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 $s_k = x_{k+1} - x_k \in \mathbb{R}^n$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \in$

$\mathbb{R}^n$, 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:

$$H_{k+1}^{\mathrm{BFGS}} = H_k^{\mathrm{BFGS}} + \frac{y_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} - \frac{H_k^{\mathrm{BFGS}} s_k s_k^{\mathrm{T}} H_k^{\mathrm{BFGS}}}{s_k^{\mathrm{T}} H_k^{\mathrm{BFGS}} s_k} \tag{2.0.3}$$

or

$$B_{k+1}^{\mathrm{BFGS}} = \left(I_{n\times n} - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) B_k^{\mathrm{BFGS}} \left(I_{n\times n} - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{2.0.4}$$

There is a simpler rank-one update that maintains symmetry of the matrix and allows it to satisfy the secant equation. Unlike the rank-two update formulae, this symmetric-rank-one, 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 symmetric rank-one update has the general form

$$H_{k+1} = H_k + \sigma\, v v^{\mathrm{T}},$$

where $v \in \mathbb{R}^n$ and $\sigma \in \{-1, 1\}$. The task now is to determine v and σ so that H_{k+1} satisfies the quasi-Newton equation, Equation (2.0.2). By substituting into Equation (2.0.2), we obtain

$$H_k s_k + [\sigma\, v^{\mathrm{T}} s_k] v = y_k. \tag{2.0.5}$$

Since the term in brackets is a scalar, v must be a multiple of $y_k - H_k s_k$, i.e. $v = \delta(y_k - H_k s_k)$ for some $\delta \in \mathbb{R}$. By substituting this form of v into Equation (2.0.5), we obtain

$$(y_k - H_k s_k) = \sigma\delta^2 [s_k^{\mathrm{T}}(y_k - H_k s_k)](y_k - H_k s_k) \tag{2.0.6}$$

and it is clear that this equation is satisfied if (and only if) we choose the parameters δ and σ to be

$$\sigma = \operatorname{sgn}(s_k^{\mathrm{T}}(y_k - H_k s_k)), \quad \delta = \pm|s_k^{\mathrm{T}}(y_k - H_k s_k)|^{-\frac{1}{2}}.$$

Hence, the only symmetric rank-one updating formula that satisfies the secant equation is given by

$$H_{k+1}^{\mathrm{SR1}} = H_k^{\mathrm{SR1}} + \frac{(y_k - H_k^{\mathrm{SR1}} s_k)(y_k - H_k^{\mathrm{SR1}} s_k)^{\mathrm{T}}}{(y_k - H_k^{\mathrm{SR1}} s_k)^{\mathrm{T}} s_k} \tag{2.0.7}$$

By applying the Sherman–Morrison-Woodbury formula (cf.), we obtain the corresponding update formula for the approximation of inverse Hessian $\nabla^2 f(x_{k+1})^{-1}$:

$$B_{k+1}^{\mathrm{SR1}} = B_k^{\mathrm{SR1}} + \frac{(s_k - B_k^{\mathrm{SR1}} y_k)(s_k - B_k^{\mathrm{SR1}} y_k)^{\mathrm{T}}}{(s_k - B_k^{\mathrm{SR1}} y_k)^{\mathrm{T}} y_k}. \tag{2.0.8}$$

By the way, (5.1.21) is a general Broyden rank-one update in which, particularly, if v = yk, (5.1.21) is called the Broyden rank-one update presented by Broyden (1965) for solving systems of nonlinear equations. It is easy to see in that even if H_k^{SR1} is positive definite, H_{k+1}^{SR1} may not have the same property (this holds also for B_k^{SR1} and B_{k+1}^{SR1}). If and only if $(y_k - H_k^{\mathrm{SR1}} s_k)^{\mathrm{T}} s_k > 0$, SR1 update retains positive definiteness. However, this condition is difficult to guarantee. This means that H_{k+1}^{SR1} or B_{k+1}^{SR1} may no longer be invertible. Moreover, $d_{k+1} = -{H_{k+1}^{\mathrm{SR1}}}^{-1} \nabla f(x_{k+1}) = -B_{k+1}^{\mathrm{SR1}} \nabla f(x_{k+1})$ is not necessarily a descent direction.

The main drawback of the SR1 update formula is that the denominator in Equation (2.0.7) or Equation (2.0.8) can vanish. In fact, even when the objective function is convex and quadratic, there may be steps on which there is no symmetric rank-one update that satisfies the quasi-Newton equation Equation (2.0.2). This disadvantage results in serious numerical difficulties, which restrict the applications of the SR1 method.

By reasoning in terms of H_k^{SR1} (similar arguments can be applied to B_k^{SR1}), we see that there are three cases:

(i) If $(y_k - H_k^{\mathrm{SR1}} s_k)^{\mathrm{T}} s_k \neq 0$, then there is a unique rank-one updating formula satisfying Equation (2.0.2), and that it is given by Equation (2.0.7).

(ii) If $y_k = H_k^{\mathrm{SR1}} s_k$, then $H_{k+1}^{\mathrm{SR1}} = H_k^{\mathrm{SR1}}$ is the only updating formula satisfying Equation (2.0.2).

