin{bmatrix}I & 0\end{bmatrix}^{\mathrm{T}}$ and $N=\begin{bmatrix}I & 0\end{bmatrix}^{\mathrm{T}}$. We have that $\{\mathrm{D}\pi(M,N)[(\dot{M},\dot{N})]:(\dot{M},\dot{N})\in\mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}\right)\}=\{\begin{bmatrix}\Omega+N_1^{\mathrm{T}} & N_2^{\mathrm{T}}\\ W & 0\end{bmatrix}:\Omega=-\Omega^{\mathrm{T}}\in\mathbb{R}^{p\times p},N_1\in\mathbb{R}^{p\times p},N_2\in\mathbb{R}^{(n-p)\times p},W\in\mathbb{R}^{(m-p)\times p}\}$, a linear subspace of dimension $p^2+(n-p)p+(m-p)p=p(m+n-p)$, which is the dimension of $\mathcal{M}(p,m\times n)$. Hence π (31) is a submersion.

The fiber of π (31) at MN^{T} is now
$$\pi^{-1}(MN^{\mathrm{T}})=\{(MR,NR):R\in\mathrm{O}(p)\},\tag{32}$$

where
$$\mathrm{O}(p)=\{R\in\mathbb{R}^{p\times p}:R^{\mathrm{T}}R=I_p\}$$

denotes the orthogonal group of degree p.

The vertical space $\mathcal{V}_{(M,N)}$ at a point $(M,N)\in\mathbb{R}_*^{m\times p}\times\mathbb{R}_*^{n\times p}$, i.e., the tangent space to the fiber $\pi^{-1}(MN^{\mathrm{T}})$ at (M,N), is given by
$$\mathcal{V}_{(M,N)}=\{(M\Omega,N\Omega):\Omega=-\Omega^{\mathrm{T}}\in\mathbb{R}^{p\times p}\}.\tag{33}$$

6.2   Riemannian metric in total space

We consider $\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}$ as a Riemannian submanifold of the Euclidean space $\mathbb{R}^{m\times p}\times\mathbb{R}^{n\times p}$. This endows $\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}$ with the Riemannian metric $\bar{g}$ defined by
$$\bar{g}_{(M,N)}\left((\dot{M},\dot{N}),(\check{M},\check{N})\right):=\mathrm{trace}\left(\dot{M}^{\mathrm{T}}\check{M}+\dot{N}^{\mathrm{T}}\check{N}\right)\tag{34}$$

for all $(\dot{M},\dot{N})$ and $(\check{M},\check{N})$ in $\mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}\right)$.

Adapting the rationale of Section 5, we will obtain in Section 6.6 below that, with this $\bar{g}$, π (31) can be turned into a Riemannian submersion.

6.3   Horizontal space

The horizontal space $\mathcal{H}_{(M,N)}$ is the orthogonal complement to $\mathcal{V}_{(M,N)}$ (33) in $\mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m)\times\mathbb{R}_*^{n\times p}\right)$ with respect to $\bar{g}$ (34). The following propositions are equivalent:

1. $(\dot{M},\dot{N})\in\mathcal{H}_{(M,N)}$,

2. $\dot{M}\in\mathrm{T}_M\mathrm{St}(p,n)$, $\dot{N}\in\mathbb{R}^{n\times p}$, $\mathrm{tr}(\tilde{\Omega}^{\mathrm{T}}(M^{\mathrm{T}}\dot{M}+N^{\mathrm{T}}\dot{N}))=0,\forall\tilde{\Omega}=-\tilde{\Omega}^{\mathrm{T}}$,

3.
$$M^{\mathrm{T}}\dot{M}=-(M^{\mathrm{T}}\dot{M})^{\mathrm{T}},\ M^{\mathrm{T}}\dot{M}+N^{\mathrm{T}}\dot{N}=(M^{\mathrm{T}}\dot{M}+N^{\mathrm{T}}\dot{N})^{\mathrm{T}},\tag{35}$$

4. $\dot{M}=M\Omega+M_\perp W$, $\dot{N}=N(N^{\mathrm{T}}N)^{-1}(-\Omega+S)+N_\perp L$, with $W\in\mathbb{R}^{(m-p)\times p}$, $\Omega=-\Omega^{\mathrm{T}}\in\mathbb{R}^{p\times p}$, $S=S^{\mathrm{T}}\in\mathbb{R}^{p\times p}$, $L\in\mathbb{R}^{(n-p)\times p}$.

