Department of Mathematical Sciences

TMA4500 - Industrial Mathematics, Specialization Project

Optimization using second order information on the Symplectic Stiefel manifold

Author:
Ole Gunnar Røsholt Hovland

Date

Table of Contents

List of Figures

List of Tables

1 Introduction

2 Theory

2.1 Basic definitions

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

The optimization algorithm we will be studying is defined on a *Riemannian manifold*. This is because the algorithms we will use are designed to utilize first and second order information, and we need to define what these concepts mean on a manifold.

Definition 1 (Tangent Space)**.**

Definition 2 (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 3 (Quotient manifold)**.** *We define the definition of quotient manifold as in* **??***p. 27]Absil, Mahony, Sepulchre Let $\mathcal{M}$ be a manifold equipped with the operation $\sim$ called the* equivalence relation*. The equivalence relation has the following properties:* ref

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

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

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

Definition 4 (Horizontal & Vertical Space)**.** *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*

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

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 w sum! i wrote it in obsidian in the H and V space file!

Definition 5 (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 6 (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 7 (Riemannian connection)**.** *The* Riemanian 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 fiels on $\mathcal{M}$, it is defined as the unique $\mathbb{R}$-bilinear smooth map on $\mathcal{M}$ with riemannian metric $\langle\cdot,\cdot\rangle_p$*

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

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 8 (Christoffel symbols)**.** *The method we will employ to completely describe a connection (as defined in Definition 7) 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 $(U, p^1, \dots, p^n) \subseteq \mathcal{M}$ by $\partial_i := \partial/\partial p^i$. 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?

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

Definition 9.

Definition 10.

Definition 11.

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 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

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

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

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

In addition, define the *Symplectic inverse* of a matrix $p \in \mathbb{R}^{2n\times 2l}$ as The symplectic group is a quotient space of the general linear group, where it is defined as the set of matrices which define the symplectic structure in the following sense.We define the real symplectic group as

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

where p^+ is the symplectic inverse of p, as defined in [**symplectic˙inverse**].

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

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

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

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

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

Define geodesics as in BZ, prop. 2.1. Maybe do the proof?

2.3 The Symplectic Stiefel manifold

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

$$\mathrm{SpSt}(2n, 2k) := \{p \in \mathbb{R}^{2n\times 2k} : p^{\mathrm{T}} J_{2n} p = J_{2k}\}.$$

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 6 such that $\mathrm{St}(2n, 2k) = \mathrm{O}(n)/\mathrm{O}(n - k)$.

This is probably unnecessary

Derive tangent space. See obsidian

2.4 Right-invariant framework

Intro for this section. now that sp and spst are defined, define right inv. framework in sp, then guide the framework through to spst.

The goal for this section is to use the right-invariant metric defined on the symplectic group to define an appropriate metric on the symplectic Stiefel manifold.

many of the usual SpSt metrics do not have geodesics, therefore... We need a proper metric for $\mathrm{SpSt}(2n, 2k)$ in the sense that we need a metric that allows us to perform optimization on this manifold. To this end we need a metric that makes it possible to derive geodesics, as opposed to (BZ sources). The following metric defined in BZ fulfils our criteria.

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

$$g_p^{\mathrm{Sp}}(X_1, X_2) := \frac{1}{2}\mathrm{tr}((X_1 p^+)^T X_2 p^+), \quad X_1, X_2 \in T_p\mathrm{Sp}(2n, \mathbb{R}). \tag{2.6}$$

It is right-invariant in the sense that $g_{pq}^{\mathrm{Sp}}(X_1 q, X_2 q) = \frac{1}{2}\mathrm{tr}((X_1 q q^+ p^+)^T X_2 q q^+ p^+) = g_p^{\mathrm{Sp}}(X_1, X_2)$ for all $p \in \mathrm{Sp}(2n, \mathbb{R})$.

