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.

2.1.1 Riemannian manifolds

As defined in [Boothby 1975] A smooth manifold (see [Lee 2012] smooth for definition) $\mathcal{M}$ 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 $p \in \mathcal{M}$.

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

2.2 The Symplectic group

We define the real symplectic group as

$$\mathrm{Sp}(2n) \coloneqq \{p \in \mathbb{R}^{2n\times 2n} : p^+ p = I_{2n}\}, \tag{2.1}$$

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

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

$$\mathfrak{sp}(2n) \coloneqq \{\Omega \in \mathbb{R}^{2n\times 2n} : \Omega^+ = -\Omega\}, \tag{2.2}$$

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

$$T_p\mathrm{Sp}(2n) = \{p\Omega, \Omega p \in \mathbb{R}^{2n\times 2n} : \Omega \in \mathfrak{sp}(2n)\}. \tag{2.3}$$

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

$$g_p(X_1, X_2) \coloneqq \frac{1}{2}\,\mathrm{tr}((X_1 M^+)^T X_2 M^+), \quad X_1, X_2 \in T_p\mathrm{Sp}(2n,\mathbb{R}). \tag{2.4}$$

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

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

$$T_p\mathrm{Sp}(2n, \mathbb{R}) = \mathrm{Ver}_p^\pi \oplus \mathrm{Hor}_p^\pi \mathrm{Sp}(2n, \mathbb{R}). \tag{2.5}$$

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

$$g_p(X_1, X_2) = \mathrm{tr}\left(X_1^T \left(I_{2n} - \frac{1}{2} J_{2n}^T p(p^T p)^{-1} p^T J_{2n}\right) X_2 (p^T p)^{-1}\right), \tag{2.6}$$

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

—

(Christoffel symbols)

$$\begin{aligned} &p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} J J p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p && (2.7)\\ &= -p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p && (2.8)\end{aligned}$$

2.5 Riemannian gradient

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 $0 = p^+ \,\mathrm{grad}\, f(p) + (\mathrm{grad}\, f(p))^+ p$ so

$$\begin{aligned} &p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} J J p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p \\ &= -p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p - p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p = 0 \end{aligned} \tag{2.10}$$

! Or should I mark where I did the simplification for clarity:

$$\begin{aligned} &p^{\mathrm{T}} J \nabla f(p) p^{\mathrm{T}} p + p^{\mathrm{T}} \overset{-I_{2n}}{\cancel{JJ}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} \overset{I_{2n}}{\cancel{J^{\mathrm{T}} J}} p \\ &= -p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p - p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} J p + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} J p = 0 \end{aligned} \tag{2.11}$$

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

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 $X \in T_p\mathrm{SpSt}(2n, 2k)$:

$$g_p(\mathrm{grad}\, f(p), X) = \mathrm{tr}\big((p^{\mathrm{T}} p (\nabla f(p))^{\mathrm{T}} + p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}})(I_{2n} - \tfrac{1}{2} G) X (p^{\mathrm{T}} p)^{-1}\big), \tag{2.12}$$

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

$$\begin{aligned} g_p(\mathrm{grad}\, f(p), X) = {}&\mathrm{tr}\big(\cancel{p^{\mathrm{T}}} p (\nabla f(p))^{\mathrm{T}} X \cancel{(p^{\mathrm{T}} p)^{-1}}\big) - \tfrac{1}{2}\mathrm{tr}\big(\cancel{p^{\mathrm{T}}} p (\nabla f(p))^{\mathrm{T}} G X \cancel{(p^{\mathrm{T}} p)^{-1}}\big) \\ &+ \mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big) - \tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} G X (p^{\mathrm{T}} p)^{-1}\big), \end{aligned} \tag{2.13}$$

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

$$
\begin{aligned}
g_p(\operatorname{grad} f(p), X) = \mathrm{d}\, f_p(X) &- \tfrac{1}{2}\mathrm{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X\big) \\
&+ \mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} J^{\mathrm{T}} X (p^{\mathrm{T}} p)^{-1}\big) \\
&- \tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) p^{\mathrm{T}} \cancelto{-I_{2n}}{J^{\mathrm{T}} J^{\mathrm{T}}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} \cancelto{-J^{\mathrm{T}}}{J} X (p^{\mathrm{T}} p)^{-1}\big).
\end{aligned}
\tag{2.14}
$$

We notice that after cancelling $p^{\mathrm{T}} p (p^{\mathrm{T}}) p^{-1}$ in the last term, the trace is equal to the second to last term. Now focusing on the second term: utilizing both the fact that for any matrix, A, (a) $\mathrm{tr}(A) = \mathrm{tr}(A^{\mathrm{T}})$, (b) the cyclic property of the trace, and (c) $J = -J^{\mathrm{T}}$, we get that

$$
\begin{aligned}
\tfrac{1}{2}\mathrm{tr}\big((\nabla f(p))^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J X\big) &\overset{\text{(a)}}{=} \tfrac{1}{2}\mathrm{tr}\big(X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1} p^{\mathrm{T}} J \nabla f(p)\big) \\
&\overset{(b),(c)}{=} -\tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) X^{\mathrm{T}} J^{\mathrm{T}} p (p^{\mathrm{T}} p)^{-1}\big)
\end{aligned}
\tag{2.15}
$$

Inserting this into our expression we end up with:

$$
\begin{aligned}
g_p(\operatorname{grad} f(p), X) = \mathrm{d}\, f_p(X) &+ \tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J \nabla f(p) \underbrace{X^{\mathrm{T}} J^{\mathrm{T}} p}_{=-X^{\mathrm{T}} J p} (p^{\mathrm{T}} p)^{-1}\big) \\
&+ \tfrac{1}{2}\mathrm{tr}\big(p^{\mathrm{T}} J^{\mathrm{T}} \nabla f(p) \underbrace{p^{\mathrm{T}} J^{\mathrm{T}} X}_{=-p^{\mathrm{T}} J X} (p^{\mathrm{T}} p)^{-1}\big) = \mathrm{d}\, f_p(X),
\end{aligned}
\tag{2.16}
$$

where the last two terms cancel by the tangent space condition ref ?? $p^{\mathrm{T}} J X = -X^{\mathrm{T}} J p$. □

2.6 Riemannian Hessian

(Christoffel symbols)

With our metric 2.6 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 $c'(0) = X$

Bibliography

'V Tensors and Tensor Fields on Manifolds' (1975). In: *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Ed. by William M. Boothby. Vol. 63. Pure and Applied Mathematics. Elsevier, pp. 174–225. DOI: https://doi.org/10.1016/S0079-8169(08)61028-4. URL: https://www.sciencedirect.com/science/article/pii/S0079816908610284.

Lee, John M. (2012). *Introduction to Smooth Manifolds*. 2nd ed. New York, NY: Springer. ISBN: 978-1-4419-9982-5.

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