

A NEW APPROACH TO GLOBAL OPTIMIZATION BY AN ADAPTED DIFFUSION

Oleg V. Poliannikov[†], Elena Zhizhina[‡], Hamid Krim^{}*

[†] University of Colorado at Denver, Department of Mathematical Sciences,
Campus Box 170, P.O. Box 173364, Denver, CO 80217-3364, USA

[‡] Institute for Information Transmission Problems of RAS,
Bol'shoy Karetnyi per. 19., Moscow 127994, Russia

^{*} Electrical and Computer Engineering Department, North Carolina State University,
Raleigh, NC 27695-7914, USA

ABSTRACT

In this paper, we study a problem of global optimization of an energy functional by a stochastic dynamics with a general diffusion coefficient. The main result is that adapting the diffusion coefficient to the shape of the functional enables the dynamics to escape wide local minima, and attracts it to narrower global minima that are missed by conventional diffusions. We discuss how to properly choose the diffusion coefficient and show numerically the superior performance of the resulting optimization algorithm.

Index Terms— Optimization methods, diffusion equations, Markov processes

1. INTRODUCTION

Gibbs field based stochastic methods have long been recognized as an effective approach to problems of global optimization, see for instance [1, 2, 3, 4, 5, 6]. Their essence can be summarized in the following way. Consider an equilibrium stochastic dynamics with a stationary Gibbs measure, where the latter is associated with some energy functional H . The dynamics is then changed so that it is no longer in equilibrium, and its limit distribution is concentrated on the set of global minima of H . This approach to optimization is generally called simulated annealing.

More precisely, consider a stochastic diffusion dynamics, whose invariant measure is given by

$$d\mu(x) = \frac{1}{Z} e^{-\beta H(x)} dm(x), \quad (1)$$

where $X = [a, b]^d \subset \mathbb{R}^d$, $H : X \rightarrow \mathbb{R}$ is the energy functional, and m - normalized Lebesgue measure on X . We seek to find the set of global minima of H . The classical technique of solving this problem is to stochastically perturb the deterministic gradient descent:

$$dX(t) = -\nabla H(X(t)) dt + \sigma(t) dW(t), \quad X(0) = X_0, \quad (2)$$

where ∇ denotes the spatial gradient, $W(t)$ is a realization of the standard Brownian motion, and X_0 is the starting point.

It is known that in order for the limiting measure to concentrate on the set of global minima of H , the diffusion coefficient should slowly decay to zero. We conventionally refer to the behavior of the function σ as the cooling schedule of our dynamics. We call this standard dynamics (spatially) homogeneous because the diffusion coefficient does not depend on the state variable $X(t)$.

In this paper, we propose a new diffusion process whose distinguishing feature is a spatially inhomogeneous diffusion coefficient. It is important that the stationary Gibbs distribution of the newly introduced dynamics be identical to that of the homogeneous diffusion. It is, however, shown that by appropriately constructing the inhomogeneous diffusion, one can improve the speed of convergence of the overall dynamics to the stationary distribution.

The inhomogeneous diffusion coefficient that leads to the optimal speed of convergence depends on the functional H at hand. Its exact form for a general H continues to be an open problem. Of particular interest in many applications including signal and image filtering, is a situation where the global minimum of H is so narrow that a standard diffusion tends to overlook it. We demonstrate that it is possible to adapt the diffusion to the cost functional and to hence alleviate this problem. The performance of the adapted diffusion is shown to offer superior performance in comparison to its classical counterpart.

This paper is organized as follows. In the second section we formulate the problem mathematically and state results on stochastic dynamics and its approximations by non-homogeneous Markov chains. In the third section we compare the convergence properties of the modified diffusions and of the Langevin dynamics using numerical simulations. In the fourth section, we conclude and discuss future work.

2. PROBLEM STATEMENT AND THEORETICAL RESULTS

Let $X = [-R, R] \subset \mathbb{R}$, and denote by $F(X)$ a space of functions on X with periodical boundary conditions. Let $H : X \rightarrow \mathbb{R}$ be a function bounded from below. For definitiveness, assume that

$$H(x) \geq 0, \forall x \in [-R, R], \min_{x \in [-R, R]} H(x) = 0. \quad (3)$$

Proposition 1 ([7]). *Consider a continuous time Markov process on X whose infinitesimal generator \mathcal{L}_a is given by*

$$\begin{aligned} \mathcal{L}_a f(x) &= \frac{1}{2} \delta(a) \omega(x) f''(x) \\ &- \left(\omega(x) H'(x) - \frac{1}{2} \delta(a) \omega'(x) \right) f'(x), \quad f \in F(X). \end{aligned} \quad (4)$$

Then the stationary distribution $\tilde{\gamma}^a$ has a Gibbs density, i.e.

$$d\tilde{\gamma}^a(x) = \frac{e^{-\frac{2H(x)}{\delta(a)}}}{Z_{\delta(a)}} dm(x), \quad (5)$$

where $Z_{\delta(a)}$ is a normalization constant, and m is the normalized Lebesgue measure (length) on X .

