Exploring the common concepts of adaptive MCMC and Covariance Matrix Adaptation schemes

Christian L. Müller

Institute of Theoretical Computer Science and Swiss Institute of Bioinformatics,
ETH Zurich, CH-8092 Zurich, Switzerland
christian.mueller@inf.ethz.ch

Abstract. In the field of scientific modeling, one is often confronted with the task of drawing samples from a probability distribution that is only known up to a normalizing constant and for which no direct analytical method for sample generation is available. Since the past decade, adaptive Markov Chain Monte Carlo (MCMC) methods gained considerable attention in the statistics community in order to tackle this black-box (or indirect) sampling scenario. Common application domains are Bayesian statistics and statistical physics. Adaptive MCMC methods try to learn an optimal proposal distribution from previously accepted samples in order to efficiently explore the target distribution. Variable metric approaches in black-box optimization, such as the Evolution Strategy with covariance matrix adaptation (CMA-ES) and Gaussian Adaption (GaA), use almost identical ideas to locate putative global optima. This extended abstract summarizes the common concepts in adaptive MCMC and covariance matrix adaptation schemes. We also present how both types of methods can be unified within the Gaussian Adaptation framework and propose a unification of both fields as "grand challenge" for future research.

Keywords. Adaptive MCMC, Gaussian Adaptation, CMA-ES, covariance matrix adaptation

1 Black-box sampling and Black-box optimization

In many areas of science and engineering a modeler is often confronted with systems where no prior information is available about intrinsic system properties. Such systems are called black-box systems. The transfer characteristics of a black-box system can only be analyzed through (multivariate) system input $\mathbf{x}$ and corresponding (scalar) output $f(\mathbf{x})$ (see Fig. 1 a). Black-box optimization is concerned with the determination of the input that optimizes the output of the black box. In black-box sampling, the system represents a probability distribution from which unbiased samples need to be generated. Sketches of both scenarios are depicted in Fig. 1 b. and c, respectively. In this contribution

Dagstuhl Seminar Proceedings 10361
Theory of Evolutionary Algorithms
http://drops.dagstuhl.de/opus/volltexte/2010/2813

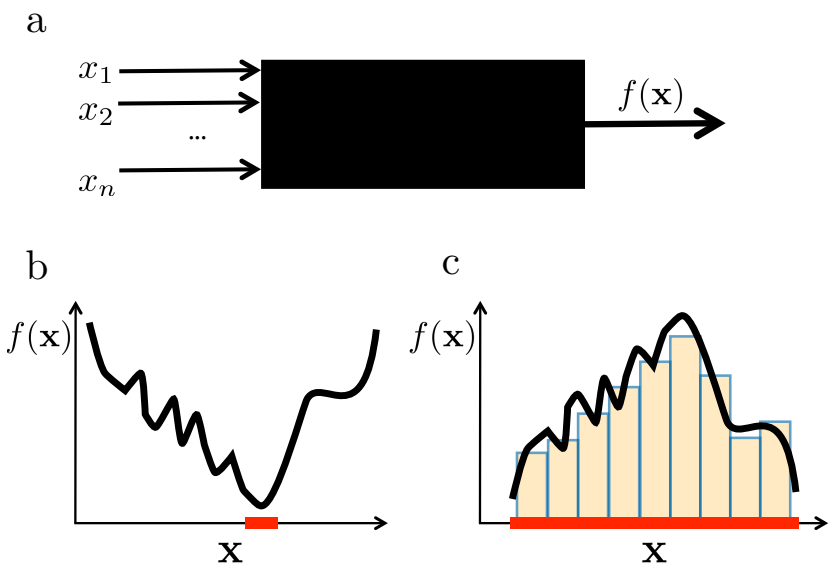

Fig. 1: a. Sketch of a black-box system with multivariate input $\mathbf{x} \in \mathbb{R}^n$ and scalar output $f(\mathbf{x}) \in \mathbb{R}$. b. Black-box optimization aims at finding the input (marked in red) that optimizes (here minimizes) the system output $f(\mathbf{x})$. b. In black-box sampling input samples are generated over the whole input domain (marked in red) proportional to the target probability. The resulting histogram of the produced ensemble (sketched in blue) should approximate the true underlying target probability $f(\mathbf{x})$.