In summary,
$$\mathcal{H}_{(M,N)}=\{(\dot{M},\dot{N}):M^{\mathrm{T}}\dot{M}=-(M^{\mathrm{T}}\dot{M})^{\mathrm{T}},\ M^{\mathrm{T}}\dot{M}+N^{\mathrm{T}}\dot{N}=(M^{\mathrm{T}}\dot{M}+N^{\mathrm{T}}\dot{N})^{\mathrm{T}}\}.\tag{36}$$

6.4 Horizontal lift

Proceeding as in Section 5.3 but now with the horizontal space (36) and taking into account that $M^{\mathrm{T}}M = I$, we obtain that the horizontal lift of $\dot{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathrm{St}(p,m) \times \mathbb{R}^{n \times p}_*$ is given by

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = M\Omega + M_\perp M_\perp^{\mathrm{T}} \dot{X}_{MN^{\mathrm{T}}} N(N^{\mathrm{T}}N)^{-1} \tag{37a}$$

$$\dot{X}_{\uparrow\mathrm{N}(M,N)} = N(N^{\mathrm{T}}N)^{-1}(S - \Omega) + N_\perp N_\perp^{\mathrm{T}} \dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}} M \tag{37b}$$

$$\text{where} \quad \Omega(N^{\mathrm{T}}N + I) + S = M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N, \ \Omega = -\Omega^{\mathrm{T}}, \ S = S^{\mathrm{T}}. \tag{37c}$$

Equation (37c) is equivalent to

$$\Omega(N^{\mathrm{T}}N + I) + (N^{\mathrm{T}}N + I)\Omega = M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N - N^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}}M, \tag{38a}$$

$$S = M^{\mathrm{T}}\dot{X}_{MN^{\mathrm{T}}}N - \Omega(N^{\mathrm{T}}N + I). \tag{38b}$$

(To see that (37c) implies (38), add the first equation of (37c) to its transpose. The equivalence can then be deduced from the fact that both systems of equations have one and only one solution, the former because the horizontal lift is known to exist and be unique, and the latter because (38a) is a Sylvester equation with a unique solution since, $(N^{\mathrm{T}}N + I)$ being symmetric positive definite, $(N^{\mathrm{T}}N + I)$ and $-(N^{\mathrm{T}}N + I)$ have no common eigenvalue.) As for the first two equations of (37), using (37c), they can be rewritten as

$$\dot{X}_{\uparrow\mathrm{M}(M,N)} = \dot{X}_{MN^{\mathrm{T}}}N(N^{\mathrm{T}}N)^{-1} - M(\Omega + S)(N^{\mathrm{T}}N)^{-1} \tag{39a}$$

$$\dot{X}_{\uparrow\mathrm{N}(M,N)} = \dot{X}_{MN^{\mathrm{T}}}^{\mathrm{T}}M + N\Omega. \tag{39b}$$

In summary,

Proposition 6.1 *Consider the submersion π (31) and the horizontal distribution (36). Let $(M,N) \in \mathrm{St}(p,m) \times \mathbb{R}^{n \times p}_*$ and let $\dot{X}_{MN^{\mathrm{T}}} \in \mathrm{T}_{MN^{\mathrm{T}}}\mathcal{M}(p, m \times n)$. Then the horizontal lift of $\dot{X}_{MN^{\mathrm{T}}}$ at (M,N) is $\dot{X}_{\uparrow(M,N)} = (\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})$ given by (39), where Ω is the solution of the Sylvester equation (38a) and S is given by (38b).*

6.5 Constitutive equation of horizontal lifts

From Proposition 6.1, routine manipulations lead to the following constitutive equation for horizontal lifts:

$$\dot{X}_{\uparrow\mathrm{M}(MR,NR)} = \dot{X}_{\uparrow\mathrm{M}(M,N)}R, \quad \dot{X}_{\uparrow\mathrm{N}(MR,NR)} = \dot{X}_{\uparrow\mathrm{N}(M,N)}R. \tag{40}$$

Hence we have the following counterpart of Proposition 5.2.

