The nonsmooth Newton method on Riemannian manifolds

C. Lageman, **U. Helmke**†, **J.H. Manton**‡*

1 Introduction

Solving nonlinear equations in Euclidean space is a frequently occurring problem in optimization and control theory, with many applications to e.g. tracking and filtering, optimal control, observer theory and robust control. A classical solution approach uses the Newton algorithm which converges locally quadratically fast to a nondegenerate zero of the vector field F.

In the last two decades there has been a considerable interest in intrinsic optimization algorithms on Riemannian manifolds, including a Riemannian variant of the Newton iteration, see [5, 10, 11, 7, 1]. Such algorithms use the intrinsic Riemannian structure of the constraint set, such as geodesics and covariant derivatives. Thus the Riemannian Newton iteration for a smooth vector field F is given by

$$x_{k+1} = \exp_{x_k}(-(\nabla F(x_k))^{-1}F(x_k)),$$

where F is a smooth vector field, exp the exponential map and ∇ denotes the covariant derivative.

In Euclidean space, it is possible to extend the Newton algorithm to non-differentiable vector fields [9, 8]. Using tools from nonsmooth analysis, one can replace the standard differential by a generalized differential, e.g. Clarke's generalized

*Centre for Mathematics and its Applications, Australian National University, Canberra, Australia. Work supported by the Australian Research Council Centre of Excellence for Mathematics and its Applications.

†Department of Mathematics, University Würzburg, Würzburg, Germany. Partially supported by BMBF-Project "Development and Implementation of novel mathematical methods for identification and control of technical systems".

‡Department of Information Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia.

Jacobian. Under suitable regularity conditions of F, local superlinear convergence of such algorithms can be established.

Motivated by recent work of [6, 2], and applications in e.g. distributed control and optimization [4], we present here a extension of the nonsmooth Euclidean Newton method of Qui and Sun [9] to Riemannian manifolds. The algorithm is locally superlinearily convergent. A numerical example illustrates the feasibility of the method.

2 Nonsmooth covariant derivatives on Riemannian manifolds

Let M denote a smooth, complete Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle$ and tangent bundle TM. The vector space of linear endomorphisms on a tangent space T_xM, $x \in M$, is denoted by $L(T_xM)$. We denote by $L(TM)$ the bundle of vector spaces $L(T_xM)$ over M. Let ∇ denote the Levi-Civita connection on M and $\exp_x$the exponential map at $x \in M$. We denote by τ_{γ,t_1,t_2} the parallel transport along a smooth curve $\gamma \colon \mathbb{R} \to M$ from $\gamma(t_1)$ to $\gamma(t_2)$. The Riemannian distance on M is denoted by dist and a ball around $x \in M$ of radius r by $B_r(x)$. $\mathcal{P}(S)$ denotes the set of subsets of S.
A vector field X on M is called **Lipschitz continuous**, if for any $x \in M$ there is a constant $C > 0$ and a neighborhood $U \subset T_xM$ of 0 such that $\exp_x : T_xM \to M$ is bijective on U and for all $v \in U$

$$\|\tau_{\exp_x(tv),1,0} X(\exp_x(v)) - X(x)\| \leq C\|v\|.$$

Note that our definition of Lipschitz continuity implies that the vector field is Lipschitz continuous in Riemannian normal coordinates, and even in arbitrary local charts. It follows from Rademacher's theorem [3], that the set of points $\Omega(X)$, where X is not differentiable, has measure zero. For points $x \in M \setminus \Omega(X)$ and $v \in T_xM$ the **covariant derivative** $\nabla_v X(x)$ can be defined by

$$\lim_{t \to 0} \frac{\tau_{\gamma,t,0}(X(\gamma(t))) - X(x)}{t}$$

for a smooth curve $\gamma \colon \mathbb{R} \to M$ with $\gamma(0) = x$, $\gamma'(0) = v$. For any $x \in M \setminus \Omega(X)$, we therefore obtain a well defined map $\nabla X(x) : T_xM \to T_xM$, given by $v \mapsto \nabla_v X(x)$. In order to extend this definition to arbitrary points $x \in M$, we introduce a set-valued generalization of the covariant derivative for nonsmooth vector fields.

Definition 1. *Let X be a Lipschitz continuous vector field on M. The* **generalized covariant derivative** *of X at $x \in M$ is*

$$\nabla X(x) = \operatorname{co}\{A \in L(T_xM) \mid \exists (x_k) \subset M \setminus \Omega(X), x_k \to x, A = \lim_{k \to \infty} \nabla X(x_k)\}, \quad (1)$$

where co *denotes the convex hull of a subset of $L(T_xM)$. We denote by $\nabla X \colon M \to \mathcal{P}(L(TM))$ the set-valued map $x \mapsto \nabla X(x)$.*

Note that, for any $x \in M \setminus \Omega(X)$, the generalized covariant derivative is single valued and thus coincides with the above definition. In particularly, it shares many properties with Clarke's earlier concept of a generalized Jacobian. Similar to the nonsmooth Newton algorithm in Euclidean case, cf. [9], we have to impose a regularity conditon to ensure that the Newton iteration is well-defined on a neighborhood of a solution of $X(x) = 0$.