we focus on black-box systems with real-valued multivariate input $\mathbf{x} \in \mathbb{R}^n$ and scalar output $f(\mathbf{x}) \in \mathbb{R}$.
Although black-box optimization and sampling face similar difficulties in practice, the scientific disciplines that investigate these scenarios are largely disjoint. Black-box optimization problems are generally investigated by the Evolutionary Computation community and, to less extent, by the Operations Research community. Black-box sampling, on the other hand, is a subfield of computational statistics. It comes at no surprise that a number of important ideas have been independently developed in the different communities. The most striking example is the emergence of iterative sampling schemes that continuously adapt the first and second moments of a multivariate Gaussian "search" or "proposal" distribution.

2 Covariance matrix adaptation schemes for black-box optimization

Two state-of-art continuous black-box optimizers employ the adaptation concept in their search for putative optimal solutions: Hansen's Evolution Strategy with

Covariance Matrix Adaptation (CMA-ES) [1, 2, 3] and Kjellström's Gaussian Adaptation algorithm [4, 5, 6, 7, 8].

Standard CMA-ES samples in each generation a population of candidate solutions from a multivariate Gaussian distribution. The sample points are ranked according to their objective function values, and a certain percentage of the best solutions are selected and used to both adapt the mean and the covariance matrix of the search distribution. The (1+1)-CMA-ES [9] is a variant that only produces a single sample per generation that is selected if it carries a lower objective function value than the previous.

Gaussian Adaptation is similar to the (1+1)-CMA-ES but uses threshold acceptance as selection mechanism. In each generation, samples are accepted and used for adaptation if their objective function values are below a threshold c_{T}. The threshold is decreased during the search process.

Both algorithms include a mechanism that adapts the global scale (or step size) of the search distribution. An important conceptual difference between CMA-ES and GaA is the purpose of covariance adaptation: While CMA-ES is designed to increase the likelihood of generating successful *search directions*, GaA adapts the covariance such as to maximize the *entropy* of the search distribution under the constraint that acceptable search points are found with a predefined, fixed acceptance probability.

3 Adaptive MCMC methods for black-box sampling

A large class of black-box samplers is based on the generation of so-called *Markov Chains*. Consider a sequence of random variables $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(g)}, \ldots, \mathbf{x}^{(N)}$ defined on a state space $\mathcal{X}$. This sequence is called a Markov Chain, if it satisfies the *Markov property*:

$$P(\mathbf{x}^{(g+1)} = \mathbf{y} \,|\, \mathbf{x}^{(g)} = \mathbf{x}, \ldots, \mathbf{x}^{(0)} = \mathbf{z}) = P(\mathbf{x}^{(g+1)} = \mathbf{y} | \mathbf{x}^{(g)} = \mathbf{x}) \,, \tag{1}$$

with $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$. This means that the next state $\mathbf{x}^{(g+1)}$ *only* depends on the current state $\mathbf{x}^{(g)}$ where $P(\cdot|\cdot)$ denotes the transition probability from a given state to the next one. In continuous state spaces $\mathcal{X}$, the transition probability function is modeled by a *transition density function* $A(\mathbf{x}, \mathbf{y})$.

Since Metropolis and co-workers introduced the Metropolis algorithm [10] and Hastings' generalization [11], the density $A(\mathbf{x}, \mathbf{y})$ of most MCMC method comprises two components, a proposal distribution $q(\cdot|\cdot)$ and an acceptance criterion α_{MH}. Green and Han [12] were among the first to employ the isotropic Gaussian distribution as $q(\cdot|\cdot)$ for continuous target distributions $\pi(\mathbf{x})$. The Metropolis-Hastings acceptance criterion is:

$$\alpha_{\mathrm{MH}}(\mathbf{x}, \mathbf{y}) = \min \left(1, \, \frac{\pi(\mathbf{y}) q(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{x}) q(\mathbf{x}|\mathbf{y})} \right) . \tag{2}$$

The transition density function for the general Metropolis-Hastings (MH) algorithm thus reads:

$$A(\mathbf{x}, \mathbf{y}) = q(\mathbf{x}|\mathbf{y}) \alpha_{\mathrm{MH}}(\mathbf{x}, \mathbf{y}) \,. \tag{3}$$

