DEPARTMENT OF MATHEMATICAL SCIENCES

TMA4500 - INDUSTRIAL MATHEMATICS, SPECIALIZATION PROJECT

Riemannian Optimization using second order information on the Symplectic Stiefel manifold

Author:
Ole Gunnar Røsholt Hovland

What date should I use?

Table of Contents

List of Figures

List of Tables

1 Introduction

2 Introductory Theory

2.1 Foundational definitions

This section is designed to be a reference work to set notation, and to ensure that the reader has the necessary background to understand the optimization algorithms we will be studying.

Definition 1 (General Linear group)**.** *The* real General Linear group *is defined as the set of all invertible matrices in* $\mathbb{R}^{n\times n}$*, denoted by* $\mathrm{GL}(n)$*. [4, Example 9.11]*

Definition 2 (Orthogonal group)**.** *The* real Orthogonal group *is defined as the set of all orthogonal matrices in* $\mathbb{R}^{n\times n}$*, denoted by* $\mathrm{O}(n)$*. [5, p. 3]*

Definition 3 (Quotient manifold)**.** *We define the definition of quotient manifold as in [1, p. 27]. Let* $\mathcal{M}$ *be a manifold equipped with the operation* $\sim$ *called the* equivalence relation*. The equivalence relation has the following properties:*

1. *(reflexive)* $p \sim p$ *for all* $p \in \mathcal{M}$*,*
2. *(symmetric)* $p \sim q$ *if and only if* $q \sim p$ *for all* $q, p \in \mathcal{M}$*, and*
3. *(transitive) given* $p \sim q$ *and* $q \sim r$ *this implies that* $p \sim r$ *for all* $p, q, r \in \mathcal{M}$*.*

Given the set $[p] := \{q \in \mathcal{M} \colon q \sim p\}$ *called the* equivalence class *of all points equivalent to* p*, the set*

$$\mathcal{M}/\sim\, := \{[p] \mid p \in \mathcal{M}\}$$

is called the quotient of $\mathcal{M}$ by $\sim$*. It is the set of all equivalence classes of* $\sim$ *in* $\mathcal{M}$*. The mapping* $\pi \colon \mathcal{M} \to \mathcal{M}/\sim$ *called the* natural- *or* canonical projection*, defined by* $p \mapsto [p]$ *.*

Definition 4 (Tangent Space)**.** *Following [4, Def. 8.33], for a point* p *on a smooth manifold* $\mathcal{M}$*, denote the set of smooth curves [4, Def. 8.5] passing through* p *at* $t = 0$ *as* C_p*. This means that* $\alpha(0) = p$ *for all* $\alpha \in C_p$*. For* $\alpha, \beta \in C_p$ *we say that they are equivalent if*

$$(\phi \circ \alpha)'(0) = (\phi \circ \beta)'(0),$$

meaning their derivatives match in a coordinate chart *(defined as in [7, p. 4]) if their derivatives in the coordinate chart at zero are equal. Denote this equivalence relation as* $\alpha \sim \beta$*. It has analogous properties to the equivalence relation in Definition 3. The equivalence class is defined as* $[\alpha] = \{\beta \in C_p \mid \alpha \sim \beta\}$*. Every equivalence class is called a* tangent vector *to* $\mathcal{M}$ *at* p*. The* tangent space *at* p *is the quotient set*

$$T_p\mathcal{M} = C_p/\sim\, = \{[\alpha] \mid \alpha \in C_p\}\,.$$

Definition 5 (Riemannian manifold)**.** *As defined in [3, def 2.6, p. 179]: a* smooth manifold $\mathcal{M}$*, as defined in [7, p. 13], is a* Riemannian manifold *if we can define a field of symmetric, positive definite, bilinear forms* $g(\cdot,\cdot)$*, called the Riemannian metric. By field we mean that* g_p *is defined on the tangent space* $T_p\mathcal{M}$ *at each point* $p \in \mathcal{M}$ *[3, def 2.1, p. 178]. We will assume that* g *is smooth, meaning that it is of class* $\mathcal{C}^\infty$*.*