Proposition 6.2 *Consider the submersion π (31) and the horizontal distribution (36). Then a tangent vector field $\mathrm{St}(p,m) \times \mathbb{R}^{n \times p}_* \ni (M,N) \mapsto \dot{X}_{\uparrow(M,N)} \in \mathrm{T}_{(M,N)}\left(\mathrm{St}(p,m) \times \mathbb{R}^{n \times p}_*\right)$ is a horizontal lift if and only if (40) holds for all $(M,N) \in \mathrm{St}(p,m) \times \mathbb{R}^{n \times p}_*$ and all $R \in \mathrm{O}(p)$.*

6.6 Riemannian submersion

From Proposition 6.2 and the properties of the trace, it is direct that $\bar{g}$ (34) satisfies the invariance condition

$$\bar{g}_{(M,N)}(\dot{X}_{\uparrow(M,N)}, \check{X}_{\uparrow(M,N)}) = \bar{g}_{(MR,NR)}(\dot{X}_{\uparrow(MR,NR)}, \check{X}_{\uparrow(MR,NR)}). \tag{41}$$

Hence one consistently defines a Riemannian metric g on $\mathcal{M}(p, m \times n)$ by

$$g_{MN^\mathrm{T}}(\dot{X}_{MN^\mathrm{T}}, \check{X}_{MN^\mathrm{T}}) = \bar{g}_{(M,N)}(\dot{X}_{\uparrow(M,N)}, \check{X}_{\uparrow(M,N)}), \tag{42}$$

and $\pi : (\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*, \bar{g}) \to (\mathcal{M}(p, m \times n), g)$ is a Riemannian submersion.

6.7 Horizontal projection

We now obtain an expression for the projection $P^\mathrm{h}_{(M,N)}(\dot{M}, \dot{N})$ of $(\dot{M}, \dot{N}) \in \mathrm{T}_{(M,N)}\left(\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*\right)$ onto the horizontal space (36) along the vertical space (33). Since the projection is along the vertical space, we have

$$P^\mathrm{h}_{(M,N)}(\dot{M}, \dot{N}) = (\dot{M} + M\Omega, \dot{N} + N\Omega) \tag{43}$$

for some $\Omega = -\Omega^\mathrm{T} \in \mathbb{R}^{p \times p}$. It remains to obtain Ω by imposing horizontality of (43). The characterization of horizontal vectors given in (35) yields the Sylvester equation

$$(N^\mathrm{T} N + I)\Omega + \Omega(N^\mathrm{T} N + I) = \dot{M}^\mathrm{T} M - M^\mathrm{T} \dot{M} + \dot{N}^\mathrm{T} N - N^\mathrm{T} \dot{N}. \tag{44}$$

In summary:

Proposition 6.3 *The projection $P^\mathrm{h}_{(M,N)}(\dot{M}, \dot{N})$ of $(\dot{M}, \dot{N}) \in \mathrm{T}_{(M,N)}\left(\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*\right)$ onto the horizontal space (36) along the vertical space (33) is given by (43) where Ω is the solution of the Sylvester equation (44).*

6.8 Riemannian connection on the total space

Let P^St_M denote the orthogonal projection from $\mathbb{R}^{m \times p}$ onto $\mathrm{T}_M \mathrm{St}(p, m)$, given by (see [AMS08, Example 5.3.2])

$$P^\mathrm{St}_M \dot{M} = (I - MM^\mathrm{T})\dot{M} + M\mathrm{skew}(M^\mathrm{T} \dot{M}) = \dot{M} - M\mathrm{sym}(M^\mathrm{T} \dot{M}), \tag{45}$$

where $\mathrm{skew}(Z) := \frac{1}{2}(Z - Z^\mathrm{T})$ and $\mathrm{sym}(Z) := \frac{1}{2}(Z + Z^\mathrm{T})$. We also let $P^{\mathrm{St}\times\mathbb{R}}_{(M,N)}$ denote the orthogonal projection from $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$ onto $\mathrm{T}_{(M,N)}\left(\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*\right)$, given by

$$P^{\mathrm{St}\times\mathbb{R}}_{(M,N)}(\dot{M}, \dot{N}) = (P^\mathrm{St}_M \dot{M}, \dot{N}). \tag{46}$$