Consider an approximation in time of the above diffusion process. It is a Markov chain X_n given by

$$\begin{aligned} X_{n+1} = X_n - \left(\omega(X_n) H'(X_n) - \frac{1}{2} \delta(a) \omega'(X_n) \right) a \\ + \sqrt{a \delta(a) \omega(X_n)} W_n, \end{aligned} \quad (6)$$

where $a, \delta(a) > 0$ are parameters, $\omega(x)$ is a fixed nonnegative smooth function and (W_n) is an i.i.d random sequence, and $\mathbb{E}[W_n] = 0, \text{var}[W_n] = 1$.

Proposition 2 ([8]). *For any $a, \delta(a) > 0$ Markov chain (6) has a stationary distribution, which will be denoted γ^a .*

We now state the main result that shows that as $a \rightarrow 0$, the stationary measure γ^a becomes concentrated on the set of global minima of H . Toward that end, we introduce the following notation. We denote by $\mathcal{O}_\varepsilon(x)$ the ε -neighborhood of any $x \in X$, and by $\mathcal{U}_\varepsilon(H)$ the union of ε -neighborhoods of all global minima of H :

$$\mathcal{O}_\varepsilon(x) = X \cap (x - \varepsilon, x + \varepsilon), \quad \mathcal{U}_\varepsilon(H) = \bigcup_{x: H(x)=0} \mathcal{O}_\varepsilon(x). \quad (7)$$

Theorem 1 ([7]). *Let $\varepsilon > 0$ and assume the existence of $C_0 > 0$, such that $\delta(a) \geq -\frac{C_0}{\ln a}$. Then $\gamma^a(\mathcal{U}_\varepsilon(H)) \rightarrow 1$ as $a \rightarrow 0$ and $\delta(a) \rightarrow 0$.*

The analysis of the convergence along with the form of the modified drift suggest that the function ω could be chosen inversely proportional to H' :

$$\begin{aligned} \omega(X) &\equiv \omega(H'(X)) \\ &= \begin{cases} 1/|H'(X)|, & |H'(X)| > k \\ 1/k, & |H'(X)| \leq k, \end{cases} \end{aligned} \quad (8)$$

where the suitable choice of parameter k ensures the stability of the numerical algorithm in the neighborhood of local minima, where $H'(X) \approx 0$. We demonstrate in the following section that this choice of ω leads to a faster convergence and to hence better optimized results.

3. NUMERICAL SIMULATIONS

In this section we present simulation results, which demonstrate the performance of the newly proposed modified diffusion versus the standard dynamics. In both simulation cases we use the same cooling schedule:

$$a_n = a_0 e^{-Kn}, \quad (9)$$

where $K > 0, K \ll 1$, e.g. $K = 0.1$. Functions H are chosen such that its global minima are narrow relative to other local minima, which is one of the most challenging situations in applications. We observe that by the choice of ω in (8), wide minima are naturally disfavored, as the jump size is large in such areas. Points then tend to concentrate in the vicinities of the global minima, where the value of H' is larger, and hence the random jumps are naturally suppressed.

3.1. 1D example

Consider the function

$$H(X) = X^2 \cos 10X, \quad X \in [-\pi, \pi], \quad X_0 = 0. \quad (10)$$

We let 100 points start from X_0 and evolve each according to its own realization of the diffusion process. We note that for smaller a_0 , the classical diffusion fails to leave local minima, while higher values of a_0 result in uniform coverage of the entire domain, similar to the high-temperature regime. The modified diffusion provides, as expected, a great improvement at locating the global minima (see Figures 1-3).

3.2. 2D example

We now consider a two dimensional example

$$H(X, \theta) = \beta \Phi_1(X, \theta) + (1 - \beta) \Phi_2(X), \quad (11)$$

where Φ_1 and Φ_2 are ‘‘data fidelity’’ and ‘‘smoothness’’ terms correspondingly:

$$\begin{aligned} \Phi_1(X, \theta) &= \|X - \theta\|^2, \\ \Phi_2(X) &= -\frac{1}{1 + \frac{1}{d^2}(X_1 - X_2)^2}, \end{aligned} \quad (12)$$

$$X = (X_1, X_2)^\top$$

with $\theta \in (-5, 5)^\top, X_0 = \theta, \beta = 0.01, d = 0.5, a_0 = 0.5$.

We observe (see Figures 4-6) that as in previous examples, particles tend to cluster around the global minimum under the modified diffusion, whereas they hover around a local minimum for a small discretization step, and spread over the entire computation domain if the step is too large.

