Model-Based Relative Entropy Stochastic Search

Abbas Abdolmaleki
IEETA, University of Aveiro
DSI, University of Minho
LIACC, University of Porto
Portugal
abbas.a@ua.pt

Rudolf Lioutikov
IAS,TU Darmstadt, Darmstadt
Lioutikov@ias.tu-darmstadt.de

Nuno Lau
DETI, IEETA, University of Aveiro, Portugal
nunolau@ua.pt

Luis Paulo Reis
DSI, University of Minho
LIACC, University of Porto
Portugal
lpreis@dsi.uminho.pt

Jan Peters
IAS,TU Darmstadt, Darmstadt
MPI for Intelligent Systems
Stuttgart, Germany
peters@ias.tu-darmstadt.de

Gerhard Neumann
CLAS, TU Darmstadt, Darmstadt, Germany
neumann@ias.tu-darmstadt.de

1. INTRODUCTION

Stochastic search algorithms [1, 2] are black box optimizers of an objective function $R(\boldsymbol{\theta}) : \mathbb{R}^n \rightarrow \mathbb{R}$. The goal is to find one or more parameter vectors $\boldsymbol{\theta} \in \mathbb{R}^n$ which have the highest possible objective value. The only accessible information on $R(\boldsymbol{\theta})$ are (possibly noisy) evaluations $\{R^{[k]}\}_{k=1\ldots N}$ of parameter vectors $\{\boldsymbol{\theta}^{[k]}\}_{k=1\ldots N}$, where k is the index of the sample and N is number of samples. Stochastic search algorithms typically maintain a search distribution $\pi(\boldsymbol{\theta})$ over the parameter space $\boldsymbol{\theta}$ of the objective function $R(\boldsymbol{\theta})$. The search distribution $\pi(\boldsymbol{\theta})$ is implemented as a multivariate Gaussian distribution, i.e., $\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In each iteration, the search distribution $\pi(\boldsymbol{\theta})$ is used to create samples $\boldsymbol{\theta}^{[k]}$ of the parameter vector $\boldsymbol{\theta}$. Subsequently, the (possibly noisy) evaluation $R^{[k]}$ of $\boldsymbol{\theta}^{[k]}$ is obtained by querying the objective function. Subsequently using the samples $\{\boldsymbol{\theta}^{[k]}, R^{[k]}\}_{k=1\ldots N}$, a new stochastic search distribution is computed. Information-theoretic search distribution updates [3] bound the Kullback Leibler divergence between two subsequent search distributions. Using a KL-bound for the update of the search distribution is a common approach in stochastic search. However, such information theoretic bounds could so far only be approximately applied either by using Taylor-expansions of the KL-divergence resulting in natural evolutionary strategies (NES) [2], or sample-based approximations, resulting in the relative entropy policy search (REPS) [3] algorithm. In this paper, we present a novel stochastic search algorithm which is called MOdel-based Relative-Entropy stochastic search (MORE). Our algorithm bounds the KL divergence of the new and old search distribution in closed form. In order to do so, we locally learn a simple, quadratic surrogate of the objective function. The quadratic surrogate allows us to compute the new search distribution analytically where the KL divergence of the new and old distribution is bounded. Therefore, we only exploit the surrogate model locally which prevents the algorithm to be misled by inaccurate optima introduced by an inaccurate surrogate model. In addition to solving the search distribution update in closed form, we also upper-bound the entropy of the new search distribution to ensure that exploration is sustained in the search distribution throughout the learning progress, and, hence, premature convergence is avoided. We provide a comparison of stochastic search algorithms.

GECCO'16 Companion July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07…$15.00

DOI: http://dx.doi.org/10.1145/2908961.2930952

2. MODEL-BASED RELATIVE ENTROPY STOCHASTIC SEARCH

Similar as in [3], we can formulate an optimization problem to obtain a new search distribution that maximizes the expected objective value while upper-bounding the KL-divergence and lower-bounding the entropy of the distribution, i.e.,

$$\max_{\pi} \int \pi(\boldsymbol{\theta}) \mathcal{R}_{\boldsymbol{\theta}} d\boldsymbol{\theta} \tag{1}$$

$$\text{s.t. } \mathrm{KL}\big(\pi(\boldsymbol{\theta})||q(\boldsymbol{\theta})\big) \leq \epsilon, \qquad H(\pi) \geq \beta, \qquad 1 = \int \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

where $\mathcal{R}_{\boldsymbol{\theta}}$ denotes the expected objective when evaluating parameter vector $\boldsymbol{\theta}$. The term $H(\pi) = -\int \pi(\boldsymbol{\theta}) \log \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$ denotes the entropy of the distribution π and $q(\theta)$ is the old distribution. The parameters ϵ and β are user-specified parameters to control the exploration-exploitation trade-off of the algorithm. We can obtain a closed form solution for $\pi(\boldsymbol{\theta})$ by optimizing the Lagrangian for the optimization problem given above. This solution is given as

