TECHNISCHE UNIVERSITÄT CHEMNITZ

Technische Universität Chemnitz
Fakultät für Mathematik
Professur Numerische Mathematik
(Partielle Differentialgleichungen)

bachelor thesis

# The Efficiency of the Riemannian Symmetric Rank-One Method in Julia

submitted in fulfilment of the requirements for the degree of

Bachelor of Science



| | |
|---|---|
| submitted by | Tom-Christian Riemer |
| student number | 444019 |
| 1. referee | PD Dr. Ronny Bergmann |
| 2. referee | |
| | |
| date of submission | 31.01.2021 |

# Contents

# 1 Introduction

In the past decades, a notable interest has grown for the problem of minimizing a smooth objective function $f$ on a Riemannian manifold, which offers efficient alternative formulations to many problems. The applications are many and varied, they occur in engineering and science, which include the following fields: algorithmic questions pertaining to linear algebra, signal processing, data mining, statistical image analysis, financial mathematics, nanostructures, model reduction of dynamical systems and more. Optimization on Riemannian manifolds, also called Riemannian optimization, concerns finding an optimum of a real-valued function $f$ defined over a manifold. It can be thought of as unconstrained optimization on a constrained space. As such, optimization algorithms on manifolds are not fundamentally different from classical algorithms for unconstrained optimization in $\mathbb{R}^n$. On an Euclidean space, various methods of solving unconstrained optimization problems are known. The concepts of these algorithms can be used for the Riemannian optimization if many definitions are reconsidered. This reconsideration is crucial because the ideas are not extended simply from the Euclidean setup. The book Absil, Mahony, Sepulchre, 2007 provides a comprehensive introduction to this area, with an emphasis on providing the necessary background in differential geometry instrumental to algorithmic development.

Many manifold-based algorithms have been proposed or are under development. The reason for this is that they bring significant benefits, such as that all the iterates stay on the manifold, i.e., they satisfy the constraints (this property allows us to stop the iteration early), that they have the convergence properties of unconstrained optimization algorithms while operating on a constrained set, that there is no need to consider Lagrange multipliers or penalty functions and more Huang, 2013, p. 2-3. The idea of quasi-Newton methods on Riemannian manifolds is also not new. The first research paper to focus this topic was Gabay, 1982, but it was barely noticed. Nevertheless, a generalization of quasi-Newton methods in general and the BFGS method in particular is becoming more and more popular, since the many positive properties can be transferred to the Riemannian setting. In the Euclidean setting, the BFGS method is a well-known quasi-Newton method that has been viewed for many years as the best quasi-Newton method for solving unconstrained optimization problems, therefore much attention has been paid to generalizing this method to Riemannian manifolds.

This thesis is intended to deal with the BFGS method on Riemannian manifolds. We are interested in whether, and above all, how the BFGS method can be generalized for the application on Riemannian manifolds. We want to summarize the currently known Riemannian BFGS methods. Their core aspects should be discussed and their convergence results should be presented. Furthermore, we are interested in how a BFGS method on Riemannian manifolds can be implemented efficiently and which requirements have to be taken into account. An implementation of such a method should happen and its performance should be compared with results of other BFGS methods on Riemannian manifolds.

# 2 The Euclidean Symmetric Rank-One Quasi-Newton Method

In the Euclidean optimization a key problem is minimizing a real-valued function $f$ over the Euclidean space $\mathbb{R}^n$ ($n \geq 1$), i.e. our focus and efforts are centred on solving

$$\min f(x), \quad x \in \mathbb{R}^n \tag{2.0.1}$$

where $f\colon\ \mathbb{R}^n \to \mathbb{R}$ is a smooth function. In this chapter we focus on smooth functions, by which we generally mean functions whose second derivatives exist and are continuous or formally $f \in C^2(\mathbb{R}^n)$, unless otherwise stated. Equation (2.0.1) is called a (nonlinear) unconstrained optimization problem. In this work we consider numerical methods belonging to the class of quasi-Newton methods, which in turn belong to the class of line search methods. These can be formulated as algorithms where the next iterate is obtained by the iterative update scheme