Definition 2. *Let X be a Lipschitz continuous vector field on M. We call X regular at a point x, if all linear maps $A \in \nabla X(x)$ are bijective.*

It is easily seen, that regularity of a Lipschitz continuous vector field X in a point $x \in M$ implies regularity on a neighborhood of x. Moreover, this neighborhood can be chosen such that there is a uniform bound on the operator norm $\|A^{-1}\|$ for all elements A of ∇X.

However, even in Euclidean space, the Lipschitz continuity of a vector field is not sufficient to guarantee convergence of the nonsmooth Newton algorithms at a regular point. Thus an additional regularity assumption on X is needed. This is given by the concept of semismoothness, which relates directional derivatives to the generalized covariant derivative; see [9] for a Euclidean space version of this definition.

Definition 3. *A Lipschitz continuous vector field X on M is called* **semismooth** *at $x \in M$, if for all $w \in T_xM$ and arbitrary sequences $t_k \in \mathbb{R}^+$, $t_k \to 0$, $w_k \in T_xM$, $w_k \to w$, and $A_k \in \nabla X(\exp_x(t_k w_k))$ the limit*

$$\lim_{k\to\infty} A_k \left(\tau_{\exp_x(t_k w_k),0,1}(w_k)\right) \tag{2}$$

exists.

3 The nonsmooth Newton iteration

We can now extend the nonsmooth Newton iteration from Euclidean space, see [9], to Riemannian manifolds. The basic idea behind this construction is simple enough: the generalized Jacobian is replaced by our generalized covariant derivative and parallel translation along a line is replaced by that along a geodesic.

Definition 4. *Let X be a semismooth vector field on M. Any sequence (x_k) in M satisfying*

$$x_{k+1} = \exp_{\mathrm{x_k}}(-\mathrm{A_k^{-1}X(x_k)}) \text{ with } \mathrm{A_k} \in \nabla\mathrm{X(x_k)}$$

is called a **nonsmooth Newton iteration***.*

Under the assumptions of semismoothness and regularity we get local superlinear convergence of this Newton methods as in the Euclidean case.

Theorem 5. *Let X be a semismooth vector field on M. Assume that x^* is a zero of X and X is regular at x^*. Then there is a neighborhood U of x^* such that any* *nonsmooth Newton iteration* (x_k) *starting in* U *is well-defined for all* $k \in \mathbb{N}$ *and converges locally superlinearily fast to* x^**:*

$$\lim_{k\to\infty} \frac{\operatorname{dist}(x_{k+1}, x^*)}{\operatorname{dist}(x_k, x^*)} = 0. \tag{3}$$

Proof. The proof is by an extension of the arguments in the Euclidean case. One can show that the semi-smoothness implies that there is a bounded neighborhood V of x^* and a continuous, nonnegative function $g\colon \mathbb{R} \to \mathbb{R}$ with $\lim_{t\to 0} g(t)/t = 0$ such that for all $x \in V$, $v \in T_xM$

$$\left\|X(y) - \tau_{\exp_x(tv),0,1}(X(x)) + A\tau_{\exp_x(tv),0,1}(v)\right\| \leq g(\operatorname{dist}(x,y)). \tag{4}$$

Choosing a suitably small neighborhood $W \subset V$ of x^*, the inequality (4) yields for all $x \in W$

$$\|A^{-1}X(x) + \exp_x^{-1}(x^*)\| \leq Cg(\operatorname{dist}(x, x^*)),$$

where $C > 0$ is a constant. Since V is bounded and $\exp_x(A^{-1}X(x)), x^* \in V$, there is a constant $K > 0$ such that

$$\operatorname{dist}(\exp_x A^{-1}X(x), x^*) \leq K\|A^{-1}X(x) + \exp_x^{-1}(x^*)\| \leq CKg(\operatorname{dist}(x, x^*)).$$

Using the boundedness assumptions and assuming the W is suitably small, there exists $K > 0$ such that for any $x_k \in W$

$$\operatorname{dist}(x_{k+1}, x^*) < CKg(\operatorname{dist}(x_k, x^*)). \tag{5}$$

The superlinear convergence now follows directly from the asymptotic properties of g for $t \to 0$. □

The Newton-Kantorovich convergence result for the nonsmooth Newton iteration on Euclidean space, [9], can also be extended to the Riemannian case. We omit the proof here.

