
Stochastic Optimization of Black-box Functions on Riemannian Manifolds Using Kernel Adaptive Sequential Monte Carlo

Abstract

This paper puts forward a new stochastic method for global optimization of black-box functions on a Riemannian manifold \mathcal{M} , which is rather challenging yet useful in many areas. The proposed algorithm is based on *Kernel Adaptive Sequential Monte Carlo* (KASMC) by exploiting the closeness between simulated annealing and sequential Monte Carlo (SMC), and thus is referred to as KASMC optimizer. At each Markov chain Monte Carlo (MCMC) transition of KASMC optimizer, samples are first proposed in a reproducing kernel Hilbert space (RKHS) of the “envelop space” of \mathcal{M} , after tempered acceptance criterion, then they are mapped back to the envelop space and \mathcal{M} respectively by two projection steps. The strength of KASMC optimizer, compared to classical simulated annealing, stems from its two adaptive tuning mechanisms. The first one is adaptive tempering, which automatically constructs temperature schedule through cooling; the second one is adaptive MCMC transition in RKHS, which is critical to handle nonlinearity of support of functions on \mathcal{M} . Results in our experiments demonstrates the promising applicability of KASMC optimizer.

1 Introduction

Global optimization on a black-box function f has been a challenging task despite its wide usage in many disciplines [Rios and Sahinidis, 2013, Conn et al., 2009]. Difficulties are usually considered from two aspects: *i.* the analytic form of f is unknown and function values can be only accessed point-wisely, thus no function landscape property (e.g. convexity) can be

exploited for numerical solutions; *ii.* function evaluation at one point is usually at high cost, which narrows algorithms to be practical only if it can find optimal (or with tolerant imprecision) with an affordable iteration budget. In this paper, we consider an even more “nasty” case, where f is defined on a Riemannian manifold \mathcal{M} . Since the support of f is \mathcal{M} instead of \mathbb{R}^d , search is supposed to be more careful since exploration and exploitation is restricted to \mathcal{M} . Very often the shape of \mathcal{M} is irregular, which worsens the situation. However, this task arises in many areas [Papazov and Burschka, 2011, Byrne and Girolami, 2013], e.g. rotation optimization to find best shape registration, simplex optimization to best fit categorical distributions. Therefore, it is rather desirable to come up with a general algorithm which can efficiently find good solutions with guaranteed precision.

Simulated annealing (SA) is a classical stochastic optimization for black-box functions [Kirkpatrick et al., 1983]. However, its applicability is limited to common case where f is defined on \mathbb{R}^d . In present paper, we propose a novel stochastic method to extend simulated annealing to handle f on Riemannian manifolds. The proposed algorithm is based on *Kernel Adaptive Sequential Monte Carlo* (KASMC) by exploiting the close relation between SA and sequential Monte Carlo (SMC), and therefore is referred to as KASMC optimizer. Similar to SA, KASMC optimizer increases the determinism of proposal acceptance criterion by monotonically decreasing a temperature parameter. However, in KASMC, sequential tempering is implemented in a SMC instead, which can be considered as a multi-particle simulated annealing [?] with weighting and resampling. At each Markov chain Monte Carlo (MCMC) transition in the SMC, samples are first proposed in a reproducing kernel Hilbert space (RKHS) of the “envelop space” of \mathcal{M} , after tempered acceptance criterion, then they are mapped back to the envelop space and \mathcal{M} respectively by two projection steps. The adaptive proposal construction in RKHS is inspired by MCMC Kameleon [Sejdinovic et al., 2014].

The strength of KASMC optimizer stems from its two adaptive tuning mechanisms. The first one is adaptive tempering, which, based on *effective sampling size* (ESS), automatically constructs temperature schedule through cooling; the second one is adaptive MCMC transition in RKHS, which is critical to handle nonlinearity of support of f on \mathcal{M} . Based on previous study, which includes the finite-time performance guarantees of simulated annealing on continuous domains [Lecchini-visintini et al., 2007] and convergence property of adaptive SMC [Beskos et al., 2013], KASMC optimizer is expected to work efficiently and effectively. In our experiments, the proposed KASMC optimizer was evaluated on a practical task: optimal rotation search for 3D registration. Two Riemannian manifolds were tried: unit quaternion manifold and $SO(3)$ manifold. According to our empirical results, KASMC optimizer consistently outperforms SA and other non-adaptive counterparts.