$$x_{k+1} = x_k + \alpha_k d_k.$$

This means these methods start with an initial point $x_0 \in \mathbb{R}^n$ and produce a sequence of iterates $\{x_k\}_k$ that we hope will converge towards a minimum of Equation (2.0.1). The algorithms follow the strategy of first determining a search direction $d_k \in \mathbb{R}^n$ and then a suitable stepsize $\alpha_k > 0$ is searched for along this search direction $d_k$.
In quasi-Newton methods,

$$d_k = -{H_k}^{-1} \nabla f(x_k) = -B_k \nabla f(x_k)$$

is chosen as search direction, where the matrix $H_k \in \mathbb{R}^{n \times n}$ approximates the action of the objective's Hessian $\nabla^2 f(\cdot)$ in the direction of $s_k$ at the current iterate $x_k$ and $B_k = {H_k}^{-1}$, which means that $B_k$ approximates the action of $\nabla^2 f(x_k)^{-1}$ in the direction of $s_k$. These matrices are not calculated anew in each iteration, but $H_k$ or $B_k$ are updated to new matrices $H_{k+1}, B_{k+1} \in \mathbb{R}^{n \times n}$ using the information obtained during the iteration about the curvature of the objective function $f$. It is required that matrices generated by the update fulfil the so-called quasi-Newton equation, which reads as

$$H_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k) \quad \text{or} \quad B_{k+1}(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k.$$

For the sake of simplicity, we introduce the notations $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, thus we obtain

$$H_{k+1}s_k = y_k \quad \text{or} \quad B_{k+1}y_k = s_k. \tag{2.0.2}$$

The fulfillment of the quasi-Newton equation is the distinguishing feature of quasi-Newton methods. This means that quasi-Newton methods, like steepest descent, require only the gradient of the objective function to be supplied at each iterate.
The idea now is to find a convenient formula for updating the matrix H, which produces a matrix that satisfies the quasi-Newton equation and also carries other positive properties for the method. Instead of recomputing the approximate Hessian (or inverse Hessian) from scratch at every iteration, we apply a simple modification that combines the most recently observed information about the objective function with the existing knowledge embedded in our current Hessian approximation Nocedal, Wright, 2006, p. 139.
There are different formulas for updating the matrix, which of course differentiates the quasi-Newton methods. Probably the best-known method is based on the BFGS update, where the updated matrix Bk+1 (or Hk+1) differs from its predecessor Bk (or Hk ) by a rank-2 matrix:

$$H_{k+1}^{BFGS} = H_k^{BFGS} + \frac{y_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} - \frac{H_k^{BFGS} s_k s_k^{\mathrm{T}} H_k^{BFGS}}{s_k^{\mathrm{T}} H_k^{BFGS} s_k}$$

or

$$B_{k+1}^{BFGS} = \left(I_{n\times n} - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) B_k^{BFGS} \left(I_{n\times n} - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{2.0.3}$$

There is a simpler rank-1 update that maintains symmetry of the matrix and allows it to satisfy the secant equation. Unlike the rank-two update formulae, this symmetric-rank-1, or SR1, update does not guarantee that the updated matrix maintains positive definiteness. Good numerical results have been obtained with algorithms based on SR1, so we derive it here and investigate its properties.

The sequence of Hessian approximations generated by the SR1 method converges to the true Hessian under mild conditions, in theory; in practice, the approximate Hessians generated by the SR1 method show faster progress towards the true Hessian than do popular alternatives (BFGS or DFP), in preliminary numerical experiments.

# 3 The Riemannian Symmetric Rank-One Quasi-Newton Method

## 3.1 Preliminaries

## 3.2 Preliminaries

# Literature

Absil, P.-A.; R. Mahony; R. Sepulchre (2007). *Optimization Algorithms on Matrix Manifolds*. USA: Princeton University Press.

Gabay, D. (1982). "Minimizing a differentiable function over a differential manifold". *Journal of Optimization Theory and Applications* 37. DOI: `10.1007/BF00934767`.

Huang, W. (2013). "Optimization Algorithms on Riemannian Manifolds with Applications". URL: `http://purl.flvc.org/fsu/fd/FSU_migr_etd-8809`.

Nocedal, J.; S. J. Wright (2006). *Numerical Optimization*. Second Edition. Springer. DOI: `10.1007/978-0-387-40065-5`.