Definition 6 (Vector field on Riemannian manifold)**.** *Following Appendix A of [6], a smooth vector field* $\mathcal{X} \colon \mathcal{M} \to T\mathcal{M}$*,* $p \mapsto \mathcal{X}(p) \in T_p\mathcal{M}$ *on a Riemannian manifold* $\mathcal{M}$ *can be expressed through local coordinates as*

$$\mathcal{X}(p) = \sum_{i=1}^{n} \alpha_i \partial_i =: \alpha^{\mathrm{T}} \partial,$$

where $\alpha \in \mathbb{R}^n$*, and* ∂ *is the canonical basis of* $T_p\mathcal{M}$*.*

Definition 7 (Horizontal & Vertical Space)**.** *Using Definition 3, given a Riemannian manifold $\overline{\mathcal{M}}$ with Riemannian metric $\overline{g}$, denote a quotient manifold of $\overline{\mathcal{M}}$ as $\mathcal{M} = \overline{\mathcal{M}}/\sim$. Following the definitions in Absil et al. [1, p. 43], for a point $p \in \mathcal{M}$, the equivalence class $[p] = \pi^{-1}(p)$ induces an embedded submanifold of $\overline{\mathcal{M}}$ (see Definition 3), hence it admits a tangent space,*

$$\mathcal{V}_{\overline{p}} = T_{\overline{p}}(\pi^{-1}(p))$$

named the vertical space *at $\overline{p}$. Canonically chosen as the orthogonal complement of $\mathcal{V}_{\overline{p}}$ in $T_{\overline{p}}\overline{\mathcal{M}}$, the* horizontal space *[1, p. 48] is defined as*

$$\mathcal{H}_{\overline{p}} \coloneqq \mathcal{V}_{\overline{p}}^{\perp} = \{Y_{\overline{p}} \in T_{\overline{p}}\overline{\mathcal{M}} \mid \overline{g}(Y_{\overline{p}}, Z_{\overline{p}}) = 0 \quad \forall \quad Z_{\overline{p}} \in \mathcal{V}_{\overline{p}}\}.$$

The horizontal lift *at $\overline{p} \in \pi^{-1}(p)$ of a tangent vector $X_p \in T_p\mathcal{M}$ is the unique tangent vector $X_{\overline{p}} \in \mathcal{H}_{\overline{p}}$ that satisfies $\mathrm{D}\pi(\overline{p})[X_{\overline{p}}] = X_p$. Note that given the horizontal space on $\overline{\mathcal{M}}$, $\mathcal{H}_{\overline{p}} \oplus \mathcal{V}_{\overline{p}} = T_{\overline{p}}\mathcal{M}$, where $\oplus$ denotes the Whitney sum.*

define or ref w sum? i wrote it in obsidian in the H and V space file.

Definition 8 (Riemannian connection)**.** *The* Riemannian connection, *also known as the* Levi-Civita connection, *is the unique affine connection which is torsion free, and metric compatible [9, Def. 6.4]. In Appendix A of [6], denoting $\mathfrak{X}(\mathcal{M})$ as the space of smooth vector fields on $\mathcal{M}$, it is defined as the unique $\mathbb{R}$-bilinear smooth map on $\mathcal{M}$ with riemannian metric $\langle\cdot,\cdot\rangle_p$*

$$\nabla\colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}), \quad (X, Y) \mapsto \nabla_X Y,$$

such that the following properties hold. Given $X, Y, Z \in \mathfrak{X}(\mathcal{M})$, and $f \in \mathcal{C}^\infty(M)$, $\nabla_X Y$ has the following properties:

1. *(first argument linearity) $\nabla_{fX} Y = f\nabla_X Y$,*
2. *(Leibnitz) $\nabla_X(fY) = (Xf)Y + f\nabla_X Y$,*
3. *(torsion free) $\nabla_X Y - \nabla_Y X = [X, Y]$, where $[\cdot,\cdot]$ is the Lie bracket, and*
4. *(metric compatibility) $Z\langle X, Y\rangle = \langle\nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$.*