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 2.6. Split $T_p\mathrm{Sp}(2n, \mathbb{R})$ into to parts: the horizontal- and vertical part, with respect to g_p^{Sp} and π:

$$T_p\mathrm{Sp}(2n) = \mathrm{Ver}_p^{\pi}\,\mathrm{Sp}(2n) \oplus \mathrm{Hor}_p^{\pi}\,\mathrm{Sp}(2n). \tag{2.7}$$

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 $g_p : T_p\mathrm{SpSt}(2n, 2k) \times T_p\mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, $g_p(X_1, X_2) := g_p^{\mathrm{Sp}}((X_1)_p^{\mathrm{hor}}, (X_2)_p^{\mathrm{hor}})$. More explicitly

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

$$g_p(X_1, X_2) = \mathrm{tr}\left(X_1^T \left(I_{2n} - \frac{1}{2} J_{2n}^T p (p^T p)^{-1} p^T J_{2n} \right) X_2 (p^T p)^{-1} \right), \tag{2.8}$$

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

Define geodesics as in BZ

2.5 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 $f : \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, the Riemannian gradient with respect to g_p is given by*

$$\mathrm{grad}\, f(p) = \nabla f(p) p^T p + J_{2n} p (\nabla f(p))^T J_{2n} p, \tag{2.9}$$

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 ref ?? that $0 = p^+ \,\mathrm{grad}\, f(p) + (\mathrm{grad}\, f(p))^+ p$. 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 $g_p(\mathrm{grad}\, f(p), X) = \mathrm{d}\, f_p(X) = \mathrm{tr}((\nabla f(p))^{\mathrm{T}} X)$ for all $X \in T_p\mathrm{SpSt}(2n, 2k)$:

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

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

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

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

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

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, $\mathrm{tr}(A) = \mathrm{tr}(A^{\mathrm{T}})$, and for the second equality we utilize the cyclic property of the trace, and (2.1),

$$\begin{aligned}\tfrac{1}{2}\mathrm{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p(p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} JX\big) &= \tfrac{1}{2}\mathrm{tr}\big(X^{\mathrm{T}} J^{\mathrm{T}} p(p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J \nabla f(p)\big)\\ &= -\tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) X^{\mathrm{T}} J^{\mathrm{T}} p(p^{\mathrm{T}} p)^{-1}\big) \qquad (2.10)\end{aligned}$$

Inserting (2.10) into our expression we end up with:

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

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

add ref to Tspspt

2.6 Riemannian Hessian

(Christoffel symbols)

Define the remaining preliminaries for Hessian. Make an intro to section

loosly following Appendix A in Jz

For two smooth vector fields, $\mathcal{X}(p), \mathcal{Y}(p) \in \mathfrak{X}(\mathcal{M})$ defined as in **??** the covariant derivative (**??** written in local coordinates becomes

add definition of smooth v fields defined in local coords

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

In preparation for the Hessian, we include [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, $c : I \to \mathcal{M}$, $t \mapsto (\gamma_1(t), \ldots, \gamma_n(t))$, where $I := [a, b] \subseteq \mathbb{R}$. Since we have an affine connection on $\mathcal{M}$, the following unique map exists:

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

where $\Gamma(T\mathcal{M}|_{c(t)})$ denotes the the vectorspace 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 fealt JZ was unclear. Is this corredct?

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

since by definition $\frac{D}{\mathrm{D}t}\dot{\gamma}(t) = 0$ and $\dot{\alpha}(t) = \dot{\gamma}(t)$. Importantly, once we have found the Christoffel symbols throug **??**bove, we can still use them for curves that are not geodesics. This is because

refs

they only depend on the Riemannian metric, and the local coordinates. To do this, we recover the Christoffel function for two different inputs

A.4 in JZ

i do not undestand the last part above.

With our metric 2.8 the Hessian at p of a smooth function $f : \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$ is the endomorphism $\operatorname{Hess} f(p) : T_p\mathrm{SpSt}(2n, 2k) \to T_p\mathrm{SpSt}(2n, 2k)$. It is defined as

$$\operatorname{Hess} f(p)[X] = \frac{d}{dt} \operatorname{grad} f(c(t))\bigg|_{t=0} + \Gamma(\operatorname{grad} f(p), X),$$

where $\operatorname{grad} f(\cdot)$ is as in 2.9. Define an arbitrary curve $c(t) \in \mathrm{SpSt}(2n, 2k)$ such that $c(0) = p$ and $c'(0) = X$

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.