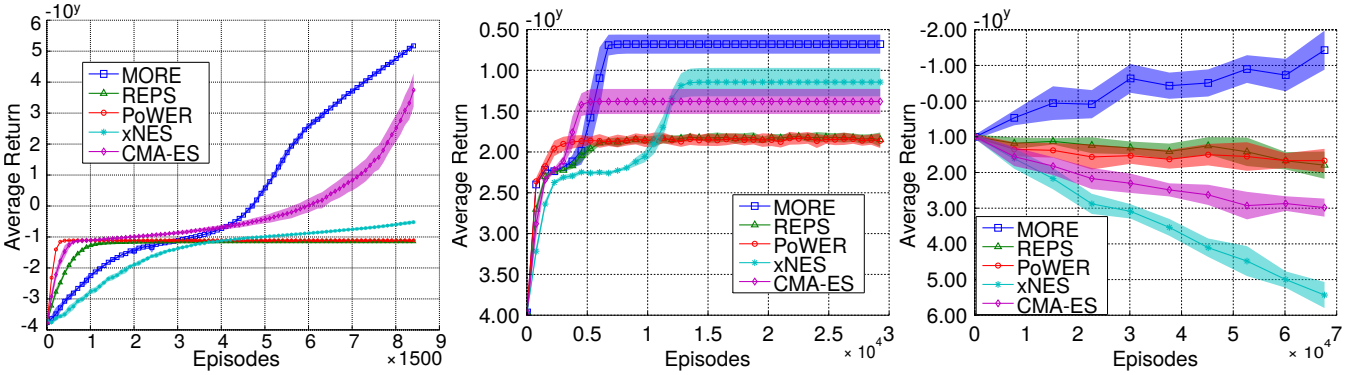
$$\pi(\boldsymbol{\theta}) \propto q(\boldsymbol{\theta})^{\eta/(\eta+\omega)} \exp\left(\frac{\mathcal{R}_{\boldsymbol{\theta}}}{\eta+\omega}\right), \tag{2}$$

where η and ω are the Lagrangian multipliers. The optimal value for η and ω can be obtained by minimizing the convex dual function $g(\eta, \omega)$ such that $\eta > 0$ and $\omega > 0$. The dual function $g(\eta, \omega)$ is given by

$$g(\eta, \omega) = \eta\epsilon - \omega\beta + (\eta+\omega) \log\left(\int q(\boldsymbol{\theta})^{\frac{\eta}{\eta+\omega}} \exp\left(\frac{\mathcal{R}_{\boldsymbol{\theta}}}{\eta+\omega}\right) d\boldsymbol{\theta}\right). \tag{3}$$

As we are dealing with continuous distributions, the entropy can also be negative. We specify β such that the relative difference of $H(\pi)$ to a minimum exploration policy $H(\pi_0)$ is decreased for a certain percentage, i.e., we change the entropy constraint to

$$H(\pi) - H(\pi_0) \geq \gamma(H(q) - H(\pi_0)) \Rightarrow \beta = \gamma(H(q) - H(\pi_0)) + H(\pi_0).$$



(a) Rosenbrock (b) Rastrigin (c) Noisy Function

Figure 1: Comparison of stochastic search methods for optimizing (a) Rosenbrock function and (b) Rastrigin function. (c) Comparison for a noisy objective function. All results show that MORE performs favourably.

We set minimum entropy $H(\pi_0)$ of search distribution to a small enough value like -75.

2.1 Analytic Solution of the Dual-Function and the Search Distribution

Using a quadratic surrogate model of the objective function, we can compute the integrals in the dual function analytically, and, hence, we can satisfy the introduced bounds exactly in the MORE framework. At the same time, we take advantage of surrogate models such as a smoothed estimate in the case of noisy objective functions. We will assume that we are given a quadratic surrogate model $\mathcal{R}_{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^T \boldsymbol{R}\boldsymbol{\theta} + \boldsymbol{\theta}^T \boldsymbol{r} + r_0$ of the objective function $\mathcal{R}_{\boldsymbol{\theta}}$ which we will learn from data [1]. Moreover, the search distribution is Gaussian, i.e., $q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{b}, \boldsymbol{Q})$. In this case the integrals in the dual function given in Equation 3 can be solved in closed form. The integral inside the log-term in Equation (3) now represents an integral over an un-normalized Gaussian distribution. Hence, the integral evaluates to the inverse of the normalization factor of the corresponding Gaussian. The dual can be written as