This paper makes four contributions. First, we reveal the closeness between SA and SMC, which sheds light on new insights into these two techniques. Second, to better fit nonlinear support (here Riemannian manifold), MCMC Kameleon is for the first time employed in adaptive SMC. Third, we put forward a simpler yet better pre-image method to replace the naive one-step gradient projection in MCMC Kameleon. Last but not least, a stochastic optimization method of black-box functions on Riemannian manifolds is proposed. Under some weak assumptions, it can be widely applied on general functions and Riemannian manifolds.

1.1 Related Work

KASMC optimizer can be related to many other work from different perspectives. The following provides a short summary of recent development of relevant study.

Black-box function global optimization. A review of historical methods for black-box function global optimization can be found in Rios and Sahinidis [2013] or Conn et al. [2009]. Out of all work, *Bayesian optimization* [Brochu et al., 2010, Adams et al., 2014] and *Hierarchical optimistic optimization* [Munos, 2014] are two particularly notable research branches which have recently attracted particularly more attention. Bayesian optimization is an extension *response surface method* with Gaussian process to interpolate the unknown function f . Hierarchical optimistic optimization is an application of *optimism principle* on *Monte Carlo tree search* (MCTS), which partitions search spaces hierarchical and conducts exploration and exploitation at different scales. A combination of these two techniques was proposed in Wang

et al. [2014]. To the best of our knowledge, no work has been conducted for Bayesian optimization or optimistic optimization on Riemannian manifolds. Only one work on Bayesian optimization with constraints [Gardner et al., 2014] can be extended to Riemannian manifolds.

SA and SMC. On the one hand, SA is a stochastic optimization and it was originally proposed as an adaption of Metropolis-Hasting algorithm to obtain samples from thermodynamic systems [Kirkpatrick et al., 1983]. On the other hand, SMC is sampler for state-space models or complex distributions [Del Moral et al., 2006]. Since SMC is more often used in filtering community, it is also referred to as *particle filter* and also its samples are called particles. Although two methods are motivated for different purposes,

Adaptive SMC [Del Moral et al., 2012, Jasra et al., 2011, Schäfer and Chopin, 2013],

MCMC on Riemannian Manifold

2 Preliminaries

2.1 Assumptions

First level headings are all caps, flush left, bold and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

2.2 Simulated Annealing

Second level headings must be flush left, all caps, bold and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

2.3 Sequential Monte Carlo Sampler

Third level headings must be flush left, initial caps, bold, and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

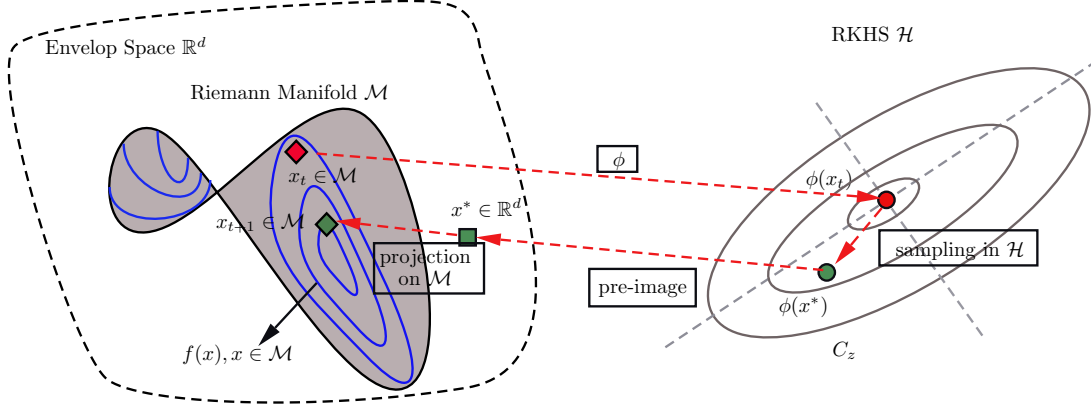


Figure 1: An overview of kernel Adaptive MCMC transition and two inverse projection steps.