Adaptive MCMC methods have been introduced in order to avoid the difficulties of fine-tuning proposal distributions in MH algorithms for specific target distributions. An adaptive MCMC method is allowed to *learn* a better proposal based on the information provided by previous sample points. The first method of this kind is the Adaptive Proposal (AP) algorithm [13] and its generalization, the Adaptive Metropolis (AM) algorithm [14]. Both algorithms employ the Gaussian density as proposal distribution. Key to both schemes is the dynamic adaptation of mean and covariance of the proposal based on the accepted states of the chain. The AM algorithm is summarized in Alg. 1

Algorithm 1: The AM algorithm

Input: Initial $\mathbf{x}^{(0)}$, $\mathbf{m}^{(0)}$, $\mathbf{C}^{(0)}$, r
Result: Unbiased sample $\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(K)}$ from target distribution $\pi(\mathbf{x})$
for $g = 0, 1, K-1$ **do**

1. Sample $\mathbf{x}^{(g+1)} \sim \mathcal{N}\left(\mathbf{x}^{(g)}, r^2\mathbf{C}^{(g)}\right)$
2. Apply Metropolis criterion $\alpha_{\mathrm{MH}}(\mathbf{x}^{(g+1)}, \mathbf{x}^{(g)}) = \min\left(1, \frac{\pi(\mathbf{x}^{(g+1)})}{\pi(\mathbf{x}^{(g)})}\right)$
3. Update

$$\mathbf{m}^{(g+1)} = \mathbf{m}^{(g)} + \gamma^{(g+1)}\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)$$

$$\mathbf{C}^{(g+1)} = \mathbf{C}^{(g)} + \gamma^{(g+1)}\left(\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)^T - \mathbf{C}^{(g)}\right)$$

end

For adaptive samplers, a prerequisite for generating unbiased samples from the target distribution is the concept of *vanishing adaptation*. This means that the adaptation of first and second moments becomes smaller in the course of the sampling process. This is realized by the sequence of numbers $\gamma^{(g)}$. Andrieu and Thoms [15] provide mathematical conditions for the sequence $\gamma^{(g)}$ in order to achieve vanishing adaptation. For instance, sequences of the form $\gamma^{(g)} = \gamma_0/g^k$ with $k \in [(1+\epsilon)^{-1}, 1]$ with $\gamma_0, \epsilon > 0$ are consistent. In the original AM algorithm Haario and co-workers simply use $\gamma^{(g)} = 1/g$. Andrieu and Thoms also provide alternative update formulae for mean and covariance (see [15] for details). Moreover, they suggest a AM variant that adapts the scale factor r. They term this generic algorithm *the AM algorithm with global adaptive scaling* summarized in Alg. 2 The adaptation of the global scale factor r provides a means to control the acceptance rate of the sampler. P^* is the user-defined fixed target acceptance rate of the sampler and $\hat{\alpha}_{\mathrm{MH}}$ is the empirical acceptance probability. Gelman's optimal rate of 0.234 for Gaussian targets could be a default choice [16]. It is

Algorithm 2: Generalized AM algorithm with global adaptive scaling

Input: Initial $\mathbf{x}^{(0)}$, $\mathbf{m}^{(0)}$, $\mathbf{C}^{(0)}$, $r^{(0)}$ and P^*
Result: Unbiased sample $\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(K)}$ from target distribution $\pi(\mathbf{x})$
for $g = 0, 1, K-1$ **do**

1. Sample $\mathbf{x}^{(g+1)} \sim \mathcal{N}\left(\mathbf{x}^{(g)}, r^{(g)\,2}\mathbf{C}^{(g)}\right)$
2. Apply Metropolis criterion $\alpha_{\mathrm{M}}(\mathbf{x}^{(g+1)}, \mathbf{x}^{(g)}) = \min\left(1, \frac{\pi(\mathbf{x}^{(g+1)})}{\pi(\mathbf{x}^{(g)})}\right)$
3. Update

