Global Optimization by Adapted Diffusion

Global Optimization by Adapted Diffusion

Oleg V. Poliannikov, Elena Zhizhina, and Hamid Krim, *Fellow, IEEE*

***Abstract*—In this paper, we study a diffusion stochastic dynamics with a general diffusion coefficient. The main result is that adapting the diffusion coefficient to the Hamiltonian allows to escape local wide minima and to speed up the convergence of the dynamics to the global minima. We prove the convergence of the invariant measure of the modified dynamics to a measure concentrated on the set of global minima and show how to choose a diffusion coefficient for a certain class of Hamiltonians.**

***Index Terms*—Nonlinear systems, optimization methods, simulated annealing, stochastic fields.**

I. Introduction

Gibbs field based stochastic methods have long been recognized as an effective approach to solve problems of global optimization, see for instance [1]–[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

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

where $\mathsf{X} = [a,b]^d \subset \mathbb{R}^d$, $H : \mathsf{X} \to \mathbb{R}$ 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

$$dX(t) = -H'(X(t))dt + \sigma(t)dW(t) \tag{2}$$

where $W(t)$ is a realization of the standard Brownian motion (see, for example, [4] and [5]). Then, following the simulated annealing regime, 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 [4], [5] 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 distinguished 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. We prove that the order of the speed of convergence cannot be improved on, but the corresponding coefficient can be in principle chosen optimally.

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 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.

These problems may arise when the cost functional consists of two terms. The first data fidelity term smoothly penalizes deviations from the given data [7]. The second smoothness term defines a relatively small subspace, to which the solution is attracted (but does not have to belong). The convex combination of these terms often results in a functional with narrow global minimum (see Section IV-A for an illustration). Alternatively, consider a system identification problem where the goal is to recover the coefficients of an unknown IIR filter based on the observed output. The coefficients are usually found by minimizing the mismatch between the synthetic response of test filters and the data. The resulting multidimensional cost functionals are multimodal, and the optimal global minimum is relatively narrow as compared to other local minima (see Section IV-C).

This paper is organized as follows. In Section II, we formulate results on stochastic dynamics and its approximations, which are non-homogeneous Markov chains. In Section III, we discuss how to choose a modified diffusion so as to adapt it to a particular form of the cost functional. In Section IV, we analyze and compare the convergence properties of the modified diffusions and of the Langevin dynamics using numerical simulations. The conclusions and description of possible extensions are deferred to Section V. Finally, the Appendix contains the proof of the main result.

Manuscript received November 22, 2009; accepted August 15, 2010. Date of publication September 02, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Christophe Pesquet. This work was also supported by the U.S. Air Force Office of Scientific Research under Grant FA 9550-07-1-0104.

O. V. Poliannikov is with the Earth Resources Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: poliann@mit.edu).

E. Zhizhina is with the Dobrushin Laboratory, Institute for Information Transmission Problems, Moscow GSP-4, 127994, Russia (e-mail: ejj@iitp.ru).

H. Krim is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: ahk@ncsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2071867

II. Theoretical Results

Let $\mathsf{X} = [-R, R] \subset \mathbb{R}$, and denote by $F(\mathsf{X})$ a space of sufficiently smooth functions on X with matching (periodic) boundary conditions. Let $H : \mathsf{X} \to \mathbb{R}$ be a smooth function bounded from below. For specificity, assume that

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

Consider a continuous time Markov process on X whose infinitesimal generator L_a is given by

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

As follows from the formula for diffusions generated by second-order differential operators of the general form [8], this process is also the solution to the diffusion equation

$$dX(t) = -\left[\omega(X(t))H'(X(t)) - \frac{1}{2}\delta(a)\omega'(X(t))\right] dt + \sqrt{\delta(a)\omega(X(t))}\, dW(t). \tag{5}$$

The function ω in (5) is assumed fixed, nonnegative and smooth. Below we will describe how by choosing a specific form of ω, we can control the behavior of the dynamic to take advantage of known features of H. The positive scalar $\delta(a) > 0$ is the annealing parameter called "temperature." As in the case of the conventional annealing, it will slowly decay to zero. The meaning of parameter a will become obvious shortly.

