NTNU

kunnskap for en bedre verden

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 (Riemannian manifold)**.** *As defined in [1, def 2.6, p. 179]: a* smooth manifold $\mathcal{M}$*, defined as in [4, p. 13], is a* Riemannian manifold *if we can define a field of symmetric, positive definite, bilinear forms g, called the Riemannian metric. By field we mean that g_p is defined on the tangent space $T_p\mathcal{M}$ at each point $p \in \mathcal{M}$, as defined in [1, def 2.1, p. 178]. We will assume that g is smooth, meaning that it is od class $\mathcal{C}^\infty$.*

Definition 2 (Quotient space)**.**

Definition 3 (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* $\mathrm{GL}(n)$*. [2, Example 9.11]*

Definition 4 (Orthogonal group)**.** *The* real Orthogonal group *is defined as the set of all orthogonal matrices in $\mathbb{R}^{n\times n}$, denoted by* $\mathrm{O}(n)$*. [3, p. 3]*

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

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

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

2.3 The Symplectic Stiefel manifold

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 2, of the orthogonal group as defined in Definition 4 such that $\mathrm{St}(2n, 2k) = \mathrm{O}(n)/\mathrm{O}(n-k)$.

This is probably unnecessary

2.4 Right-invariant framework

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.

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. 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.

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

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

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

$$p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} J J p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p \tag{2.9}$$
$$= -p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p \tag{2.10}$$

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*

$$\operatorname{grad} f(p) = \nabla f(p) p^T p + J_{2n} p (\nabla f(p))^T J_{2n} p, \tag{2.11}$$

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 $0 = p^+ \operatorname{grad} f(p) + (\operatorname{grad} f(p))^+ p$. Computing this we get

$$p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} J J p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p = 0 \tag{2.12}$$

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 $G := J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J$. 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),

$$\begin{aligned} \tfrac{1}{2}\operatorname{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X\big) &= \tfrac{1}{2}\operatorname{tr}\big(X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J \nabla f(p)\big) \\ &= -\tfrac{1}{2}\operatorname{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1}\big) \end{aligned} \tag{2.13}$$

Inserting 2.13 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 ref ??, $p^{\mathrm{T}} J X = -X^{\mathrm{T}} J p$. □

add ref to Tspspt

2.6 Riemannian Hessian

(Christoffel symbols)

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

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

where $\operatorname{grad} f(\cdot)$ is as in 2.11. Define an arbitrary curve $c(t) \in \mathrm{SpSt}(2n, 2k)$ such that $c(0) = p$ and $c'(0) = X$

Bibliography

[1] 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.

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

[3] 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.

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

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