$$\log\left(r^{(g+1)}\right) = \log\left(r^{(g)}\right) + \gamma^{(g+1)}\left(\hat{\alpha}_{\mathrm{MH}}(\mathbf{x}^{(g+1)}, \mathbf{x}^{(g)}) - P^*\right)$$

$$\mathbf{m}^{(g+1)} = \mathbf{m}^{(g)} + \gamma^{(g+1)}\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)$$

$$\mathbf{C}^{(g+1)} = \mathbf{C}^{(g)} + \gamma^{(g+1)}\left(\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)\left(\mathbf{x}^{(g+1)} - \mathbf{m}^{(g)}\right)^T - \mathbf{C}^{(g)}\right)$$

end

amazing that this scheme is *exactly* the generic framework of GaA and CMA-ES-like optimization algorithms, yet in the context of black-box sampling. Note that update of the covariance and the global step size are (like in CMA-ES) not decoupled, thus hampering an efficient learning of the global scale in this context. As a remedy Andrieu and Thoms present variations of this scheme, such as, e.g., component-wise update of mean, covariance and scale factors, all of which have well-known equivalents in the optimization context since many years.

4 Unifying black-box optimization and sampling through Gaussian Adaptation

To the best of our knowledge, the close relationship between the presented black-box optimizers and adaptive MCMC methods has not been previously recognized. Gaussian Adaptation provides now a straight-forward means to combine black-box optimization and adaptive MCMC in a unifying framework. We present a GaA variant for black-box sampling that can be seen as a specific instance of an adaptive MCMC with global adaptive scaling [17]: the *Metropolis Gaussian Adaptation* (M-GaA).

4.1 Metropolis Gaussian Adaptation: an adaptive MCMC method

When using GaA for optimization, sample points with function values higher than the threshold c_{T} are strictly rejected and points with lower values accepted. For black-box sampling this hard threshold is now replaced by the Metropolis acceptance-rejection scheme $\alpha_{\mathrm{MH}}(\mathbf{x}^{(g+1)}, \mathbf{x}^{(g)})$. In cases where the continuous target probability distribution $\pi(\mathbf{x})$ is only known up to a normalization constant, $f(\mathbf{x}) \propto \pi(\mathbf{x})$ is used in the acceptance criterion. We further move GaA's

mean directly to the accepted sample $\mathbf{x}^{(g+1)}$. For the remaining parameters the standard settings are used. This yields a sampling algorithm with adaptive Gaussian proposals. Moreover, M-GaA possesses the convenient feature of setting the acceptance probability P *a priori*. The standard setting is $P = 0.234$. This renders M-GaA an adaptive MCMC sampler with global adaptive scaling *and* decoupling of covariance orientation and scale because the updated covariance is constantly normalized. Similar to the AP algorithm, M-GaA does not yet embed the concept of vanishing adaptation, leading to a scheme where ergodicity cannot be proven. It is, however, straightforward to include this concept into M-GaA. A detailed description of M-GaA can be found in [17].

4.2 Test scenarios for M-GaA

Different test scenarios for benchmarking adaptive MCMC algorithms are available in the literature. We highlight two such scenarios where M-GaA has been applied and compared to other samplers. One benchmark is based on the tests provided by Haario and co-workers for the AP and the AM algorithm [13, 14] and revisited in Andrieu and Thoms' tutorial [15]. The second benchmark is from Neal's article on Slice Sampling [18]. We describe these test scenarios here. Details about numerical results for M-GaA and alternative samplers are available in [17] and [19].

Haario's distributions. Following the protocol outlined in [13] the following three test target distributions are known to be suitable for adaptive MCMC methods:

π_1: Uncorrelated Gaussian distribution
π_2: Moderately twisted Gaussian distribution
π_3: Strongly twisted Gaussian distribution

Distribution π_1 is a centered n-dimensional multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C_1})$ with $\mathbf{C_1} = diag(100, 1, \ldots, 1)$. It thus has the shape of an axis-aligned hyper-ellipsoid with an axes aspect ratio of 10. The twisted Gaussians are constructed as follows: Let g be the density of π_1. The density function of a twisted Gaussian with twisting parameter $b > 0$ is then given by

$$g_b = g(\Phi_b(\mathbf{x}))\,, \tag{4}$$

where $\Phi_b(\mathbf{x}) = (x_1, x_2 + bx_1^2 - 100b, x_3, \ldots, x_n)$. Φ_b thus only affects the second coordinate, and the determinant of its Jacobian is unity [13]. It is easy to compute probability regions of g_b and to verify that the expectation value of g_b is $\mathbf{0}$ for all b. Haario et al. used $b = 0.03$ for π_2 and $b = 0.1$ for π_3. Fig. 2 shows the contour lines of the 68.3% and 99% probability regions of π_1 to π_3. Haario et al. also suggested the following quality measures for sampling algorithms:

1. mean($\|E\|$): The mean distance of the expectation values from their true value ($\mathbf{0}$), averaged over N repetitions

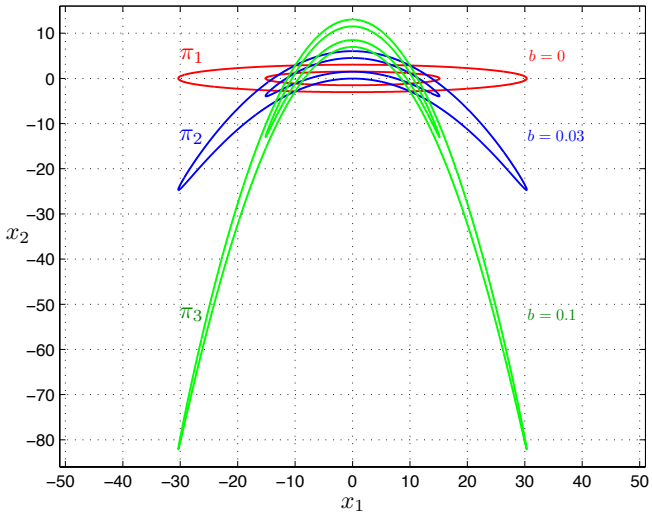

Fig. 2: 68.3% and 99% probability regions of the three test target distributions π_1 (red), π_2 (blue), and π_3 (green) in 2D. The parameter b controls the distortion of the Gaussian density (see main text for details).

2. std($\|E\|$): The standard deviation of the distance of the expectation values from their true value, averaged over N repetitions
3. err($\leq 68.3\%$): The mean error (in %) of the percentage of sampled points that hit the probability region inside the 68.3% contour
4. std($\leq 68.3\%$): The standard deviation of err($\leq 68.3\%$)
5. err($> 99\%$): The mean error (in %) of the percentage of sampled points that hit the probability region outside the 99% contour.
6. std($> 99\%$): The standard deviation of err($> 99\%$)

For these target distributions, adaptive MCMC algorithms like M-GaA show superior performance with respect to the presented quality measures. We refer to [13, 14, 15, 17] for further details.

Neal's funnel distribution. In his seminal paper on Slice Sampling [18], Neal introduced a funnel-shaped test distribution $\pi_{\text{f}}(\mathbf{x})$. It is a ten-dimensional distribution with variables $\mathbf{x} = (v, x_1, x_2, \ldots, x_9)$. The marginal distribution of v is $\mathcal{N}(0, 3^2)$. Conditional on a given value of v, the variables x_1 to x_9 are independent, with the conditional distribution for each being Gaussian with $\mathcal{N}(0, e^v)$. The resulting shape of the distribution resembles a single ten-dimensional funnel, with increasing values for v from one end to the other. A contour plot of $\pi_{\text{f}}(\mathbf{x})$ is provided in upper right part of Fig. 3. Neal states that "Such a distribution is typical of priors for components of Bayesian hierarchical models: x_1 to x_9 might, for example, be random effects for nine subjects, with v being the log of the variance of these random effects. If the data happens to be largely uninformative, the problem of sampling from the posterior will be similar to

that of sampling from the prior, so this test is relevant to actual Bayesian inference problems." A direct sampling method is straight-forward and consists of first sampling $v^{(g)}$ from $\mathcal{N}(0, 3^2)$ and then, conditionally sampling all $x_i^{(g)}$ from $\mathcal{N}(0, e^{v^{(g)}})$. A representative sample of size $2e4$ along with the corresponding marginal histograms for v and x_i are depicted in lower and left part of Fig. 3. For standard Metropolis-based schemes, the difficulty of sampling from $\pi_{\mathrm{f}}(\mathbf{x})$ is