In JZ this was stated wrongly

citation

Definition 9 (Christoffel symbols)**.** *The method we will employ to completely describe a connection (as defined in Definition 8) locally is to describe them through* Christoffel symbols. *Following the definition of [9, p. 100], let ∇ be an affine connection on $\mathcal{M}$. Denote a coordinate vector field on the coordinate open set $(U, p^1, \ldots, p^n) \subseteq \mathcal{M}$ by $\partial_i \coloneqq \partial/\partial p^i$. In this coordinate frame there exist the numbers called Christoffel symbols, Γ^k_{ij}, defined through the following*

do i need a source to define this?

$$\nabla_{\partial_i}\partial_j = \sum_{k=1}^{n} \Gamma^k_{ij}\partial_k \eqqcolon \boldsymbol{\Gamma}^{\mathrm{T}}_{ij}\partial.$$

Definition 10 (Retraction)**.** *Following [4, Def. 3.47], a retraction on a smooth manifold $\mathcal{M}$ is a smooth map,*

$$\mathcal{R}\colon T\mathcal{M} \to \mathcal{M}, \quad (p, X) \mapsto \mathcal{R}_p(X)$$

such that every curve generated from $c(t) = \mathcal{R}_p(tX)$ satisfies $c(0) = p$ and $\dot{c}(0) = X$. Equivalently the conditions can be stated as in [4, p. 40] without the use of curves. For all $p \in \mathcal{M}$, $\mathcal{R}_p(0) = p$, and $\mathrm{D}\,\mathcal{R}_p(0)\colon T_p\mathcal{M} \to T_p\mathcal{M}$, $\mathrm{D}\,\mathcal{R}_p(0)[X] = X$ is the identify map.

Definition 11.

Definition 12.

For the rest of this paper we denote $\mathcal{M}$ as being a Riemannian manifold.

After the basic definitions, talk about how the rest of the theory is a highlighted summary through BZ and JZ. The goal is to look at the findings in JZ, however it relies heavely on theory derived in BZ. It will be mentioned of some parts are from other works, or if they are original work.

2.2 The Symplectic group

The real *symplectic group* is the space overarching the symplectic Stiefel manifold, and we will look at this space first. To be able to define the symplectic group we first need some preliminary definitions. Define the *symplectic identity* as the following block matrix,

Why do we need sp? We will use quotient properties to map stuff to Spst

$$J_{2n} := \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix},$$

where I_n denotes the $n \times n$ identity matrix. J_{2n} has some properties we will take advantage of frequently:

$$J_{2n}^{\mathrm{T}} = -J_{2n} = J_{2n}^{-1} \tag{2.1}$$

The symplectic group is defined as the set of matrices which define the symplectic structure in the following sense. We define the real symplectic group as

$$\mathrm{Sp}(2n) := \{p \in \mathbb{R}^{2n\times 2n} \mid p^+ p = I_{2n}\}, \tag{2.2}$$

where $^+$ is defined as the *symplectic inverse* of any matrix $q \in \mathbb{R}^{2n\times 2k}$ such that

$$q^+ := J_{2k}^{\mathrm{T}} q^{\mathrm{T}} J_{2n}. \tag{2.3}$$

The Lie algebra of $\mathrm{Sp}(2n)$ is the symplectic groups' tangent space at the identity. It is given by

$$\mathfrak{sp}(2n) := \{\Omega \in \mathbb{R}^{2n\times 2n} \mid \Omega^+ = -\Omega\}, \tag{2.4}$$

where Ω is called the Hamiltonian matrix ref ??. Now we can define the tangent space of $\mathrm{Sp}(2n)$ at any point p as

Do I want to find a historical reference?