Proposition 1: If the process defined by (4) has a unique stationary distribution $\tilde{\gamma}^a$, then the latter has a Gibbs density, i.e.,

$$d\tilde{\gamma}^a(x) = \frac{e^{-2H(x)/\delta(a)}}{Z_{\delta(a)}}\, dm(x) \tag{6}$$

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

Proof: Follows from the equality

$$\int \mathsf{L}_a f(x) d\tilde{\gamma}^a(x) = 0 \tag{7}$$

for any $f \in D(\mathsf{L}_a)$. ■

The diffusion process that corresponds to the infinitesimal generator (4) is written as an approximation in time of the diffusion process (5). It is a Markov chain X_n given by

$$X_{n+1} = X_n - \left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a + \sqrt{a\delta(a)\omega(X_n)}\, W_n \tag{8}$$

where $a > 0$ is the discretization step, and (W_n) is an i.i.d. random sequence, and $\mathbb{E}[W_n] = 0$, $\mathrm{var}[W_n] = 1$.

Proposition 2: For any $a > 0$ such that $\delta(a) > 0$, Markov chain (8) has a stationary distribution, which will be denoted γ^a.

Proof: Follows from general facts concerning Markov processes on compact spaces. See, for example, [9, Ch. 1]. ■

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

$$\mathsf{O}_\varepsilon(x) = \mathsf{X} \cap (x - \varepsilon, x + \varepsilon),\ \mathsf{U}_\varepsilon(H) = \bigcup_{x: H(x)=0} \mathsf{O}_\varepsilon(x). \tag{9}$$

Theorem 1: For an arbitrary $\varepsilon > 0$, assume the existence of $C_0 > 0$, such that $\delta(a) \geq -C_0/\ln a$. Then $\gamma^a(\mathsf{U}_\varepsilon(H)) \to 1$ as $a \to 0$ and $\delta(a) \to 0$.

Proof: See the Appendix. ■

III. Speed of Convergence and Modified Diffusion Coefficient

It is seen from (37)–(39) that the rate of convergence of the modified diffusion remains similar, since the parameter $\delta(a)$ in the cooling procedure decreases in the same way as for the classical dynamic. The speed of convergence, however, may be improved by minimizing the coefficients in (37) and (38). We note that if $\tilde{X}_n \approx X_n$, the coefficients in (37) and (38) have the same expression:

$$K(\omega) \equiv \mathbb{E}_{\gamma^a}\left[\omega(X_n)\ (H'(X_n))^3\right]. \tag{10}$$

This together with the form of the modified drift suggest that the function ω could be chosen inversely proportional to H':

$$\omega(X) \equiv \omega\left(H'(X)\right) = \begin{cases} \frac{1}{|H'(X)|}, & |H'(X)| > k \\ \frac{1}{k}, & |H'(X)| \leq k \end{cases} \tag{11}$$

where the suitable choice of parameter k ensures the stability of the numerical algorithm in the neighborhood of local mimima, where $H'(X) \approx 0$.

The diffusion coefficient so constructed suppresses random jumps when the gradient of the cost function is large, and it reinforces them when the gradient is small. As a result, the process naturally explores narrow steep cavities of the cost functional in more detail than the standard (homogeneous) diffusion would. In the next section, we show that this results in far superior performance of the optimization algorithm for the class of cost functionals under consideration.

IV. 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:

$$a_n = a_0 e^{-Kn}, \quad n \in \{1, \ldots, N\} \tag{12}$$

where $K > 0$, $K \ll 1$, and N is the total number of iterations. Note that the sequence $\{a_n\}$ decays monotonously. The parameter K controls the rate of decay, and it should be sufficiently small as to allow proper mixing in the sense of Proposition 2 at each temperature level, while being large enough to ensure effective convergence to zero within N iterations. The total number of iterations N is the same for both diffusions. 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

