Applied Mathematics Letters 78 (2018) 42–50

Contents lists available at ScienceDirect

Applied Mathematics Letters

www.elsevier.com/locate/aml

Convergence analysis of Riemannian Gauss–Newton methods and its connection with the geometric condition number

Paul Breiding[a], Nick Vannieuwenhoven[b,*]

[a] Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany
[b] KU Leuven, Department of Computer Science, Leuven, Belgium

ARTICLE INFO

Article history:
Received 7 August 2017
Received in revised form 17 October 2017
Accepted 17 October 2017
Available online 24 November 2017

Keywords:
Riemannian Gauss–Newton method
Convergence analysis
Geometric condition number
CPD

ABSTRACT

We obtain estimates of the multiplicative constants appearing in local convergence results of the Riemannian Gauss–Newton method for least squares problems on manifolds and relate them to the geometric condition number of Bürgisser and Cucker (2013).

© 2017 Elsevier Ltd. All rights reserved.

Introduction

Many problems in science and engineering are *parameter identification problems* (PIPs). Herein, there is a parameter domain $\mathcal{M} \subset \mathbb{R}^M$ and a function $\Phi : \mathcal{M} \to \mathbb{R}^N$. Given a point $\mathbf{y}$ in the image of Φ, the PIP asks to identify parameters $\mathbf{x} \in \mathcal{M}$ such that $\mathbf{y} = \Phi(\mathbf{x})$; note that there could be several such parameters. For example, computing QR, LU, Cholesky, polar, singular value and eigendecompositions of a given matrix $A \in \mathbb{R}^{m\times n} \simeq \mathbb{R}^{mn}$ are examples of this. In other cases we have a tensor $\mathfrak{A} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ and need to compute CP, Tucker, block term, hierarchical Tucker, or tensor trains decompositions [1].

If the object $\widetilde{\mathbf{y}} \in \mathbb{R}^N$ whose parameters should be identified originates from applications, then usually $\widetilde{\mathbf{y}} \notin \Phi(\mathcal{M})$. Nevertheless, in this setting one seeks parameters $\mathbf{x} \in \mathcal{M}$ such that $\mathbf{y} := \Phi(\mathbf{x})$ is as close as possible to $\widetilde{\mathbf{y}}$, e.g., in the Euclidean norm. This can be formulated as a nonlinear least squares problem:

$$\widetilde{\mathbf{y}} \mapsto \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{M}} \tfrac{1}{2}\|\Phi(\mathbf{x}) - \widetilde{\mathbf{y}}\|^2. \tag{1}$$

Here, we deal with functions Φ that offer differentiability guarantees, so that continuous optimization methods can be employed for solving (1). Specifically, we assume that $\mathcal{M}$ is a smooth embedded submanifold[1]

* Corresponding author.
E-mail addresses: breiding@mis.mpg.de (P. Breiding), nick.vannieuwenhoven@cs.kuleuven.be (N. Vannieuwenhoven).

https://doi.org/10.1016/j.aml.2017.10.009
0893-9659/© 2017 Elsevier Ltd. All rights reserved.

of $\mathbb{R}^M$ and that Φ is a smooth function on $\mathcal{M}$ [3, Chapters 1 and 2]. Hence, (1) is a *Riemannian optimization problem* that can be solved using, e.g., Riemannian Gauss–Newton (RGN) methods [4]; see the next section.

The sensitivity of $\mathbf{x} \in \mathcal{M}$ with respect to perturbations of $\mathbf{y} = \Phi(\mathbf{x})$ might impact the performance of these RGN methods. Let $\Psi : \mathcal{X} \to \mathcal{Y}$ be a smooth map between manifolds $\mathcal{X}$ and $\mathcal{Y}$, and let $\mathrm{T}_{\mathbf{x}}\mathcal{X}$ denote the tangent space to the manifold $\mathcal{X}$ at $\mathbf{x} \in \mathcal{X}$. We recall from [5, Section 14.3] that the geometric condition number $\kappa(\mathbf{x})$ characterizes to first-order the sensitivity of the output $\mathbf{y} = \Psi(\mathbf{x})$ to input perturbations as the spectral norm of the derivative operator $\mathrm{d}_{\mathbf{x}}\Psi : \mathrm{T}_{\mathbf{x}}\mathcal{X} \to \mathrm{T}_{\Psi(\mathbf{x})}\mathcal{Y}$; that is, $\kappa(\mathbf{x}) := \|\mathrm{d}_{\mathbf{x}}\Psi\| := \max_{\mathbf{t}\in\mathrm{T}_{\mathbf{x}}\mathcal{X}} \|\mathrm{d}_{\mathbf{x}}\Psi(\mathbf{t}/\|\mathbf{t}\|)\|$. In the case of PIPs, the geometric condition number is derived as follows. Assume that there exists an open neighborhood $\mathcal{N}$ of $\mathbf{x} \in \mathcal{M}$ such that $\mathcal{M} = \Phi(\mathcal{N})$ is a smooth manifold with $m = \dim \mathcal{M} = \dim \mathcal{N}$. Since $\Phi|_{\mathcal{N}} : \mathcal{N} \to \mathcal{M}$ is a smooth map between manifolds, the inverse function theorem for manifolds [3, Theorem 4.5] entails that there exists a unique inverse function $\Phi_{\mathbf{x}}^{-1}$ whose derivative satisfies $\mathrm{d}_{\Phi(\mathbf{x})}\Phi_{\mathbf{x}}^{-1} = (\mathrm{d}_{\mathbf{x}}\Phi)^{-1}$, provided that $\mathrm{d}_{\mathbf{x}}\Phi$ is injective. Hence, the geometric condition number of the *parameters*[2] $\mathbf{x}$ is