$$T_p\mathrm{Sp}(2n) = \{p\Omega, \Omega p \in \mathbb{R}^{2n\times 2n} \mid \Omega \in \mathfrak{sp}(2n)\}. \tag{2.5}$$

maybe explain a little more how we get this tangent space

2.3 The Symplectic Stiefel manifold

Now that we have defined the symplectic group, we introduce the manifold of interest, the real symplectic Stiefel manifold. It is defined as

$$\mathrm{SpSt}(2n, 2k) := \{p \in \mathbb{R}^{2n\times 2k} \mid p^+ p = I_{2k}\}, \tag{2.6}$$

where p^+ is as in (2.3). Following [2, Prop. 3.1], it is explicitly connected to the symplectic group in the sense that $\mathrm{SpSt}(2n, 2k)$ is diffeomorphic to the following quotient manifold of $\mathrm{Sp}(2n)$:

$$\mathrm{SpSt}(2n, 2k) \cong \mathrm{Sp}(2n)/\mathrm{Sp}(2(n-k)),$$

where the notion of quotient manifold is as in Definition 3. It has dimension $\dim(\mathrm{SpSt}(2n, 2k)) = (2n - 2k + 1)k$.

The following piece of insight can give some further intuition on what the Symplectic Stiefel manifold is. We note that the Stiefel manifold is a quotient space, as defined in Definition 3, of the orthogonal group as defined in Definition 2 such that $\mathrm{St}(2n, 2k) = \mathrm{O}(n)/\mathrm{O}(n-k)$.

Is a version of this info too much of a detour?

The expression for the tangent space follows straightforwardly from the definition of $\mathrm{SpSt}(2n, 2k)$. Assume we have a curve, $c(t) \in \mathrm{SpSt}(2n, 2k)$, s.t. $c(0) = p$ and $\dot{c}(0) := \frac{d}{dt}c(t)|_{t=0} = X$. Since $c(t)$ is a curve in $\mathrm{SpSt}(2n, 2k)$, by (2.6) it must satisfy the following condition:

$$c(t)^{\mathrm{T}} J_{2n} c(t) = J_{2k}. \tag{2.7}$$

Taking the derivative of (2.7) with respect to t at $t = 0$ we get

$$\dot{c}^{\mathrm{T}}(t) J_{2n} c(t) + c^{\mathrm{T}}(t) J_{2n} \dot{c}\big|_{t=0} = X^{\mathrm{T}} J_{2n} p + p^{\mathrm{T}} J_{2n} X = 0_{2k}.$$

After moving the first term over to the left hand side, and multiplying with J_{2n} from the left, we get

$$p^+ X = -X^+ p.$$

We recognize this condition as $p^+ X \in \mathfrak{sp}(2k)$ as defined in (2.4). This means that for a point p,

$$T_p\mathrm{SpSt}(2n, 2k) = \left\{X \in \mathbb{R}^{2n\times 2k} \;\middle|\; p^+ X \in \mathfrak{sp}(2k)\right\}. \tag{2.8}$$

3 Right-Invariant Framework on the Symplectic Stiefel Manifold

One of the key insights of Bendokat and Zimmermann [2, p. 11] is that using a right invariant framework one is able to construct geodesics on $\mathrm{SpSt}(2n, 2k)$. In this section we will first define a right-invariant metric on $\mathrm{Sp}(2n)$ and its corresponding geodesics, then transport this metric to $\mathrm{SpSt}(2n, 2k)$. This will allow us to define geodesics on $\mathrm{SpSt}(2n, 2k)$.

3.1 Right-invariant metric

We begin by defining the point-wise right-invariant metric on $\mathrm{Sp}(2n)$ as the mapping $g_p^{\mathrm{Sp}} \colon T_p\mathrm{Sp}(2n) \times T_p\mathrm{Sp}(2n) \to \mathbb{R}$,

$$g_p^{\mathrm{Sp}}(X_1, X_2) := \frac{1}{2}\operatorname{tr}((X_1 p^+)^T X_2 p^+), \quad X_1, X_2 \in T_p\mathrm{Sp}(2n). \tag{3.1}$$