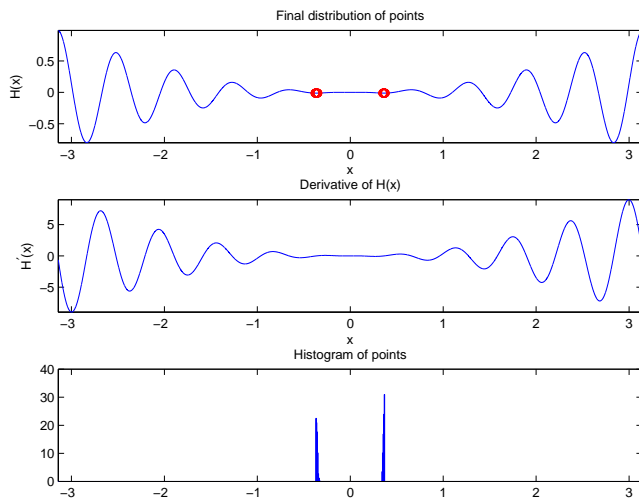


Fig. 1. Classical 1D diffusion: $X_0 = 0, a_0 = 0.5$

4. CONCLUSIONS

Problems of global optimization are of great theoretical as well as practical importance. Gibbs fields-based methods have a tremendous potential because of their tractability and ease of implementation. The main shortcoming of these methods is the slow speed of convergence to the global minimum.

In this paper we have considered diffusion dynamics with a general diffusion coefficient. We have shown that the speed of convergence of a particle to a global minimum may be improved by choosing the diffusion coefficient adapted to a particular functional H . The resulting dynamics then outperforms its classical counterpart at finding the global minimum of an energy functional.

We refer to our proposed approach as *adapted diffusion based algorithm* since the choice of the function $\omega(x)$ depends on the energy function H . The main goal of this paper is to show that we can construct a non-homogeneous diffusion which escapes wide local minima and stays at deep and tight minima of the energy function H . When solving examples where the global minimum is very narrow, we proceed by cooling the dynamics where the value of H' is large and heating it up when $H' \approx 0$. In so doing, we force the diffusion to favor the global minimum and reject other local extrema.

As may be seen from demonstrated simulations, our proposed algorithm reveals deep and tight minima, whereas the classical diffusion escapes these minima for a short time and prefers to stay at local wide minima. That is an important property, which is desirable in global optimization problems for functionals with deep and tight minima.

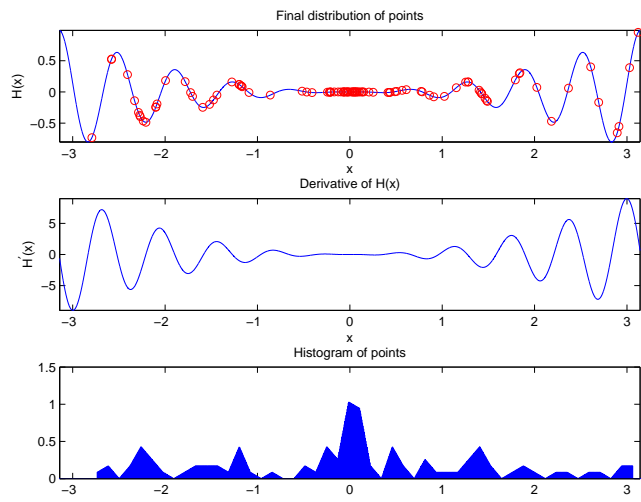


Fig. 2. Classical 1D diffusion: $X_0 = 0, a_0 = 50$

5. REFERENCES

- [1] S. Geman and D. Geman, “Stochastic relaxation, gibbs distribution, and the bayesian restoration of images,” *PAMI*, vol. 6, no. 6, pp. 721–741, 1984.
- [2] S. Geman and G. Reynolds, “Constrained restoration and recovery of discontinuities,” *PAMI*, vol. 14, no. 3, pp. 367–383, 1992.
- [3] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, pp. 671–680, 1983.
- [4] S. Geman and G. Hwang, “Diffusion for global optimization,” *SIAM Journal of Control and Optimization*, vol. 24, pp. 1031–1043, 1986.
- [5] T.-S. Chiang, C.-R. Hwang, and S.-J. Sheu, “Diffusion for global optimization in r^n ,” *SIAM Journal of Control and Optimization*, vol. 25, pp. 737–753, 1987.
- [6] G. Winkler, *Image analysis, random fields and Markov chain Monte Carlo methods, A mathematical introduction*, Springer, New York, 2003.
- [7] O.V. Poliannikov, E.Zhizhina, and H. Krim, “Global optimization by modified diffusion,” Submitted for publication.
- [8] T.M. Liggett, *Interacting Particle Systems*, Springer, 2005.