Since $\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*$, endowed with the Riemannian metric $\bar{g}$ (34), is a Riemannian submanifold of the Euclidean space $\mathbb{R}^{m \times p} \times \mathbb{R}^{n \times p}$, a classical result of Riemannian geometry (see [AMS08, §5.3.3]) yields that the Riemannian connection $\bar{\nabla}$ on $(\mathrm{St}(p, m) \times \mathbb{R}^{n \times p}_*, \bar{g})$ is given by

$$\bar{\nabla}_{\dot{X}}\dot{Y} = P^{\mathrm{St}\times\mathbb{R}}_{(M,N)}\partial_{\dot{X}}\dot{Y},$$

that is,

$$\left(\bar{\nabla}_{\dot{X}} \dot{Y}\right)_{\mathrm{M}} = P_M^{\mathrm{St}}(\partial_{\dot{X}} \dot{Y}_{\mathrm{M}}) \tag{47a}$$

$$\left(\bar{\nabla}_{\dot{X}} \dot{Y}\right)_{\mathrm{N}} = \partial_{\dot{X}} \dot{Y}_{\mathrm{N}} \tag{47b}$$

for all $(M, N) \in \mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p}$, all $\dot{X} \in \mathrm{T}_{(M,N)}\left(\mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p}\right)$ and all tangent vector fields $\dot{Y}$ on $\mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p}$.

6.9 Riemannian connection on the quotient space

As in Section 5.8, we can now provide an expression for the Riemannian connection ∇ on the manifold $\mathcal{M}(p, m \times n)$ endowed with the Riemannian metric g (42):

$$\begin{aligned}
(\nabla_{\dot{X}_{MN^{\mathrm{T}}}} \dot{Y})_{(M,N)} &= P_{(M,N)}^{\mathrm{h}}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}} \dot{Y}_{\uparrow}) \\
&= P_{(M,N)}^{\mathrm{h}} P_{(M,N)}^{\mathrm{St} \times \mathbb{R}} \partial_{\dot{X}_{\uparrow(M,N)}} \dot{Y}_{\uparrow},
\end{aligned}$$

with P^{h} as in (43) and $P^{\mathrm{St} \times \mathbb{R}}$ as in (46).

6.10 Riemannian Newton equation

Given $f : \mathcal{M}(p, m \times n) \to \mathbb{R}$, define $\bar{f} = f \circ \pi$, i.e.,

$$\bar{f} : \mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p} \to \mathbb{R} : (M, N) \mapsto f(MN^{\mathrm{T}}),$$

and define

$$\bar{\bar{f}} : \mathbb{R}_*^{m \times p} \times \mathbb{R}_*^{n \times p} \to \mathbb{R} : (M, N) \mapsto f(MN^{\mathrm{T}}).$$

Let $\partial \bar{\bar{f}}$ denote the Euclidean gradient of $\bar{\bar{f}}$. We have (see [AMS08, (3.37)])

$$\mathrm{grad}\, \bar{f}(M, N) = P_{(M,N)}^{\mathrm{St} \times \mathbb{R}} \partial \bar{\bar{f}}(M, N) \tag{48}$$

and (see [AMS08, (3.39)])

$$(\mathrm{grad}\, f)_{\uparrow(M,N)} = \mathrm{grad}\, \bar{f}(M, N),$$

where the left-hand side stands for the horizontal lift at (M, N) of $\mathrm{grad}\, f(MN^{\mathrm{T}})$.

We can now obtain the counterpart of the (lifted) Newton equation (28) with normalization on the M factor:

$$P_{(M,N)}^{\mathrm{h}}(\bar{\nabla}_{\dot{X}_{\uparrow(M,N)}} \mathrm{grad}\, \bar{f}) = -\mathrm{grad}\, \bar{f}(M, N), \tag{49}$$

where P^{h} is the horizontal projection given in Section 6.7, $\bar{\nabla}$ is the Riemannian connection on $(\mathrm{St}(p, m) \times \mathbb{R}_*^{n \times p}, \bar{g})$ given in Section 6.8, and $\mathrm{grad}\, \bar{f}$ is obtained from the Euclidean gradient of $\bar{\bar{f}}$ from (48).