It is right-invariant in the sense that $g_{pq}^{\mathrm{Sp}}(X_1 q, X_2 q) = \frac{1}{2}\operatorname{tr}((X_1 q q^+ p^+)^T X_2 q q^+ p^+) = g_p^{\mathrm{Sp}}(X_1, X_2)$ for all $p \in \mathrm{Sp}(2n)$. Now that we have defined g_p^{Sp}, we want to, in a sense, transport it to $\mathrm{SpSt}(2n, 2k)$ in a way that preserves the right-invariance. To achieve this we will use a *horizontal lift* to define a metric on $\mathrm{SpSt}(2n, 2k)$ through 3.1. Split $T_p\mathrm{Sp}(2n)$ into to parts: the horizontal- and vertical part, with respect to g_p^{Sp} and π:

rewrite these two sentences

$$T_p\mathrm{Sp}(2n) = \mathrm{Ver}_p^\pi \mathrm{Sp}(2n) \oplus \mathrm{Hor}_p^\pi \mathrm{Sp}(2n). \tag{3.2}$$

Define $\mathrm{Ver}_p^\pi \mathrm{Sp}(2n)$ and $\mathrm{Hor}_p^\pi \mathrm{Sp}(2n)$ in a smart way. Maybe just a reference if I do not have the space.

Rework the following paragraph. see obsidian

The point-wise right-invariant Riemannian metric on $\mathrm{SpSt}(2n, 2k)$ is defined as the mapping $g_p \colon T_p\mathrm{SpSt}(2n, 2k) \times T_p\mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, $g_p(X_1, X_2) := g_p^{\mathrm{Sp}}((X_1)_p^{\mathrm{hor}}, (X_2)_p^{\mathrm{hor}})$. More explicitly

Look at 20, Theorem 2.28 in BZ to maybe skip the derivation up to gp

$$g_p(X_1, X_2) = \operatorname{tr}\left(X_1^T \left(I_{2n} - \frac{1}{2} J_{2n}^T p (p^T p)^{-1} p^T J_{2n}\right) X_2 (p^T p)^{-1}\right), \tag{3.3}$$

for $X_1, X_2 \in T_p\mathrm{SpSt}(2n, 2k)$. For this metric, π denotes a Riemannian submersion.

pi here is weird, change this so it is more clear what it does

3.2 Geodesics

To define geodesics on $\mathrm{SpSt}(2n, 2k)$, we will first define them on $\mathrm{Sp}(2n)$.

Give more detail on the roadmap of this section. Maybe also write a motivation for why we want to introduce geodesics

Following [2, Prop. 2.1], given $p \in \mathrm{Sp}(2n)$, $X \in T_p\mathrm{Sp}(2n)$ and the right-invariant Riemannian metric (3.1), the respective geodesic $\gamma(t)$ is defined as

$$\gamma(t) := \exp_p\big(t(Xp^+ - (Xp^+)^{\mathrm{T}})\big)\exp_p\big(t(Xp^+)^{\mathrm{T}}\big)p,$$

where $\gamma(0) = p$, $\dot\gamma(0) = X$ and $^+$ is the symplectic inverse (as in (2.3)). Bendokat and Zimmermann proposes that "the proof of cite[Prop. 4.2]something can be transferred straightforwardly to [the settinge we are in]".

is this sentence superfluous?

ref

3.3 Riemannian gradient of the Symplectic Stiefel manifold

Now that we have chosen a metric, we can justify a choice for a Riemannian gradient.

Proposition 1. *Given a function $f\colon \mathrm{SpSt}(2n,2k) \to \mathbb{R}$, the Riemannian gradient with respect to g_p is given by*

add general def of R grad.?

$$\operatorname{grad} f(p) = \nabla f(p) p^T p + J_{2n} p (\nabla f(p))^T J_{2n} p, \tag{3.4}$$

where $\nabla f(p)$ is the Euclidean gradient of a smooth extension around $p \in \mathrm{SpSt}(2n,2k)$ in $\mathbb{R}^{2n\times 2k}$ at p.

Proof. We can see that this is the Riemannian gradient by the following two observations stated in [**BZ**], which we verify ourselves below.

Firstly, gradient must be in $T_p\mathrm{SpSt}(2n,2k)$, which means by (2.8) that $0 = p^+ \operatorname{grad} f(p) + (\operatorname{grad} f(p))^+ p$. Computing this we get

$$p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} J J p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p = 0,$$

where we have used $JJ = -J^{\mathrm{T}} J = -I_{2n}$ and (2.1).