$$g(\eta, \omega) = \eta\epsilon - \beta\omega + \frac{1}{2}\left(\boldsymbol{f}^T \boldsymbol{F}\boldsymbol{f} - \eta \boldsymbol{b}^T \boldsymbol{Q}^{-1}\boldsymbol{b} - \eta \log|2\pi\boldsymbol{Q}| + (\eta+\omega)\log|2\pi(\eta+\omega)\boldsymbol{F}|\right) \tag{4}$$

with $\boldsymbol{F} = (\eta\boldsymbol{Q}^{-1} - 2\boldsymbol{R})^{-1}$ and $\boldsymbol{f} = \eta\boldsymbol{Q}^{-1}\boldsymbol{b} + \boldsymbol{r}$. Hence, the dual function $g(\eta, \omega)$ can be efficiently evaluated by matrix inversions and matrix products. Note that, for a large enough value of η, the matrix $\boldsymbol{F}$ will be positive definite and hence invertible even if $\boldsymbol{R}$ is not. In our optimization, we always restrict the η values such that $\boldsymbol{F}$ stays positive definite. Nevertheless, we could always find the η value with the correct KL-divergence. We can also obtain the update rule for the new policy $\pi(\boldsymbol{\theta})$. From Equation (2), we know that the new policy is the geometric average of the Gaussian sampling distribution $q(\boldsymbol{\theta})$ and a squared exponential given by the exponentially transformed surrogate. After rearranging terms and completing the square, the new policy can be written as $\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{F}\boldsymbol{f}, \boldsymbol{F}(\eta+\omega))$, where $\boldsymbol{F}$, $\boldsymbol{f}$ are given in the previous section. objective function.

[1] In order to learn the local quadratic surrogate, we can use linear regression to fit a function of the form $f(\boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{\theta})\boldsymbol{\beta}$, where $\boldsymbol{\phi}(\boldsymbol{\theta})$ is a feature function that returns a bias term, all linear and all quadratic terms of $\boldsymbol{\theta}$.

3. EXPERIMENTS

We compare MORE with state of the art methods in stochastic search and policy search such as CMA-ES [1], NES [2], PoWER [4] and episodic REPS [3]. We use standard optimization test functions. We chose Rosenbrock function $f(\boldsymbol{x}) = \sum_{i=1}^{n-1}[100(x_{i+1} - x_i^2)^2 + (1-x_i)^2]$, and a multi-modal function which is known as the Rastgirin function $f(x) = 10n + \sum_{i=1}^{n}[x_i^2 - 10\cos(2\pi x_i)]$. All these functions have a global minimum equal $f(\boldsymbol{x}) = 0$. We use a 15 dimensional version of these functions. In our experiments, the mean of the initial distributions has been chosen randomly.

Algorithmic Comparison. We compared our algorithm against CMA-ES, NES, PoWER and REPS. In each iteration, we generated 15 new samples [2]. For MORE, REPS and PoWER, we always keep the last $L = 150$ samples, while for NES and CMA-ES only the 15 current samples are kept[3]. As we can see in the Figure 1, MORE performs favourably in terms of both the learning speed and the final performance. However, in terms of the computation time, MORE was slower than the other algorithms. Yet, MORE was sufficiently fast as one policy update took less than $1s$.

Performance on a Noisy Function. We also conducted an experiment on optimizing the Sphere function where we add multiplicative noise to the reward samples, i.e., $y = f(\boldsymbol{x}) + \epsilon|f(\boldsymbol{x})|$, where $\epsilon \sim \mathcal{N}(0, 1.0)$ and $f(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{M}\boldsymbol{x}$ with a randomly chosen $\boldsymbol{M}$ matrix. Figure 1(c) shows that MORE successfully smooths out the noise and converges, while other methods diverge.

4. REFERENCES

[1] N. Hansen, S.D. Muller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, 2003.

[2] Y. Sun, D. Wierstra, T. Schaul, and J. Schmidhuber. Efficient Natural Evolution Strategies. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation(GECCO)*, 2009.

[3] A. Kupcsik, M. P. Deisenroth, J. Peters, and G. Neumann. Data-Efficient Contextual Policy Search for Robot Movement Skills. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2013.

[4] J. Kober and J. Peters. Policy Search for Motor Primitives in Robotics. *Machine Learning*, pages 1–33, 2010.

[2] We use the heuristics introduced in [1, 2] for CMA-ES and NES.
[3] NES and CMA-ES algorithms typically discard the old samples.