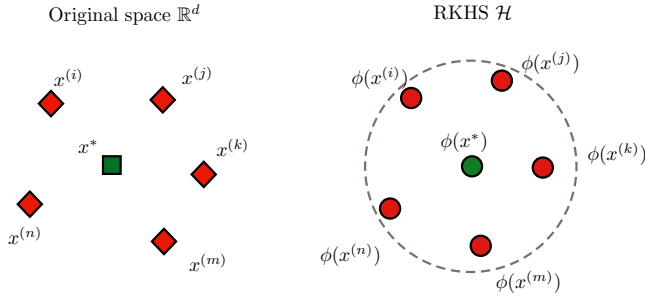


Figure 2: Pre-image by using local topology consistency.

2.4 Adaptive Tempering

2.5 Kernel Adaptive MCMC

2.6 Optimization on Riemannian Manifolds

3 Proposed Algorithm

3.1 Overview

3.2 Pre-Image

3.3 Projection onto \mathcal{M}

4 Experiments

4.1 Unit Quaternion Manifold Optimization

4.2 $SO(3)$ Manifold Optimization

References

Ryan P Adams, Zoubin Ghahramani, Matthew W Hoffman, Jasper Snoek, and Kevin Swersky. NIPS workshop on Bayesian Optimization in Academia and Industries, 2014.

A Beskos, A Jasra, N Kantas, and A Thiery. On the convergence of adaptive sequential monte carlo methods. Technical Report arXiv:1306.6462v3, arXiv.org, 2013.

Eric Brochu, Vlad M Cora, and Nando de Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. Technical Report arXiv:1012.2599, arXiv.org, December 2010.

Simon Byrne and Mark Girolami. Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.

Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2009.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, June 2006.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. An adaptive sequential monte carlo method for approximate bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.

Jacob Gardner, Matt Kusner, Kilian Q. Weinberger, John Cunningham, and Zhixiang Xu. Bayesian optimization with inequality constraints. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 937–945. JMLR Workshop and Conference Proceedings, 2014.

Ajay Jasra, David A. Stephens, Arnaud Doucet, and Theodoros Tsagaris. Inference for lévy-driven stochastic volatility models via adaptive sequential monte carlo. *Scandinavian Journal of Statistics*, 38(1):1–22, 2011.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Op-

timization by simulated annealing. *SCIENCE*, 220 (4598):671–680, 1983.

Andrea Lecchini-visintini, John Lygeros, and Jan Maciejowski. Simulated annealing: Rigorous finite-time guarantees for optimization on continuous domains. In *Advances in Neural Information Processing Systems (NIPS)*, pages 865–872. MIT Press, 2007.

Rémi Munos. From bandits to monte-carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends in Machine Learning*, 7(1):1–129, 2014.

Chavdar Papazov and Darius Burschka. Stochastic global optimization for robust point set registration. *Comput. Vis. Image Underst.*, 115(12):1598–1609, December 2011.

LuisMiguel Rios and NikolaosV. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.

Christian Schäfer and Nicolas Chopin. Sequential monte carlo on large binary sampling spaces. *Statistics and Computing*, 23(2):163–184, 2013.

D. Sejdinovic, H. Strathmann, M. Lomeli, C. Andrieu, and A. Gretton. Kernel Adaptive Metropolis-Hastings. In *ICML*, pages 1665–1673, 2014.

Ziyu Wang, Babak Shakibi, Lin Jin, and Nando de Freitas. Bayesian multi-scale optimistic optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1005–1014, 2014.