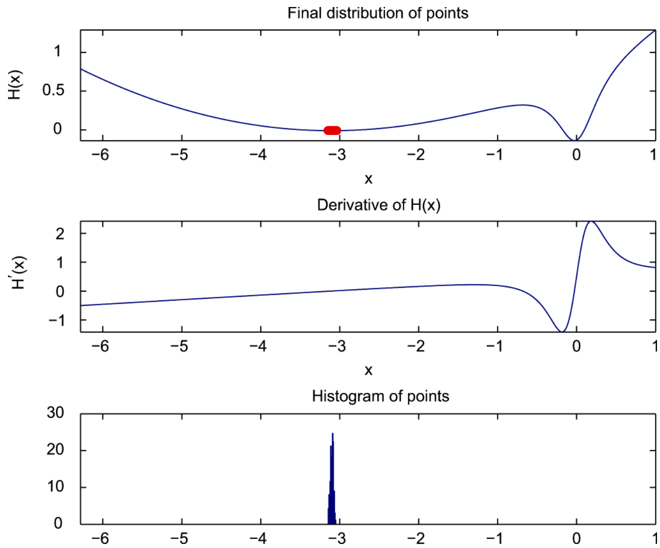


Fig. 1. Classical diffusion: $H(X) = H_1(X), X_0 = -3, a_0 = 0.5$.

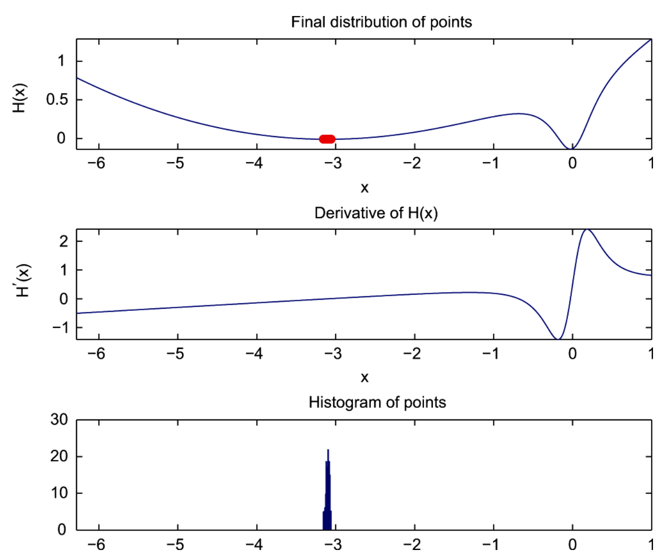


Fig. 2. Classical diffusion: $H(X) = H_1(X), X_0 = -3, a_0 = 1$.

(11), 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.

A. Example 1

Consider the function

$$H_1(X) = \alpha(X+\pi)^2 - (1-\alpha)\frac{1}{1+\frac{X^2}{d}} \tag{13}$$

where $\mathsf{X} = [-2\pi, \pi]$, $\alpha = 0.08$ and $d = 0.01$. Let the initial point be $X_0 = -3$. We let 100 points start from X_0 and evolve each according to its own realization of the diffusion process with the cooling schedule (12), where $K = 10^{-5}$ and $N = 2 \cdot 10^5$. Our simulations show that the particles first spread in a near uniform fashion all over the domain, and then tend to come back guided by the shape of H to the local minimum where they started (Figs. 1 and 2).

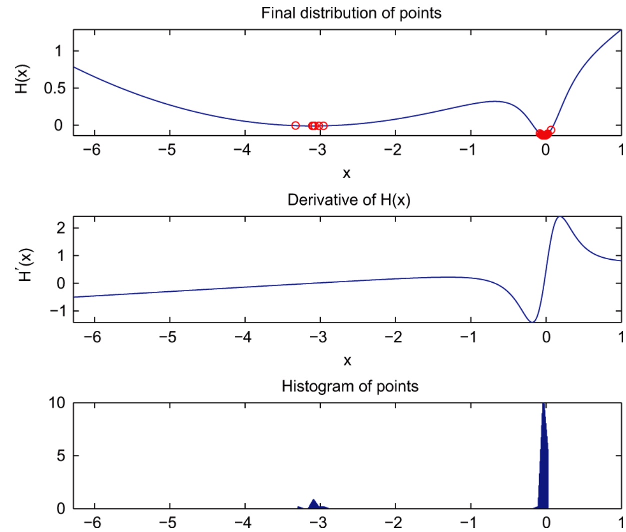


Fig. 3. Modified diffusion: $H(X) = H_1(X), X_0 = -3, a_0 = 0.5$.

