Efficient recursive estimates in a Riemannian manifold

Jialun Zhou [1] **Salem Said** [2]

1. Introduction

Stochastic approximation is an important field of research in applied mathematics (Duflo, 1996; Borkar, 2008), with a wide range of applications in the information sciences (Benveniste et al., 1990; Kushner & Yin, 2003). Stochastic approximation in Riemannian manifolds was recently studied by (Bonnabel, 2013), who proved the convergence of the Riemannian stochastic gradient algorithm, and by (Zhang & Sra, 2016), who studied the rate of convergence of this algorithm, under convexity assumptions. For the problem of computing Riemannian p-barycentres, (Arnaudon et al., 2012) previously showed the asymptotic normality of the Riemannian stochastic gradient algorithm. Our ongoing research aims to provide a detailed study of the application of stochastic approximation, to the recursive estimation of statistical parameters which belong to Riemannian manifolds. The contribution presented in the current submission extends the theory developed by (Nevilson & Hasminskii, 1973), from Euclidean space to any Riemannian manifold. The main results are stated in Section 3, and their proofs available online (Zhou & Said, 2018).

2. Problem statement

Let (P, Θ, X) be a statistical model, with parameter space Θ and sample space X, where Θ is a complete Riemannian manifold. Let $(x_n\,; n = 1, \dots)$ be i.i.d. data with distribution P_{θ^*}, for some $\theta^* \in \Theta$. Consider the recursive estimates $(\theta_n\,; n = 1, \dots)$, given by the decreasing-step-size algorithm,

$$\theta_{n+1} \;=\; \exp_{\theta_n}(\gamma_{n+1} u(\theta_n\,, x_{n+1})) \tag{1}$$

where $\exp$ is the Riemannian exponential mapping, γ_n a decreasing sequence of step sizes, and $u(\theta, x)$ is a continuous vector field on Θ, for each $x \in X$. The step sizes γ_n verify

$$\sum \gamma_n \;=\; \infty \qquad\qquad \sum \gamma_n^2 \;<\; \infty \tag{2}$$

Problem : *under what conditions on algorithm (1) will the recursive estimates θ_n be asymptotically efficient?*

[1] Master Modélisation Statistique, Université de Franche-Comté
[2] Laboratoire IMS, Université de Bordeaux. Correspondence to: Salem Said <salem.said@u-bordeaux.fr>.

3. Main results

The following assumptions are considered, where $D(\theta) = D(\theta^*|\theta)$ denotes the Kullback-Leibler divergence :

(A1) $D(\theta)$ has an isolated stationary point at θ^* .
(A2) $D(\theta)$ has Lipschitz gradient in a neighborhood of θ^* .
(A3) the gradient of $D(\theta)$ verifies the identity

$$\nabla D(\theta) \;=\; -E_{\theta^*}\, u(\theta, x) \tag{3}$$

(A4) in a neighborhood of θ^*, there is a uniform upper bound on

$$R(\theta) = E_{\theta^*} \|u(\theta, x)\|^4 \tag{4}$$

(A5) $D(\theta)$ is twice differentiable at $\theta = \theta^*$, with positive definite Hessian.

In a system of normal coordinates θ^α, with origin at θ^*, let Σ^* denote the covariance matrix of the components of $u(\theta^*, x)$, and H denote the matrix of the Hessian of $D(\theta)$ at θ^*, with $\lambda > 0$ its smallest eigenvalue.

Proposition 1 *Assume (A1) to (A5) are verified, and the recursive estimates θ_n all lie in a compact convex neighborhood of θ^*. If $\gamma_n = \frac{a}{n}$ where $2a\lambda > 1$, then*

$$\mathbb{E}\, d^2(\theta_n, \theta^*) = O\left(n^{-1}\right) \tag{5a}$$

$$d^2(\theta_n, \theta^*) = o(n^{-p}) \;\; \textit{for}\; p \in (0, 1) \tag{5b}$$

where $d(\cdot, \cdot)$ denotes Riemannian distance. Moreover, the distribution of the re-scaled coordinates $n^{1/2}\theta^\alpha$ converges to a centred normal distribution with covariance matrix Σ given by Lyapunov's equation

$$A\Sigma + \Sigma A = -a^2\Sigma^* \tag{5c}$$

where $A = \frac{1}{2}I - aH$, with I the identity matrix.

4. Discussion of main results

Proposition 1 solves the problem of Section 2. As a corollary of this proposition, if the Riemannian metric of Θ coincides with the information metric of the model P, and if $a = 1$, then the asymptotic covariance matrix Σ of (5c) is equal to the inverse of the Fisher information matrix. In other words, the recursive estimates θ_n asymptotically achieve the Cramér-Rao lower bound, and are therefore asymptotically efficient.

Acknowledgements

The authors would like to thank Pr. Jonathan H. Manton, at the University of Melbourne, for helpful discussion and advice. The current work was funded by LIA GEODESIC, the international research laboratory of CNRS and the University of Melbourne. Jialun Zhou's Masters project is funded by the SysNum cluster of excellence, part of IdEx Bordeaux.

References

Arnaudon, M., Dombry, C., Phan, A., and Le, Y. Stochastic algorithms for computing means of probability measures. *Stoch. Proc. Appl.*, 122(4):1437–1455, 2012.

Benveniste, A., Métivier, M., and Priouret, P. *Adaptive algorithms and stochastic approximation*. Springer, 1990.

Bonnabel, S. Stochastic gradient descent on Riemannian manifolds. *IEEE. Trans. Autom. Control.*, 58(9):2217–2229, 2013.

Borkar, V. *Stochastic approximation : a dynamical systems viewpoint*. Cambridge University Press, 2008.

Duflo, M. *Algorithmes stochastiques*. Springer, 1996.

Kushner, H. and Yin, G. *Stochastic approximation, recursive algorithms, and applications*. Springer, 2003.

Nevilson, M. B. and Hasminskii, R. Z. *Stochastic approximation and recursive estimation*. American Mathematical Society, 1973.

Zhang, H. and Sra, S. First-order methods for geodesically convex optimization. In *COLT*, pp. 1617–1638, 2016.

Zhou, J. and Said, S. Recursive parameter estimation in a Riemannian manifold. 2018. URL `https://arxiv.org/abs/1805.06811`.