(iii) If $y_k \neq H_k^{\mathrm{SR1}} s_k$ and $(y_k - H_k^{\mathrm{SR1}} s_k)^{\mathrm{T}} s_k = 0$, then Equation (2.0.6) shows that there is no symmetric rank-one updating formula satisfying Equation (2.0.2).

The last case suggests that numerical instabilities and even breakdown of the method can occur, which means that a rank-one update does not provide enough freedom to develop a matrix with all the desired characteristics, and that a rank-two correction is required. This reasoning leads us back to the BFGS method, in which positive definiteness (and thus non-singularity) of all Hessian approximations is guaranteed if the so called curvature condition, which requires

$$s_k^{\mathrm{T}} y_k > 0, \tag{2.0.9}$$

is satisfied. Nevertheless, the SR1 formula has the following advantages:

(i) A simple safeguard seems to adequately prevent the breakdown of the method and the occurrence of numerical instabilities.

(ii) The matrices generated by the SR1 formula tend to be good approximations to the true Hessian matrix.

(iii) In quasi-Newton methods for constrained problems, it may not be possible to impose Equation (2.0.9), and thus the BFGS update, Equation (2.0.3), is not recommended.

The vanishing of the denominator in Equation (2.0.7) or Equation (2.0.8) is a topic, which deserves more research how to modify these updates such that they possess positive definiteness. For that we introduce a strategy to prevent the SR1 method from breaking down. It has been observed in practice that SR1 performs well simply by skipping the update if the denominator is small. More specifically, the update (6.24) is applied only if

$$|(s_k - B_k^{\mathrm{SR1}} y_k)^{\mathrm{T}} y_k| \geq \ r \ \|y_k\| \|s_k - B_k^{\mathrm{SR1}} y_k\| \tag{2.0.10}$$

holds, where $r \in (0,1)$ is a small number, e.g. $r = 10^{-8}$. Most implementations of the SR1 method use a skipping rule of this kind. The condition $(s_k - B_k^{\mathrm{SR1}} y_k)^{\mathrm{T}} y_k$ occurs infrequently, since it requires certain vectors to be aligned in a specific way. When it occurs, skipping the update appears to have no negative effects on the iteration, since the skipping condition implies that $y_k^{\mathrm{T}} \tilde{G}^{-1} y_k \approx y_k^{\mathrm{T}} B_k^{\mathrm{SR1}} y_k$, where $\tilde{G}$ is the average Hessian over the last step, meaning that the curvature of B_k^{SR1} along y_k is already correct.

In summary, a general quasi-Newton method using the inverse SR1 update, Equation (2.0.8) formula is as follows:

Algorithm 1 Inverse SR1 Method

Continuously differentiable real-valued function f on $\mathbb{R}^n$, bounded below; initial iterate $x_0 \in \mathbb{R}^n$; initial $s.p.d.$ matrix $B_0^{\text{SR1}} \in \mathbb{R}^{n \times n}$; convergence tolerance $\varepsilon > 0$. Set $k = 0$.
while $\|\nabla f(x_k)\| > \varepsilon$ **do**
    Compute the search direction $d_k = -B_k^{\text{SR1}} \nabla f(x_k)$.
    Determine a suitable stepsize $\alpha_k > 0$.
    Set $x_{k+1} = x_k + \alpha_k d_k$.
    Set $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
    Compute $B_{k+1}^{\text{SR1}} \in \mathbb{R}^{n \times n}$ by means of Equation (2.0.8).
    Set $k = k + 1$.
end while
Return x_k.

One of the main advantages of the SR1 method is its ability to generate good Hessian approximations. We first consider the case of a quadratic objective function. We have the following statement:

Theorem 2.0.1 (Nocedal, Wright, 2006, Theorem 6.1.). *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}$ is the strongly convex quadratic function $f(x) = \frac{1}{2}x^{\mathrm{T}}Ax + b^{\mathrm{T}}x$, where $A \in \mathbb{R}^{n \times n}$ is symmetric positive definite. Then for any starting point x_0 and any symmetric starting matrix $B_0^{\text{SR1}} \in \mathbb{R}^{n \times n}$, the iterates $\{x_k\}_k$ generated by Algorithm 1 with constant stepsize $\alpha_k = 1$ converge to the minimizer in at most n steps, provided that $(s_k - B_k^{\text{SR1}} y_k)^{\mathrm{T}} y_k \neq 0$ for all k. Moreover, if n steps are performed, and if the search directions d_k are linearly independent, then $B_n^{\text{SR1}} = A^{-1}$.*

The distinguishing feature of the SR1 quasi-Newton method with constant stepsize is its natural quadratic termination, which means that for a quadratic function, the method terminates within n steps.
In the proof it is shown that the SR1 update possesses the so-called hereditary property, i.e.