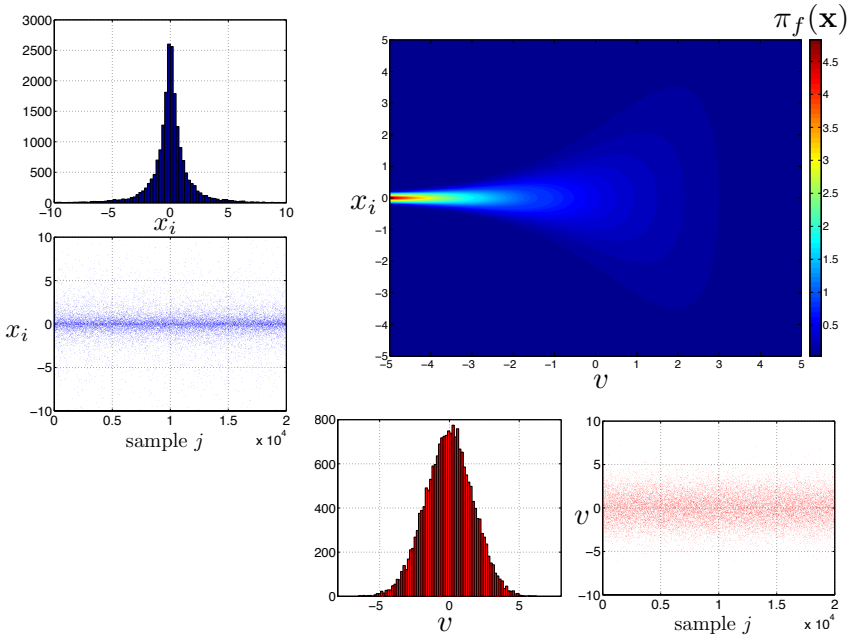

Fig. 3: Contour plot of the density $\pi_{\mathrm{f}}(\mathbf{x})$ of v and an arbitrary x_i in the top right corner. Direct samples from this distributions for v (red) and an arbitrary x_i (blue) are depicted in the lower right and upper left corner, respectively.

the low probability of accepting a proposal state when the chain is in a region of negative v, for instance $v^{(g)} = -4$. Conditional on this value, the variances of x_i would attain the tiny value of 0.018 leading to a highly peaked Gaussian. Hence, a MH algorithm with a standard multivariate Gaussian proposal $\mathcal{N}(\mathbf{0}, \mathbf{I})$ would, in the majority of cases, propose samples that are rejected. The chain gets stuck on the lower end of the funnel. The same reasoning also hinders the chain from visiting negative v values when started from positive values because these moves are almost always rejected. Hence, monitoring the $v^{(g)}$ values of the chain reveals the efficiency of the MCMC method.

This target distribution is challenging for adaptive MCMC methods. Numerical data for M-GaA are available in [19]. M-GaA can reproduce the correct

marginal for $v^{(g)}$ but diverges in all x_i. This indicates that the vanishing adaptation mechanism must be built into M-GaA in order to stabilize the covariance matrix estimation.

5 A unifying framework for black-box optimization and sampling as "grand challenge"

In this contribution we explored adaptive algorithms for black-box optimization and sampling. In identifying and analyzing the common design principles and features of CMA-ES, GaA and adaptive MCMC methods, we have been able to synthesize a novel adaptive MCMC method, M-GaA, that showed encouraging performance on the presented test problems. In our view, this is just a first step toward a unifying framework for adaptive black-box optimization and sampling. Exploring key ideas in both communities might be of mutual advantage. Besides our own work, a number of attempts have already been made in this direction. In statistics, Liang and Wong introduced an Evolutionary Monte Carlo scheme, a MCMC method that uses move sets from evolutionary algorithms [20, 21]. Several authors have recently refined such schemes [22, 23]. A combined approach to sampling and optimization is proposed in [24]. In the black-box optimization community, Vrugt and co-workers used the AM algorithm as part of an ensemble of search/sampling methods for improving multi-objective optimization [25].

