Workshop on
Nonsmooth Optimization and Applications
NOPTA 8-12 April 2024
University of Antwerp, Belgium

Riemannian trust-region methods for strict saddle functions

Florentin Goyens and Clément W. Royer
University Paris Dauphine-PSL
goyensflorentin@gmail.com

University of Antwerp

Abstract

For functions with strict saddle points and isolated minimizers, we show that the classical trust-region methods finds an approximate second-order critical point in a number of iterations that depends logarithmically on the accuracy parameter, which significantly improves known results for general nonconvex optimization.

Introduction

Consider the following optimization problem:

$$\min_{x \in \mathcal{M}} f(x) \tag{1}$$

where $\mathcal{M}$ is a (smooth) Riemannian manifolds and $f : \mathcal{M} \to \mathbb{R}$ is smooth and nonconvex.

Target points satisfy second-order necessary optimality conditions: $x \in \mathcal{M}$,

$$\|\mathrm{grad} f(x)\| \le \varepsilon_g \quad \text{and} \quad \lambda_{\min}(\mathrm{Hess} f(x)) \ge -\varepsilon_H. \tag{2}$$

- Second-order Riemannian trust-region method produces an approximate second-order critical point (2) in at most $\mathcal{O}\left(1/\min\left(\varepsilon_g^2, \varepsilon_H^3\right)\right)$ iterations [2].
- These are pessimistic worst-case bounds which do not reflect the practical behaviour of these methods on many applications.
- We improve this result in the case of **strict saddle functions**

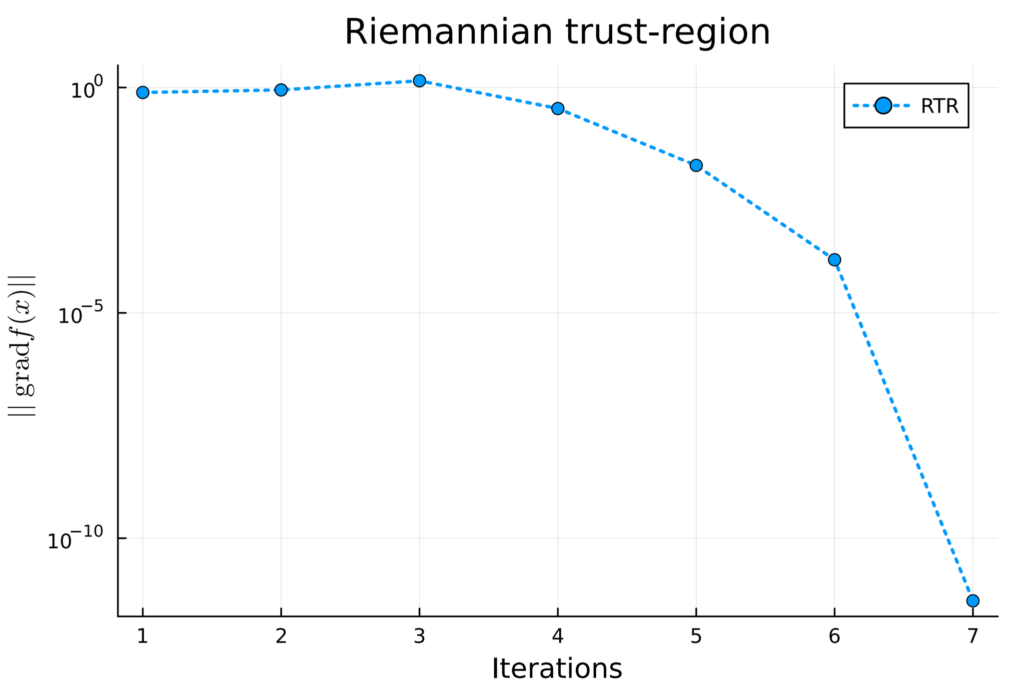

Figure 1: *Minimization of the Rayleigh quotient on the sphere of dimension* 10^3.