Secondly, the gradient also has to satisfy $g_p(\operatorname{grad} f(p), X) = \mathrm{d}\, f_p(X) = \operatorname{tr}((\nabla f(p))^{\mathrm{T}} X)$ for all $X \in T_p\mathrm{SpSt}(2n,2k)$:

$$g_p(\operatorname{grad} f(p), X) = \operatorname{tr}\big((p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}})(I_{2n} - \tfrac{1}{2} G) X (p^{\mathrm{T}} p)^{-1}\big),$$

where $G \coloneqq J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J$. Expanding this expression we obtain

$$\begin{aligned} &= \operatorname{tr}\big(p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big) - \tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} G X (p^{\mathrm{T}} p)^{-1}\big) \\ &+ \operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big) - \tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} G X (p^{\mathrm{T}} p)^{-1}\big), \end{aligned}$$

where the cancellations used the fact that the trace is invariant under circular shifts. Noting that the first term is by definition $\mathrm{d}\, f_p(X)$, and inserting the definition of G, the expression becomes

$$\begin{aligned} &= \mathrm{d}\, f_p(X) - \tfrac{1}{2}\operatorname{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X\big) \\ &+ \operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big) \\ &- \tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X (p^{\mathrm{T}} p)^{-1}\big). \end{aligned}$$

After using $J^{\mathrm{T}} J^{\mathrm{T}} = -I_{2n}$ and (2.1) on the last term, we notice that we can cancel $p^{\mathrm{T}} p (p^{\mathrm{T}}) p^{-1}$, making it equal to the second to last term. Now focusing on the second term: for the first equality we use the fact that for any matrix, A, $\operatorname{tr}(A) = \operatorname{tr}(A^{\mathrm{T}})$, and for the second equality we utilize the cyclic property of the trace, and (2.1),

$$\begin{aligned} \tfrac{1}{2}\operatorname{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X\big) &= \tfrac{1}{2}\operatorname{tr}\big(X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J \nabla f(p)\big) \\ &= -\tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1}\big) \end{aligned} \tag{3.5}$$

Inserting (3.5) into our expression we end up with:

$$\mathrm{d}\, f_p(X) = \mathrm{d}\, f_p(X) + \tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J \nabla f(p) X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1}\big) + \tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big),$$

where the last two terms cancel after applying (2.1), and the tangent space condition (2.8), $p^{\mathrm{T}} J X = -X^{\mathrm{T}} J p$. □

3.4 Riemannian Hessian of the Symplectic Stiefel manifold

Define the remaining preliminaries for Hessian. Make an intro to section. loosly following Appendix A in Jz

For two smooth vector fields, $\mathcal{X}(p) = \alpha^{\mathrm{T}}\partial$ and $\mathcal{Y}(p) = \beta^{\mathrm{T}}\partial$ defined as in Definition 6, the covariant derivative (defined through the Riemannian connection by Definition 8) written in local coordinates is

$$\nabla_{\mathcal{X}}\mathcal{Y} = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \partial_i(\beta_j)\partial_j + \alpha_i\beta_j \sum_{k=1}^{n} \Gamma_{ij}^{k}\partial_k.$$

In preparation for the Hessian, we include the description of the covariant derivative from [9, p. 96] to restrict the covariant derivative further. We want to define the covariant derivative of a vector field along a curve $c(t)$. $c(t)$ is the smooth curve, $c\colon I \to \mathcal{M}$, $t \mapsto (\gamma_1(t), \dots, \gamma_n(t))$, where $I := [a, b] \subseteq \mathbb{R}$. Since we have an affine connection on $\mathcal{M}$, the following unique map exists:

$$\frac{D}{\mathrm{d}t}\colon \Gamma(T\mathcal{M}|_{c(t)}) \to \Gamma(T\mathcal{M}|_{c(t)}),$$

where $\Gamma(T\mathcal{M}|_{c(t)})$ denotes the the vector space of all smooth vector fields along $c(t)$. If $V \in \Gamma(T\mathcal{M}|_{c(t)})$ is induced by $\mathcal{X}$, meaning $V(t) = \mathcal{X}|_{c(t)}$, then

$$\frac{DV}{\mathrm{d}t}(t) = \nabla_{\dot{c}(t)}\mathcal{X} = \dot{\alpha}(t) + \Gamma(\alpha(t), \dot{\gamma}(t)), \quad \Gamma(u, v) = \begin{bmatrix} u^{\mathrm{T}}\Gamma^{1}v \\ \vdots \\ u^{\mathrm{T}}\Gamma^{n}v \end{bmatrix}.$$