The Newton equation (49) can be considered less intricate than in the non-orthonormal case (28) because the expression for $\bar{\nabla}$ in (47) is simpler than in (27). In any case, the discussion that follows (28) applies equally: the Newton equation is merely a linear system of equations, and the Riemannian overhead requires only $O(p^2(m + n + p))$ flops.

6.11 Newton's method

Another reward that comes with the orthonormalization of the M factor is that the Riemannian exponential with respect to g (42) admits a closed-form expression. First, we point out that, in view of [EAS98, §2.2.2], the Riemannian exponential on $\mathrm{St}(p,m) \times \mathbb{R}_*^{n \times p}$ for the Riemannian metric $\bar{g}$ (34) is given by

$$\mathrm{Exp}_{(M,N)}(\dot{M}, \dot{N}) = (\begin{bmatrix} M & \dot{M} \end{bmatrix} \exp \begin{bmatrix} A & -S \\ I & A \end{bmatrix} I_{2p,p} \exp(-A),\ N + \dot{N}), \tag{50}$$

where $A := M^{\mathrm{T}} \dot{M}$, $S := \dot{M}^{\mathrm{T}} \dot{M}$, $I_{2p,p}$ denotes the $2p \times p$ matrix with ones on its main diagonal and zeros elsewhere, and where exp stands for the matrix exponential (`expm` in Matlab). Second, since by [O'N83, Corollary 7.46] horizontal geodesics in $(\mathrm{St}(p,m) \times \mathbb{R}_*^{n \times p}, \bar{g})$ map to geodesics in $(\mathcal{M}(p, m \times n), g)$, we have that

$$\mathrm{Exp}_{MN^{\mathrm{T}}}(\dot{X}_{MN^{\mathrm{T}}}) = \pi(\mathrm{Exp}_{(M,N)}(\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})), \tag{51}$$

with $(\dot{X}_{\uparrow\mathrm{M}(M,N)}, \dot{X}_{\uparrow\mathrm{N}(M,N)})$ as in Proposition 6.1. (In (51), Exp on the right-hand side is given by (50) and Exp on the left-hand side denotes the Riemannian exponential of $(\mathcal{M}(p, m \times n), g)$.)

Observe that the matrix exponential is applied in (50) to matrices of size $2p \times 2p$ and $p \times p$; hence, when $p \ll m$, the cost of computing the M component of (50) is comparable to the cost of computing the simple sum $M + \dot{M}$. Note also that, in practice, the M component of the Newton iterates may gradually depart from orthonormality due to the accumulation of numerical errors; a remedy is to restore orthonormality by taking the Q factor of the unique QR decomposition where the diagonal of the R factor is positive.

We can now formally describe Newton's method in the context of this Section 6.

Theorem 6.4 (Riemannian Newton on $\mathcal{M}(p, m \times n)$ with Riemannian metric (42)**)**
Let f be a real-valued function on the Riemannian manifold $\mathcal{M}(p, m \times n)$ (1), endowed with the Riemannian metric g (42), with the associated Riemannian connection, and with the exponential retraction (51). Then the Riemannian Newton method [AMS08, Algorithm 5] for f maps $MN^{\mathrm{T}} \in \mathcal{M}(p, m \times n)$ to $\pi(\mathrm{Exp}_{(M,N)}(\dot{X}_{\mathrm{M}}, \dot{X}_{\mathrm{N}}))$, where π is given in (31), Exp is defined in (50), and $(\dot{X}_{\mathrm{M}}, \dot{X}_{\mathrm{N}})$ is the solution $\dot{X}_{\uparrow(M,N)}$ of the Newton equation (49).

The quadratic convergence result in Theorem 5.5 still holds, replacing the reference to Theorem 5.4 by a reference to Theorem 6.4.

7 Numerical experiments

In this section, we report on numerical experiments that illustrate the impact of the choice of the geometry on the performance of second-order low-rank optimization algorithms. The salient observation is that, whereas the geometries of Sections 5 and 6 yield quite similar performance in terms of the number of outer iterations in our experiments, the latter may require many more inner iterations than the former. This should be put in perspective with the recent finding in [MMBS12, §5.1] that the geometry of Section 6 yields much poorer performance in a first-order optimization method. These conclusions are however based on a limited number of experiments conducted on one class of low-rank optimization problems. A wider set of experiments is needed to confirm these trends.