A class of strict saddle functions

Definition 1 *Let* $f : \mathcal{M} \to \mathbb{R}$ *be twice differentiable and let* $\alpha, \beta, \gamma, \delta$ *be positive constants. The function* f *is* $(\alpha, \beta, \gamma, \delta)$-strict saddle *if the manifold* $\mathcal{M}$ *satisfies* $\mathcal{M} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$, *where*

$\mathcal{R}_1 = \{x \in \mathcal{M} : \|\mathrm{grad} f(x)\| \ge \alpha\}$
$\mathcal{R}_2 = \{x \in \mathcal{M} : \lambda_{\min}(\mathrm{Hess} f(x)) \le -\beta\}$
$\mathcal{R}_3 = \{x \in \mathcal{M} :$ *there exists* $x^* \in \mathcal{M}$, *a local minimizer of* f *such that* $\mathrm{dist}(x, x^*) \le \delta$ *and* f *is geodesically* γ*-strongly convex over the set* $\{y \in \mathcal{M} : \mathrm{dist}(x^*, y) < 2\delta\}$.

Applications There are many applications...

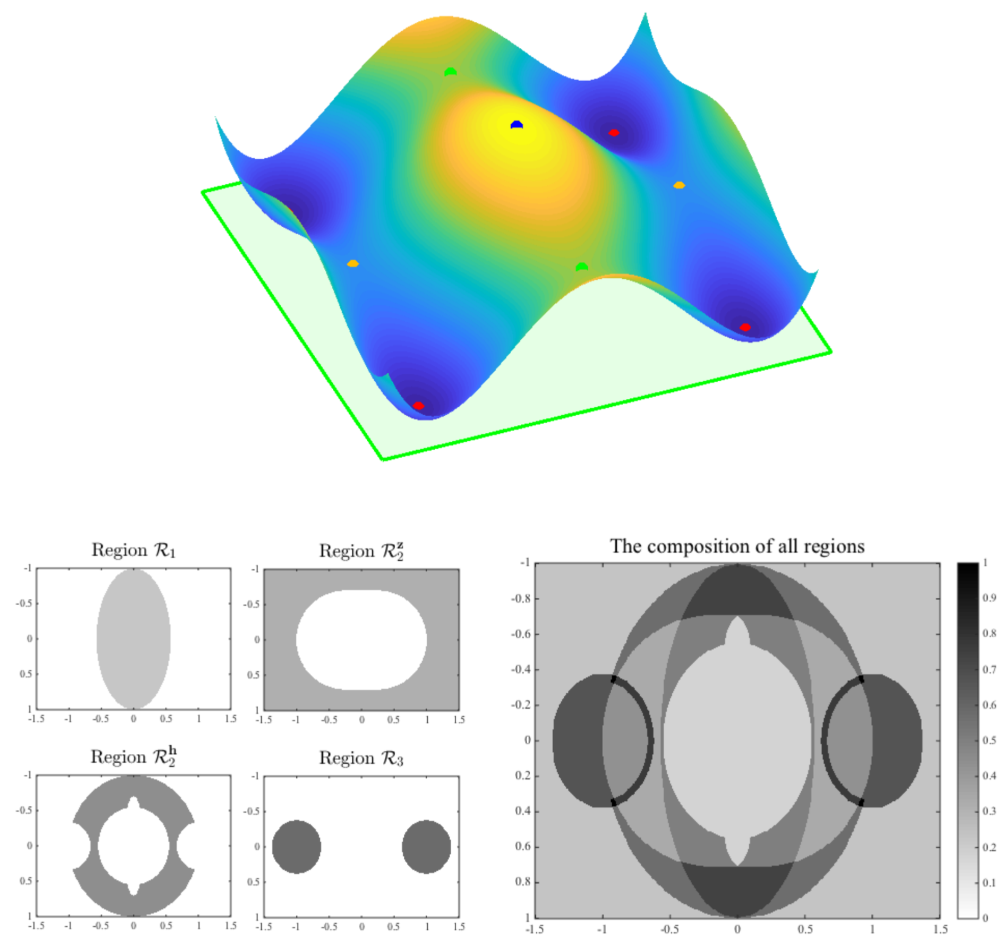

Figure 3: *Landscape of phase retrieval [1]*

Trust-region methods

Variant of Newton's method, with global convergence guarantees. At each iterate $x_k \in \mathcal{M}$, build a local quadratic model of f:

$$m_k(s) = f(x_k) + \langle \mathrm{grad} f(x_k), s\rangle + \frac{1}{2}\langle s, H_k s\rangle \tag{3}$$

where $H_k : \mathrm{T}_{x_k}\mathcal{M} \to \mathrm{T}_{x_k}\mathcal{M}$ is the Hessian of f.

Algorithm 1: Riemannian trust-region method
input : $x_0 \in \mathcal{M}$
output: Optimal point x^*.
1 **begin**
2 Compute s_k as a solution of the trust-region subproblem

$$s_k \in \operatorname*{arg\,min}_{s \in \mathrm{T}_{x_k}\mathcal{M}} m_k(s) \text{ subject to } \|s\| \le \Delta_k, \tag{4}$$

where m_k is the model defined by (3).

3 Compute $\rho_k = \dfrac{f(x_k) - f(R_{x_k}(s_k))}{m_k(0) - m_k(s_k)}$ and set

$$x_{k+1} = \begin{cases} R_k(x_k) & \text{if } \rho_k \ge \eta_1 \\ x_k & \text{otherwise.} \end{cases}$$

4 Set $\Delta_{k+1} =$

$$\begin{cases} \min(\tau_2 \Delta_k, \bar{\Delta}) & \text{if } \rho_k > \eta_2 \quad \text{[very successful]} \\ \Delta_k & \text{if } \eta_2 \ge \rho_k \ge \eta_1 \quad \text{[successful]} \\ \tau_1 \Delta_k & \text{otherwise.} \quad \text{[unsuccessful]} \end{cases}$$

5 **end**

Main Results

Theorem 1 *Under reasonable assumptions, Algorithm 1 produces an iterate satisfying* (2) *in at most*

$$\frac{C_1}{\min\left(\alpha^2, \alpha^{4/3}\beta, \alpha^{4/3}\gamma, \alpha^{2/3}\gamma^2, \beta^3, \beta^2\gamma, \beta\gamma^2, \gamma^3, \gamma^2\delta\right)} + 1 + \log_2\log_2\left(C_2\gamma^2\varepsilon_g^{-1}\right)$$

successful iterations, where the constant $C_1, C_2 > 0$ *depends on smoothness and algorithmic constants.*

- Complexity depends on landscape parameters and not on accuracy ε.
- Newton's method on strongly convex function requires $\mathcal{O}\left(\gamma^{-5} + \log\log(\gamma^3\varepsilon^{-1})\right)$ iterations in the worst-case [Boyd and Vandenberghe, 2004].
- If α, β, γ are known, we show a similar result for inexact minimization of the subproblem (4) using truncated conjugate cradients (tCG).
- Extends the ideas of [1] and [3] to generic manifolds and a generic second-order method.

Conclusions

Strict saddle properties improve the worst-case complexity of optimization algorithms.

Acknowledgements

The authors would like to express their gratitude to ...
This work was carried out with the support of ...

References

[1] Sun J., Qu Q., Wright J, A geometric analysis of phase retrieval, Foundations of Computational Mathematics 18(5): 1131-1198, 2018.

[2] Boumal N., Absil P.-A., Cartis C., Global rates o fconvergence for nonconvex optimization on manifolds, IMA Journal of Numerical Analysis, 39(1):1–33, 2019.

[3] O'Neill M., Wright S. J., A line-search descent algorithm for strict saddle functions with complexity guarantees, Journal of Machine Leaning and Research, 24(10):1–34, 2023.

Download

PAPER POSTER