$\Gamma(u, v)$ is called the *Christoffel function.* If $\dot{c}(t)$ is a geodesic, the expression above reduces to

i felt JZ was unclear. Is this correct?

refs

$$\ddot{\gamma}(t) = -\Gamma(\dot{\gamma}(t), \dot{\gamma}(t)), \tag{3.6}$$

since by definition geodesics must satisfy $\frac{D}{\mathrm{D}t}\dot{\gamma}(t) = 0$ and $\dot{\alpha}(t) = \dot{\gamma}(t)$. Importantly, once we have found the Christoffel symbols through (3.6), we can still use them for curves that are not geodesics. This is because the Christoffel symbols only depend on the Riemannian metric, and the local coordinates. To do this, we recover the Christoffel function for two different inputs through polarization [5, p. 312]

$$\Gamma(X, Y) = \frac{1}{4}\big(\Gamma(X + Y, X + Y) - \Gamma(X - Y, X - Y)\big),$$

where $X, Y \in \Gamma(T\mathcal{M}|_{c(t)})$.

i do not understand polarization.

To find the Christoffel symbols for $\mathrm{SpSt}(2n, 2k)$ with respect to the right invariant metric g defined in (3.3), we differentiate the geodesic formula from ref, and use (3.6) to achieve the following formula,

ref geod.

$$\Gamma(X, X) = -\ddot{\gamma}(0) = -(\overline{\Omega}(X) - \overline{\Omega}(X)^{\mathrm{T}})(X + \overline{\Omega}(X)^{\mathrm{T}}p) - (\overline{\Omega}(X)^{\mathrm{T}})^{2}p.$$

Here $X = \dot{\gamma}(0) \in T_p\mathrm{SpSt}(2n, 2k)$, $p \in \mathrm{SpSt}(2n, 2k)$, and $\overline{\Omega}(X)$ is as in ref. With our metric g, the Hessian at p of a smooth function $f\colon \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$ is the endomorphism

define this somewhere

$$\operatorname{Hess} f(p)\colon T_p\mathrm{SpSt}(2n, 2k) \to T_p\mathrm{SpSt}(2n, 2k),$$
$$\operatorname{Hess} f(p)[X] = \frac{\mathrm{d}}{\mathrm{d}t} \operatorname{grad} f(c(t))\Big|_{t=0} + \Gamma(\operatorname{grad} f(p), X),$$

where $\operatorname{grad} f(\cdot)$ is as in 3.4. $c(t) \in \mathrm{SpSt}(2n, 2k)$ is an arbitrary curve such that $c(0) = p$ and $c'(0) = X$

3.5 Moving from Theory to Application

Additional stuff to bridge the gap between theory and application. I.e. cayley transformation (p.20 in BZ, and prop. 5.2). Approx to the Riemannian hessian

4 Algorithms

I've implemented steepest descent and trustregion

4.1 Gradient descent

Armijo line search bou23, p.58

Algorithm 1 Riemannian Gradient descent

Input: Initial point $p_0 \in \mathrm{SpSt}(2n, 2k)$, objective function $f\colon \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, retraction $\mathcal{R}$, parameters $\beta, \gamma \in (0, 1)$, steplength range $0 < \gamma_{\min} < \gamma_{\max}$, initial step size $\gamma_0 = f(p_0)$, maximum number of iterations $N \in \mathbb{N}$, step parameters $h_{\min} < h_{\max} \in \mathbb{Z}$, tolerance parameters $\epsilon, \epsilon_x, \epsilon_f > 0$, Riemannian metric $\langle \cdot, \cdot \rangle_p$, with gradient grad_f where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.

```
for 0 ≤ k ≤ N do
    X_k = − grad f(X_k)
    γ_k = max(γ_min, (γ_k, γ_max))
    Armijo, t_k = ...
    p_{k+1} = R_{p_k}(t_k X_k)
    if ||grad f(p_k)||_F < ε then
        if |f(p_k) − f(p_{k+1})| / |f(p_k) + 1| < ε_f and ||p_k − p_{k+1}||_F / √(2n) < ε_x then
            Break
        end if
    end if
end for
```