This section, written after the first version of this paper became public, conveniently benefited from recent developments in this very active area of research. In particular, the geometry of Section 6 was implemented in the Manopt toolbox [BM13] by Nicolas Boumal and Bamdev Mishra.

As in [MMBS12], we conduct numerical experiments on low-rank matrix completion problems. The objective function f on $\mathcal{M}(p, m \times n)$ is given by

$$f(X) = \frac{1}{2}\|\mathcal{P}_\Omega(X - C)\|_F^2,$$

where C is a given $m \times n$ matrix, $\Omega \subset \{1, \ldots, m\} \times \{1, \ldots, n\}$ is the set of indices of the observed entries, $\mathcal{P}_\Omega$ sets to zero the entries not in Ω while leaving the other entries unchanged, and $\|\cdot\|_F$ denotes the Frobenius norm.

We tackle the optimization problem (2) with a Riemannian trust-region Newton approach [ABG07, AMS08]. Specifically, we follow the Riemannian trust-region (RTR) meta-algorithm [AMS08, Alg. 10] where the trust-region subproblems are solved approximately using the truncated CG (tCG) method [AMS08, Alg. 11] with the inner stopping criterion [AMS08, (7.10)], i.e.,

$$\|r_{j+1}\| \leq \|r_0\| \min(\|r_0\|^\theta, \kappa) \tag{52}$$

with r_j denoting the residual in the jth step of the tCG inner iteration. The resulting algorithm depends on the choice of (i) the Riemannian metric on $\mathcal{M}(p, m \times n)$, (ii) the affine connection on $\mathcal{M}(p, m \times n)$ that enters the definition [AMS08, (7.2)] of the Hessian involved in the quadratic model [AMS08, (7.1)], and (iii) the retraction on $\mathcal{M}(p, m \times n)$. For the affine connection, we invariably choose the Riemannian connection, which is fully specified by the Riemannian metric. For the Riemannian metric and the retraction, we use those provided in Sections 5 (Riemannian metric induced by (12) and non-exponential retraction (29)) and 6 (Riemannian metric induced by (34) and exponential retraction (51)).

Working out the lifted expression of the gradient and Hessian of f in both geometries is rather straightforward. In the process, one obtains that $\partial \bar{f}(M, N)$ (or $\partial \bar{\bar{f}}(M, N)$ in the notation of Section 6) is given by $(\mathcal{P}_\Omega(MN^{\mathrm{T}} - C)N, (\mathcal{P}_\Omega(MN^{\mathrm{T}} - C))^{\mathrm{T}}M)$.

The experiments are run with Matlab R2011b using the Manopt toolbox [BM13]. Matrix C is generated as $C = AB^{\mathrm{T}}$, where A of size $m \times p$ and B of size $n \times p$ are drawn from the standard normal distribution. The index set Ω is chosen uniformly at random with a sampling ratio of $5d/(mn)$, where $d = p(m + n - p)$ is the dimension of $\mathcal{M}(p, m \times n)$. The initial iterate $X_0 = M_0 N_0^{\mathrm{T}}$ is chosen as follows: compute the QR decomposition $A = M_* R$, set $N_* = BR^{\mathrm{T}}$ (hence $C = M_* N_*^{\mathrm{T}}$), compute the QR decomposition $M_* + \epsilon E_M = M_0 \tilde{R}$ where $\epsilon > 0$ and E_M is drawn from the standard normal distribution, finally set $N_0 = N_* + \epsilon E_N$ where E_N is drawn from the standard normal distribution. Parameters θ and κ in the inner stopping criterion (52) are set to 1 (aiming at quadratic convergence) and 10^{-1}, respectively.

In order to assess the convergence of the algorithms to a stationary point of the optimization problem, we monitor the evolution of $\|\mathrm{grad}\, f(M_k N_k^{\mathrm{T}})\|$, where $\mathrm{grad}\, f$ stands for the gradient in the embedded geometry [Van13] and $\|\cdot\|$ is the Frobenius norm. In the context of our comparative experiments, this is a more neutral comparison measure than the norm of the gradient in one of the two proposed geometries. We also report the number of inner (tCG) iterations.