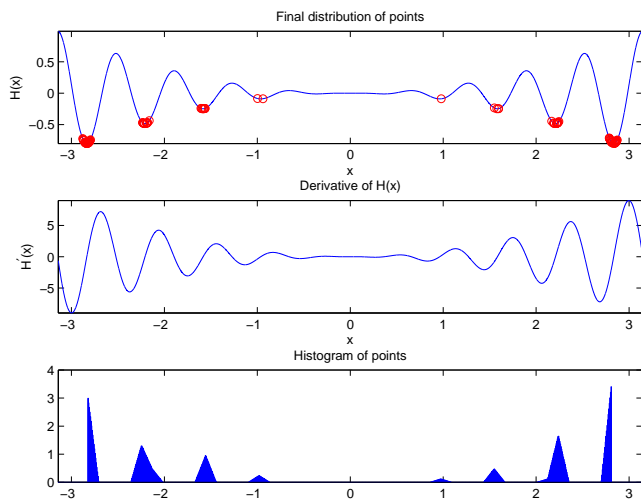


Fig. 3. Modified 1D diffusion: $X_0 = 0, a_0 = 0.5$

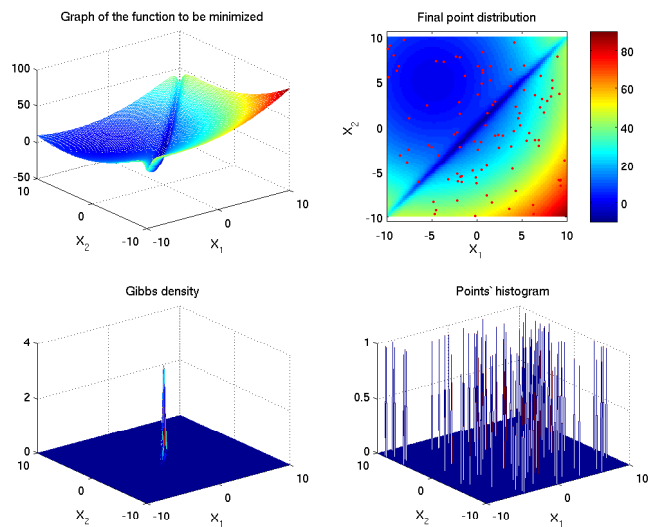


Fig. 5. Classical 2D diffusion: $X_0 = (-5, 5)^\top, a_0 = 20$

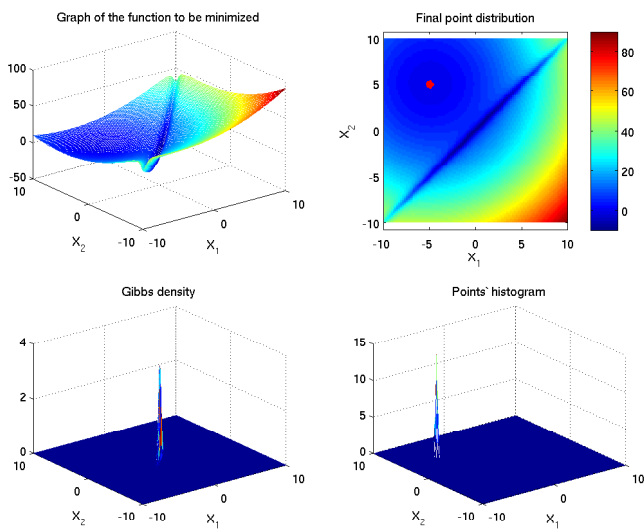


Fig. 4. Classical 2D diffusion: $X_0 = (-5, 5)^\top, a_0 = 0.9$

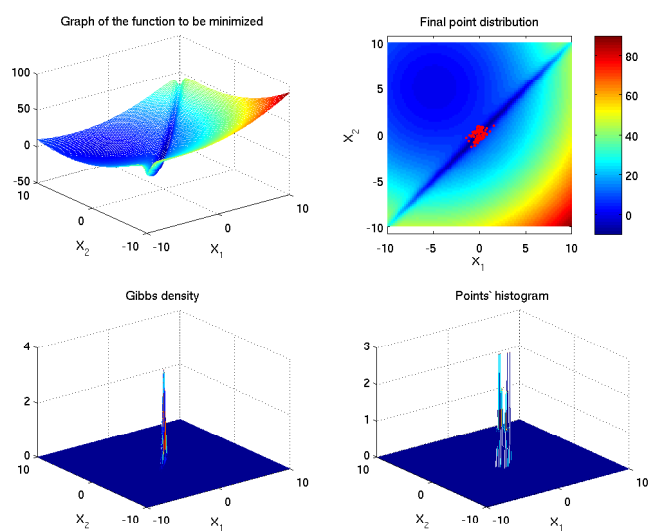


Fig. 6. Modified 2D diffusion: $X_0 = (-5, 5)^\top, a_0 = 0.5$