$$B_k^{\text{SR1}} y_i = s_i, \quad i = 0, \cdots, k-1. \tag{2.0.11}$$

This relation shows that when f is quadratic, the quasi-Newton equation, Equation (2.0.2), is satisfied along all previous search directions, regardless of how the line search is performed. A result like this can be established for the BFGS update, Equation (2.0.4), only under the restrictive assumption that an exact line search is used.
For general nonlinear functions, the SR1 update continues to generate good Hessian approximations under certain conditions:

Theorem 2.0.2 (Nocedal, Wright, 2006, Theorem 6.2.). *Suppose that f is twice continuously differentiable, and that its Hessian is bounded and Lipschitz continuous in a neighborhood of a point x^*. Let $\{x_k\}_k$ be any sequence of iterates such that $x_k \to x^*$ for some $x^* \in \mathbb{R}^n$. Suppose in addition that the inequality Equation (2.0.10) holds for all k, for some $r \in (0, 1)$, and that the steps s_k are uniformly linearly* *independent. Then the matrices H_k^{SR1} generated by Equation (2.0.7) satisfy*

$$\lim_{k\to\infty} \|H_k^{\text{SR1}} - \nabla^2 f(x^*)\| = 0.$$

“Uniformly linearly independent steps” means, that the steps do not tend to fall in a subspace of a dimension less than n. This assumption is usually, but not always, satisfied in practice.

3 The Riemannian Symmetric Rank-One Quasi-Newton Method

3.1 Preliminaries

The goal in this chapter is to solve Riemannian optimization problems, which consider finding an optimum of a real-valued function f defined on a Riemannian manifold, i.e.

$$\min f(x), \quad x \in \mathcal{M} \tag{3.1.1}$$

where $\mathcal{M}$ is a Riemannian manifold. From now on we assume that $\mathcal{M}$ is a n-dimensional geodesically complete Riemannian manifold. We further assume that the manifold $\mathcal{M}$ is embedded in a real-valued space (e.g. $\mathcal{M} \subseteq \mathbb{R}^m$) and connected. Further we assume that $f\colon\ \mathcal{M} \to \mathbb{R}$ is a twice continuously differentiable function, i.e. $f \in C^2(\mathcal{M})$.

$$x_{k+1} = \exp_{x_k}(\alpha_k \eta_k), \tag{3.1.2}$$

$$f(\exp_{x_k}(\alpha_k \eta_k)) \leq f(x_k) + c_1 \alpha_k g_{x_k}(\operatorname{grad} f(x_k), \eta_k) \tag{3.1.3}$$

$$g_{\exp_{x_k}(\alpha_k \eta_k)}(\operatorname{grad} f(\exp_{x_k}(\alpha_k \eta_k)), \mathrm{P}_{x_k, \alpha_k \eta_k}(\eta_k)) \geq c_2 g_{x_k}(\operatorname{grad} f(x_k), \eta_k). \tag{3.1.4}$$

with $0 < c_1 < c_2 < 1$.

$$\widetilde{\mathcal{H}}_k = \mathrm{P}_{x_{k+1} \leftarrow x_k} \circ \mathcal{H}_k \circ \mathrm{P}_{x_k \leftarrow x_{k+1}} \colon\ \mathcal{T}_{x_{k+1}}\mathcal{M} \to \mathcal{T}_{x_{k+1}}\mathcal{M}, \tag{3.1.5}$$

$$\mathcal{H}^{RBFGS}_{k+1}[\cdot] = \widetilde{\mathcal{H}}^{RBFGS}_k[\cdot] + \frac{y_k y^\flat_k[\cdot]}{s^\flat_k[y_k]} - \frac{\widetilde{\mathcal{H}}^{RBFGS}_k[s_k] s^\flat_k(\widetilde{\mathcal{H}}^{RBFGS}_k[\cdot])}{s^\flat_k(\widetilde{\mathcal{H}}^{RBFGS}_k[s_k])}. \tag{3.1.6}$$

$$\mathcal{B}^{RBFGS}_{k+1}[\cdot] = \Big(\mathrm{id}_{\mathcal{T}_{x_{k+1}}\mathcal{M}}[\cdot] - \frac{s_k y^\flat_k[\cdot]}{s^\flat_k[y_k]}\Big)\widetilde{\mathcal{B}}^{RBFGS}_k[\cdot]\Big(\mathrm{id}_{\mathcal{T}_{x_{k+1}}\mathcal{M}}[\cdot] - \frac{y_k s^\flat_k[\cdot]}{s^\flat_k[y_k]}\Big) + \frac{s_k s^\flat_k[\cdot]}{s^\flat_k[y_k]}. \tag{3.1.7}$$

3.2 Preliminaries

4 Numerics

the Rayleigh quotient minimization problem on the sphere $\mathbb{S}^{n-1}$. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the unit-norm eigenvector, $v \in \mathbb{R}^n$, corresponding to the smallest eigenvalue, defines the two global minima, $\pm v$, of the Rayleigh quotient

$$\begin{aligned} f\colon\ \mathbb{S}^{n-1} &\to \mathbb{R} \\ x &\mapsto x^{\mathrm{T}} A x \end{aligned} \tag{4.0.1}$$

with its gradient

$$\operatorname{grad} f(x) = 2(Ax - xx^{\mathrm{T}} A x).$$

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.