In a first set of experiments, we chose $m = n = 100$ and $p = 5$. We set $\epsilon = 10^{-8}$, a small value, in order to focus on the asymptotic behavior of the algorithms. In this setting, we

observed in our experiments that the boundary of the trust region was never reached, which means that the trust-region method is equivalent to an inexact Newton method, where the inexactness comes from the fact that the Newton equation is solved approximately by means of a (linear) CG iteration stopped by the criterion (52). The results for a typical run are provided in Table 1.

iter	Section 5		Section 6	
	$\|\text{grad}\|$	inner iter	$\|\text{grad}\|$	inner iter
0	1.185670e-05	0	1.185670e-05	0
1	1.366235e-08	10	1.065745e-07	7
2	1.677951e-14	24	9.751445e-13	128

Table 1: Numerical experiments with $m = n = 100$, $p = 5$, $\epsilon = 10^{-8}$.

In a second set of experiments, we chose $m = 10$, $n = 100$, $p = 5$ and $\epsilon = 10^{-6}$. The results for an exemplative run are displayed in Table 2.

iter	Section 5		Section 6	
	$\|\text{grad}\|$	inner iter	$\|\text{grad}\|$	inner iter
0	1.814751e-04	0	1.814751e-04	0
1	2.095396e-06	9	2.195920e-05	7
2	7.887987e-11	14	1.097058e-13	16

Table 2: Numerical experiments with $m = 10$, $n = 100$, $p = 5$, $\epsilon = 10^{-6}$.

Finally, in order to compare the global behavior of the algorithms, we conducted experiments with $\epsilon = 1$; see Table 3 for an illustrative instance.

The superlinear (quadratic) local convergence predicted by the theory is clearly visible in all the experiments. The geometry of Section 6 is seen to yield slightly faster convergence on some problem instances with respect to the number of (outer) iterations. However, the geometry of Section 6 also tends to produce many more inner (tCG) iterations. This remark points in the direction of [MMBS12, §5.1] where the geometry of Section 6 was observed to yield much slower convergence in a first-order method.

8 Conclusion

We have reached the end of a technical hike that led us to give in Theorem 6.4 what is, to the best of our knowledge, the first closed-form description of a purely Riemannian Newton method on the set of all matrices of fixed dimension and rank. By "closed-form", we mean that, besides calling an oracle for Euclidean first and second derivatives, the method only needs to perform elementary matrix operations, solve linear systems of equations, and compute (small-size) matrix exponentials. By "purely Riemannian", we mean that it uses the tools provided by Riemannian geometry, namely, the Riemannian connection (instead of any other affine connection) and the Riemannian exponential (instead of any other retraction).

The developments strongly rely on the theory of Riemannian submersions and are based on factorizations of low rank matrices X as MN^{T}, where one of the factors is orthonormal.

iter	Section 5		Section 6	
	$\|\text{grad}\|$	inner iter	$\|\text{grad}\|$	inner iter
0	1.636720e+02	0	1.636720e+02	0
1	1.636720e+02	3	1.636720e+02	2
2	1.636720e+02	3	1.636720e+02	2
3	1.636720e+02	3	1.636720e+02	2
4	1.636720e+02	3	1.636720e+02	2
5	1.636720e+02	3	1.636720e+02	2
6	1.636720e+02	3	5.085657e+01	1
7	1.216769e+02	2	1.563970e+01	4
8	1.216769e+02	4	1.060595e+01	2
9	6.217289e+01	3	6.700386e+00	9
10	3.728346e+01	3	5.495128e+00	4
11	2.345336e+01	6	1.381830e+00	15
12	6.441494e+00	4	5.603172e-01	15
13	9.937363e-01	4	2.611366e-02	35
14	7.861386e-02	4	4.312611e-03	25
15	5.051588e-03	5	8.784949e-05	48
16	4.540472e-04	5	1.530663e-08	103
17	2.253013e-05	6	4.196742e-14	193
18	4.037523e-08	11		
19	1.639055e-13	21		

Table 3: Numerical experiments with $m = n = 100$, $p = 5$, $\epsilon = 1$.

Relaxing the orthonormality constraint is more appealing for its symmetry (the two factors are treated alike), but it did not allow us to obtain a closed-form expression for the Riemannian exponential.

Acknowledgment

References

[ABG07]   P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, July 2007. `doi:10.1007/s10208-005-0179-9`.