Theorem 6. *Let* X *be a semismooth, Lipschitz continuous vector field on* M*. Assume that we are given a ball* $B_r(x_0)$ *such that the following conditions hold*

- *for all* $x \in B_r(x_0), A \in \nabla X(x)$ *satisfies* $\|A^{-1}\| < C_1$*,*
- *for all* $x \in B_r(x_0)$*,* $v \in T_xM$*, such that* $\exp_x(tv) \in B_r(x_0)$ *for all* $t \in [0,1]$*,*

$$\left\|Av - \lim_{t\to 0} \frac{\tau_{1,0,\exp_x(tv)}(X(\exp_x(tv))) - X(x)}{t}\right\| \leq C_2 \operatorname{dist}(x, \exp_x(tv))$$

- *for all* $x \in B_r(x_0)$*,* $v \in T_xM$*, such that* $\exp_x(tv) \in B_r(x_0)$ *for all* $t \in [0,1]$*,*

$$\left\|\tau_{1,0,\exp_x(tv)}(X(\exp_x(tv))) - X(x) - \lim_{t\to 0} \frac{\tau_{1,0,\exp_x(tv)}(X(\exp(tv)) - X(x)}{t}\right\| \\ \leq C_3 \operatorname{dist}(x, \exp(tv))$$

- $C_1(C_2 + C_3) < 1$ *and* $C_2\|X(x_0)\| < r$.

Then the nonsmooth Newton iteration starting in x_0 *converges to the unique solution* x^* *of* $X(x) = 0$ *in* $B_r(x_0)$.

4 Examples

We illustrate the nonsmooth Newton algorithm by simple examples. Consider any Lipschitz continuous vector field Y on $\mathbb{R}^n$ and let S^{n-1} denote the set of Euclidean norm unit vectors in $\mathbb{R}^n$. We endow S^{n-1} with the Riemannian metric induced by the Euclidean one. The orthogonal projection at $x \in S^{n-1}$ onto the tangent space T_xS^{n-1} is the linear map $\pi(x)\colon \mathbb{R}^n \to T_xS^{n-1}$, defined by $\pi(x) = I_n - xx^\top$. Thus the projected vector field

$$X(x) = (I_n - xx^\top)Y(x)$$

defines a vector field on S^{n-1}. It is easy to see that the generalized covariant derivative of X on S^{n-1} then is given by

$$\nabla_v X(x) = (I_n - xx^\top)DY(x)(v) - x^\top Y(x)v,$$

where $DY(x)$ denotes Clarke's generalized derivative. We consider now two special situations.

4.1 Piecewise linear vector fields

Assume that we are given polytopes P_j in $\mathbb{R}^n$, $j = 1, \ldots, k$, of the form $C_jx \geq 0$ where C_j is a $m_j \times n$ matrix, $m_j \in \mathbb{N}$. Note that these polytopes are unbounded. We further assume that if $P_i \cap P_j \neq \{0\}$ then $P_i \cap P_j$ is a lower dimensional polytope P_l. Let Y be a continuous, piecewise linear vector field such that $Y(x) = B_jx$ holds for $x \in P_j$. Note that this implies, for any subpolytope $P_l = P_i \cap P_j$, the condition $B_lx = B_ix = B_jx$ for all $x \in P_l$. The Clarke generalized Jacobian of Y is readily calculated as

$$DY(x)(v) = \operatorname{co}\{B_jv \mid x \in P_j\}.$$

Consider the restriction $X(x) = \pi(x)Y(x)$ of Y to S^{n-1} This yields the generalized covariant derivative

$$\nabla_v X(x) = \operatorname{co}\left\{(I_n - xx^\top)B_jv - x^\top B_jxv \mid x \in P_j, \dim P_j = n\right\}.$$

Based on this information, we can immediately implement the nonsmooth Newton iteration.

4.2 Component-wise minimum vector fields

Let us consider a vector field

$$Y(x) = \min\{h_1(x), \ldots, h_k(x)\}$$

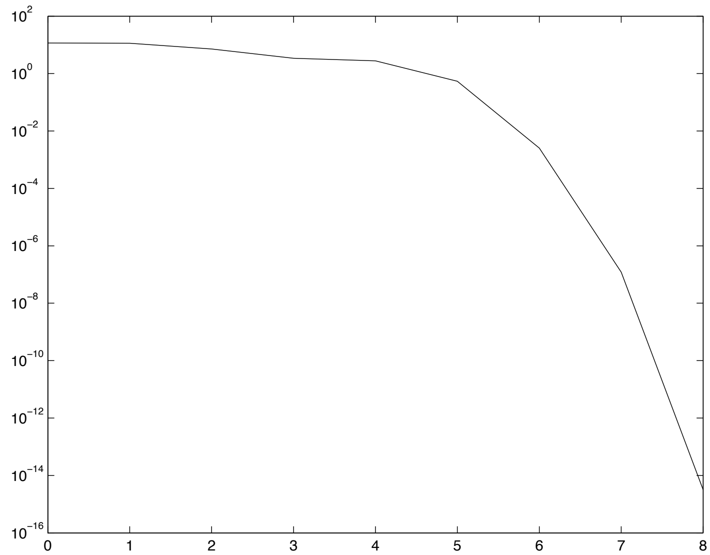

Figure 1. *Evolution of* $\|X\|$ *for* $n = 4$ *and* $k = 5$.