Output: Iterates p_k

4.2 Trust region method

Algorithm 2 Riemannian Trust-region method

Input: Initial point $p_0 \in \mathrm{SpSt}(2n, 2k)$, objective function $f\colon \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, retraction $\mathcal{R}$, maximum number of iterations $N \in \mathbb{N}$, maximal radius $\overline{\Delta} > 0$, initial radius $\Delta_0 \in (0, \overline{\Delta})$, ratio threshold $\gamma_{\min} > 0$
tolerance parameters $\epsilon, \epsilon_x, \epsilon_f > 0$, Riemannian metric $\langle \cdot, \cdot \rangle_p$, with gradient grad_f where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.

for $0 \le k \le N$ **do**

$X_k = \mathrm{argmin}_{X \in T_{p_k}\mathrm{SpSt}(2n,2k)}\, m_k(X)$, subject to $||X||_{p_k} \le \Delta_k$ with m_k as in todo: ref

$\hat{p} = \mathcal{R}_{p_k}(X_k)$

$\gamma_k = \big(f(p_k) - f(\hat{p})\big) / \big(m_k(0) - m_k(X_k)\big)$

Compute new iterate

$$p_{k+1} = \begin{cases} \hat{p} & \text{if } \gamma_k > \gamma_{\min} \\ p_k & \text{otherwise} \end{cases}$$

Compute new trust-region radius

$$\Delta_{k+1} = \begin{cases} \frac{1}{4}\Delta_k & \text{if } \gamma_k < \frac{1}{4} \\ \min\left\{2\Delta_k, \overline{\Delta_k}\right\} & \text{if } \gamma > \frac{3}{4} \text{ and } ||X_k|| = \Delta_k \\ \Delta_k & \text{otherwise} \end{cases}$$

Todo: add tolerances

end for

Output: Iterates p_k

5 Numerical Experiments

6 Conclusion

Note: If I've only written ISBN, It's because I couldn't find the DOI.

Question: Is BZ not published?

Bibliography

[1] P-A Absil, R Mahony and R Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. ISBN: 9780691132983.

[2] Thomas Bendokat and Ralf Zimmermann. *The real symplectic Stiefel and Grassmann manifolds: metrics, geodesics and applications*. 2021. arXiv: arXiv:2108.12447v1.

[3] William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Vol. 63. Elsevier Science (USA), 1975. DOI: https://doi.org/10.1016/S0079-8169(08)61028-4.

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

[5] Alan Edelman, Tomás A. Arias and Steven T. Smith. 'The Geometry of Algorithms with Orthogonality Constraints'. In: *SIAM Journal on Matrix Analysis and Applications* 20 (1998). DOI: 10.1137/s0895479895290954. arXiv: arXiv:physics/9806030v1.

[6] Rasmus Jensen and Ralf Zimmermann. *Riemannian optimization on the symplectic Stiefel manifold using second-order information*. 2024. arXiv: arXiv:2404.08463v2.

[7] John M. Lee. *Introduction to Smooth Manifolds*. 2nd ed. Springer New York, NY, 2012. DOI: 10.1007/978-1-4419-9982-5.

[8] Department of Marine Technology NTNU. *IMT Software Wiki - LaTeX*. URL: https://www.ntnu.no/wiki/display/imtsoftware/LaTeX (visited on 15th Sept. 2020).

[9] Loring W. Tu. *Differential Geometry*. Springer International Publishing, 2017. DOI: 10.1007/978-3-319-55084-8.