[ADM⁺02]   Roy L. Adler, Jean-Pierre Dedieu, Joseph Y. Margulies, Marco Martens, and Mike Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, July 2002. `doi:10.1093/imanum/22.3.359`.

[ADY09]   Luca Amodei, Jean-Pierre Dedieu, and Jean-Claude Yakoubsohn. A dynamical approach to low-rank approximation of a matrix, 2009. communication at the 14th Belgian-French-German Conference on Optimization, 14–18 September 2009.

[AMR88]   R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

[AMS08]   P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. URL: `http://sites.uclouvain.be/absil/amsbook/`.

[BA11]   Nicolas Boumal and P.-A. Absil. RTRMC: A Riemannian trust-region method for low-rank matrix completion. In J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F.C.N. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24 (NIPS)*, pages 406–414. 2011. URL: `http://perso.uclouvain.be/pa.absil/RTRMC/`.

[BM13]   Nicolas Boumal and Bamdev Mishra. The Manopt toolbox. `http://www.manopt.org`, 2013. version 1.0.1.

[dC92]   Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[DKM12]   Wei Dai, E. Kerman, and O. Milenkovic. A geometric approach to low-rank matrix completion. *IEEE T. Inform. Theory*, 58(1):237 –247, jan. 2012. `doi:10.1109/TIT.2011.2171521`.

[DMK11]   Wei Dai, O. Milenkovic, and E. Kerman. Subspace evolution and transfer (set) for low-rank matrix completion. *IEEE T. Signal Proces.*, 59(7):3120–3132, july 2011. `doi:10.1109/TSP.2011.2144977`.

[DS83]   John E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice Hall Series in Computational Mathematics. Prentice Hall Inc., Englewood Cliffs, NJ, 1983.

[EAS98]   Alan Edelman, Tomás A. Arias, and Steven T. Smith.   The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998. `doi:10.1137/S0895479895290954`.

[GR85]   A. Griewank and G. W. Reddien. The approximation of simple singularities. In *Numerical boundary value ODEs (Vancouver, B.C., 1984)*, volume 5 of *Progr. Sci. Comput.*, pages 245–259. Birkhäuser Boston, Boston, MA, 1985.

[HM94]   Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994. With a foreword by R. Brockett.

[HS95]   Uwe Helmke and Mark A. Shayman. Critical points of matrix least squares distance functions. *Linear Algebra Appl.*, 215:1–19, 1995. `doi:10.1016/0024-3795(93)00070-G`.

[KMO10]   Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries, 2010. `arXiv:0906.2027v2`.

[MAS12]   B. Mishra, K. Adithya Apuroop, and R. Sepulchre. A Riemannian geometry for low-rank matrix completion, 2012. `arXiv:1211.1550v2`.

[Mey11]   Gilles Meyer. *Geometric optimization algorithms for linear regression on fixed-rank matrices*. PhD thesis, University of Liège, 2011.

[MMBS11]   B. Mishra, G. Meyer, F. Bach, and R. Sepulchre. Low-rank optimization with trace norm penalty, 2011. `arXiv:1112.2318v1`.

[MMBS12]   B. Mishra, G. Meyer, S. Bonnabel, and R. Sepulchre. Fixed-rank matrix factorizations and Riemannian low-rank optimization, 2012. `arXiv:1209.0430v1`.

[MMS11]   B. Mishra, G. Meyer, and R. Sepulchre. Low-rank optimization for distance matrix completion. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 4455–4460, dec. 2011. `doi:10.1109/CDC.2011.6160810`.

[O'N66]   Barrett O'Neill. The fundamental equations of a submersion. *Michigan Math. J.*, 13:459–469, 1966.

[O'N83]   Barrett O'Neill. *Semi-Riemannian Geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[SE10]   Lennart Simonsson and Lars Eldén. Grassmann algorithms for low rank approximation of matrices with missing values. *BIT Numer. Math.*, 50(1):173–191, 2010. `doi:10.1007/s10543-010-0253-9`.

[Smi94]   Steven T. Smith. Optimization techniques on Riemannian manifolds. In Anthony Bloch, editor, *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Inst. Commun.*, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[Van13]   B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM J. Optim.*, 23(2):1214–1236, 2013.