with $h_1(x), \dots, h_k(x) \in \mathbb{R}^n$ homogeneous polynomials and the minimum taken component wise. We denote by $h_{jl}(x)$ the lth row of h_j. Standard rules for the Clarke generalized gradient of a minimum function, cf. [3], imply that

$$DY(x)(v) = \begin{pmatrix} \mathrm{co}\{dh_{j1}(x)(v) \mid h_{j1}(x) = \max\{h_{11}(x), \dots, h_{k1}(x)\}, j \in \{1, \dots, k\}\} \\ \vdots \\ \mathrm{co}\{dh_{jn}(x)(v) \mid h_{jn}(x) = \max\{h_{1n}(x), \dots, h_{kn}(x)\}, j \in \{1, \dots, k\}\} \end{pmatrix}.$$

We can now calculate $\nabla_v X$ for this Y by the formulas given above. Furthermore, it is easy to see that X is semismooth. Using $\nabla_v X$ and the well-known formula for geodesics on the sphere, we can implement our nonsmooth Newton algorithm. For the special case that all polynomials have degree 1, i.e. $h_j(x) = B_j x$ and $b_{j1}, \dots, b_{jn}$ denoting the rows of B_j, we have to following algorithm.

1. Let $B_1, \dots, B_k \in \mathbb{R}^{n\times n}$ and $x_0 \in S^{n-1}$.

2. Choose column vectors

$$G_i \in \begin{pmatrix} \mathrm{co}\{b_{j1} \mid b_{j1}x_i = \max\{b_{11}x_i, \dots, b_{k1}x_i\}, j \in \{1, \dots, k\}\} \\ \vdots \\ \mathrm{co}\{b_{jn} \mid b_{jn}x_i = \max\{b_{1n}x_i, \dots, b_{kn}x_i\}, j \in \{1, \dots, k\}\} \end{pmatrix}.$$

3. Determine $v_i \in T_{x_i}S^{n-1}$ from the equation

$$((I_n - x_i x_i^\top)G_i - x_i^T Y(x_i))v_i = -X(x_i)$$

4. Set
$$x_{i+1} = \cos(\|v_i\|)x_i + \frac{\sin(\|v_i\|)}{\|v_i\|}v_i$$
and go to step 2.

In Figure 1 we show the results of this algorithm for 10 randomly chosen 10×10 matrices. The algorithm is terminated when $\|X\|$ is below 10^{-13}. The diagram shows the evolution of $\|X\|$. Thus, in the example the algorithm converges locally superlinearily as expected. However, since Newton algorithms are only locally convergent, the region of attraction can be reached arbitrarily late. In our simulations, we experienced that superlinear convergence occurs only in a region where Y coincides with a fixed $\tilde{B}x$, with $\tilde{B}$ consisting of rows of different B_i. This can be explained by realizing that generically the zeros of Y lie in the interior of such regions. It it also observed that the size of the regions of attraction increases with increasing k and decreases with increasing n.

Bibliography

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, New Jersey, 2008.

[2] D. Azagra, J. Ferrera, and F. López-Mesas, *Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds*, J. Funct. Anal., 220 (2005), pp. 304–361.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, Inc., 1983.

[4] J. Cortes and F. Bullo, *Coordination and geometric optimization via distributed dynamical systems*, SIAM J. Control and Optimization, 44 (2005), pp. 1543–1574.

[5] U. Helmke and J. Moore, *Optimization and Dynamical Systems*, Springer, London, 1994.

[6] Y. Ledyaev and Q. J. Zhu, *Nonsmooth analysis on manifolds*, Trans. AMS, 359 (2007), pp. 3687–3732.

[7] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Proc., 50 (2002), pp. 635–650.

[8] J.-S. Pang and L. Qi, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optimization, 3 (1993), pp. 443–465.

[9] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353–367.

[10] M. Shub, *Some remarks on dynamical systems and numerical analysis*, in Dynamical Systems and Partial Differential Equations: Proceedings of the VII ELAM, L. Lara-Carrero and J. Lewowicz, eds., Caracas, 1986, Equinoccio, Universidad Simon Bolivar, pp. 69–92.

[11] S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, PhD thesis, Havard University, Cambridge, Massachusetts, 1993.