$$\kappa(\mathbf{x}) := \|\mathrm{d}_{\Phi(\mathbf{x})}\Phi_{\mathbf{x}}^{-1}\| = \|(\mathrm{d}_{\mathbf{x}}\Phi)^{-1}\| = \frac{1}{\varsigma_m(\mathrm{d}_{\mathbf{x}}\Phi)}, \tag{2}$$

where $\varsigma_m(A)$ is the mth largest singular value of the linear operator A. If the derivative is not injective, then the condition number is defined to be ∞.

In this letter, we show that the condition number of the parameters $\mathbf{x}$ in (2) appears naturally in the multiplicative constants in convergence estimates of RGN methods. Our main contribution is Theorem 1.

The Riemannian Gauss–Newton method

Recall that a Riemannian manifold $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ is a smooth manifold $\mathcal{M}$, where for each $p \in \mathcal{M}$ the tangent space $\mathrm{T}_p\mathcal{M}$ is equipped with an inner product $\langle\cdot,\cdot\rangle_p$ that varies smoothly with p; see [3, Chapter 13]. The zero element of $\mathrm{T}_p\mathcal{M}$ is denoted by 0_p. Since we deal exclusively with embedded submanifolds $\mathcal{M} \subset \mathbb{R}^M$, we take $\langle\mathbf{a},\mathbf{b}\rangle_p := \mathbf{a}^T\mathbf{b}$ equal to the standard inner product on $\mathbb{R}^M$. In the following we drop the subscript "p". The induced norm is $\|\mathbf{v}\| = \sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$. The *tangent bundle* of a manifold $\mathcal{M}$ is the smooth vector bundle $\mathcal{TM} := \{(p,\mathbf{v}) \mid p \in \mathcal{M}, \mathbf{v} \in \mathrm{T}_p\mathcal{M}\}$.

In the remainder of this letter, we let $\mathcal{M} \subset \mathbb{R}^M$ be an embedded submanifold with $m = \dim \mathcal{M} \le M$ equipped with the standard Riemannian metric inherited from $\mathbb{R}^M$. Riemannian optimization methods can be applied to the minimization of a least-squares cost function

$$f : \mathcal{M} \to \mathbb{R},\ p \mapsto \tfrac{1}{2}\|F(p)\|^2 \quad \text{with } F : \mathcal{M} \to \mathbb{R}^N. \tag{3}$$

Recall that Newton's method for minimizing f consists of choosing a $x_0 \in \mathcal{M}$ and then generating a sequence of iterates $x_1, x_2, \ldots$ in $\mathcal{M}$ according to the following process:

$$x_{k+1} \leftarrow R_{x_k}(\eta_k) \quad \text{with } \left(\nabla^2_{x_k} f\right)\eta_k = -\nabla_{x_k} f; \tag{4}$$

herein, $\nabla_{x_k} f : \mathrm{T}_{x_k}\mathcal{M} \to \mathbb{R}$ is the *Riemannian gradient*, and $\nabla^2_{x_k} f : \mathrm{T}_{x_k}\mathcal{M} \to \mathrm{T}_{x_k}\mathcal{M}$ is the *Riemannian Hessian*; for details see [4, Chapter 6]. The map $R_{x_k} : \mathrm{T}_{x_k}\mathcal{M} \to \mathcal{M}$ is a *retraction operator*.

[1] Both the optimization problem (1) and the condition number of maps between manifolds can be defined for abstract manifolds. Nevertheless, we consider embedded manifolds because it greatly simplifies the proof of the main theorem, allowing us to compare tangent spaces in the ambient space using Wedin's theorem [2, Chapter III, Theorem 3.9]. This is no longer possible for abstract manifolds, which would make the letter much more difficult to understand. In practice, many manifolds are naturally embedded.

[2] Note that this is the geometric condition number at the output rather than the input of $\Phi_{\mathbf{x}}^{-1}$. The reason is that the PIP can have several $\mathbf{x}_i \in \mathcal{M}$ as solutions. Since the RGNs will only output one of these solutions, say $\mathbf{x}_1$, the natural question is whether this computed solution $\mathbf{x}_1$ is stable to perturbations of $\Phi(\mathbf{x}_1)$.

Download English Version:

https://daneshyari.com/en/article/8054049

Download Persian Version:

https://daneshyari.com/article/8054049

Daneshyari.com