We also argue that ergodicity proofs for the AM algorithm [14] and other adaptive MCMC schemes [26] may contain valuable ideas and techniques that can be exploited in convergence proofs of CMA-like schemes.

References

1. Hansen, N., Ostermeier, A.: Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. In: Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC '96). (1996) 312–317
2. Hansen, N., Ostermeier, A.: Completely Derandomized Self-Adaption in Evolution Strategies. Evolutionary Computation **9** (2001) 159–195
3. Hansen, N., Muller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evol. Comput. **11** (2003) 1–18
4. Kjellström, G.: Network Optimization by Random Variation of Component Values. Ericsson Technics **25** (1969) 133–151
5. Kjellström, G., Taxen, L.: Stochastic Optimization in System Design. IEEE Trans. Circ. and Syst. **28** (1981)
6. Kjellström, G.: On the Efficiency of Gaussian Adaptation. J. Optim. Theory Appl. **71** (1991)
7. Kjellström, G., Taxen, L.: Gaussian Adaptation, an evolution-based efficient global optimizer. In: Comp. Appl. Math., Elsevier Science (1992) 267—276
8. Müller, C.L., Sbalzarini, I.F.: Gaussian adaptation revisited - an entropic view on covariance matrix adaptation. In C. Di Chio et al., ed.: EvoApplications. Volume I of Lecture Notes in Computer Science., Springer (2010) 432–441

9. Igel, C., Suttorp, T., Hansen, N.: A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In: GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation, New York, NY, USA, ACM (2006) 453–460
10. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics **21** (1953) 1087–1092
11. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57** (1970) 97–109
12. Green, P., Han, X.: Metropolis methods, Gaussian proposals and antithetic variables. In Barone, P., Frigessi, A., Piccioni, M., eds.: Stochastic Models, Statistical Methods and Algorithms in Image Analysis., Berlin, Germany, Springer-Verlag (1992) 142–64
13. Haario, H., Saksman, E., Tamminen, J.: Adaptive proposal distribution for random walk Metropolis algorithm. Computational Statistics **14** (1999) 375–395
14. Haario, H., Saksman, E., Tamminen, J.: An adaptive Metropolis algorithm. Bernoulli **7** (2001) 223–242
15. Andrieu, C., Thoms, J.: A tutorial on adaptive MCMC. Statistics and Computing **18** (2008) 343–373
16. Gelman, A., Roberts, G., Gilks, W.: Efficient Metropolis jumping rules. In Bernado, J.M., et al., eds.: Bayesian Statistics. Volume 5. OUP (1996) 599
17. Müller, C.L., Sbalzarini, I.F.: Gaussian adaptation as a unifying framework for continuous black-box optimization and adaptive monte carlo sampling. In: Evolutionary Computation (CEC), 2010 IEEE Congress on. (2010) 1 –8
18. Neal, R.: Slice sampling. Annals of Statistics **31** (2003) 705–741
19. Müller, C.L.: Black-box Landscapes: Characterization, Optimization, Sampling, and Application to Geometric Configuration Problems. PhD thesis, Institute of Theoretical Computer Science, Department of Computer Science, ETH Zürich (2010)
20. Liang, F., Wong, W.: Real-Parameter Evolutionary Monte Carlo With Applications to Bayesian Mixture Models. Journal of the American Statistical Association **96** (2001) 653–666
21. Liang, F., Wong, W.: Evolutionary Monte Carlo for protein folding simulations. Journal of Chemical Physics **115** (2001) 3374–3380
22. Braak, C.J.: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing **16** (2006) 239–249
23. Hu, B., Tsui, K.W.: Distributed evolutionary Monte Carlo for Bayesian computing. Computational Statistics & Data Analysis (2008)
24. Ren, Y., Ding, Y., Liang, F.: Adaptive evolutionary Monte Carlo algorithm for optimization with applications to sensor placement problems. Statistics and Computing **18** (2008) 375–390
25. Vrugt, J.A., Robinson, B.A.: Improved evolutionary optimization from genetically adaptive multimethod search. Proceedings of the National Academy of Sciences of the United States of America **104** (2007) 708–711
26. Andrieu, C., Moulines, E.: On the ergodicity properties of some adaptive MCMC algorithms. Annals of Applied Probability **16** (2006) 1462–1505