The modified diffusion on the other hand, forces a particle to leave the neighborhood of the starting point but not from that of the global minimum (Fig. 3).

B. Example 2

Consider now

$$H_2(X) = X^2 \cos 10X, \quad X \in [-\pi, \pi], \quad X_0 = 0. \tag{14}$$

Again 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 as expected does a far better job at finding the global minima (see Figs. 4–6).

C. Example 3

The main results in this paper are stated and proven for the one-dimensional case. Much like in [10], generalizing this proof to higher dimensions is very non-trivial. However, as long as we are presented with a qualitatively similar multidimensional problem, we expect the proposed approach to work well. Here, we demonstrate the performance of our algorithm for a real problem of system identification. An IIR filter is governed by the input–output relation

$$y(k) + \sum_{j=1}^{M} \beta_j y(k-j) = \sum_{i=0}^{L} \alpha_i x(k-i). \tag{15}$$

The IIR filter design problem consists of recovering the filter's unknown coefficients from its observed response $d(k)$. This is done by minimizing the misfit between $d(k)$ and the projected response of a test filter. Specifically, we seek to find the global minimum of the cost functional H given by

$$\min_{\{\alpha_i, \beta_j\}} \left\{ H(\alpha_i, \beta_j) = \mathbb{E}\left[(d(k) - y(k))^2\right] \right\}. \tag{16}$$

The expectation is computed over all possible random input realizations. The functional H in the left-hand side of (16) depends on several variables and it is typically multimodal. A

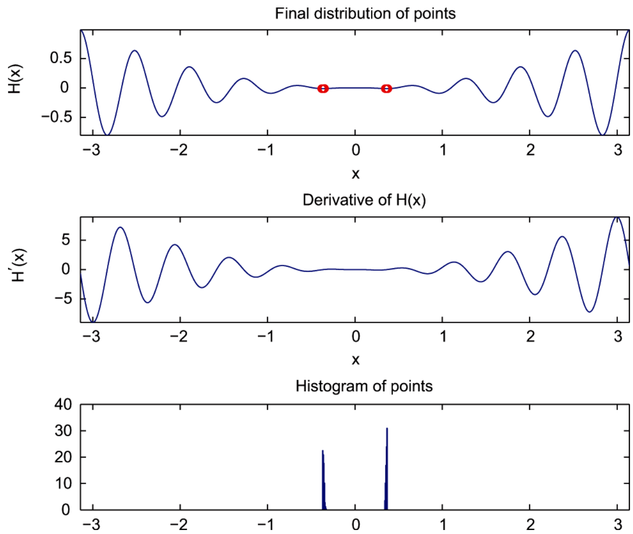


Fig. 4. Classical diffusion: $H(X) = H_2(X), X_0 = 0, a_0 = 0.5$.

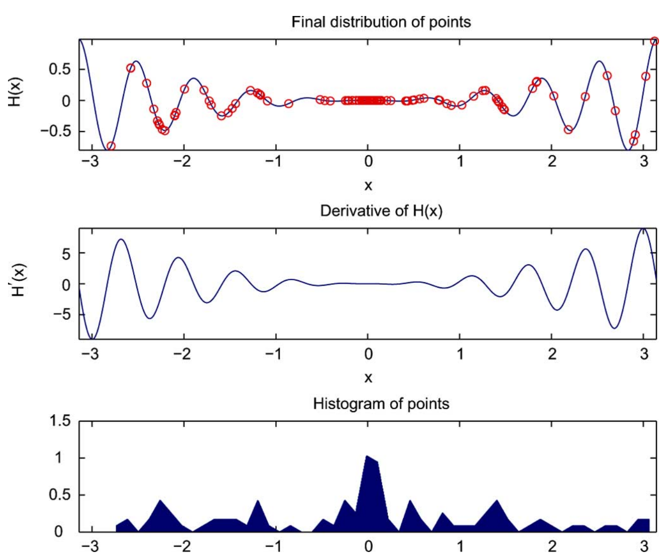


Fig. 5. Classical diffusion: $H(X) = H_2(X), X_0 = 0, a_0 = 50$.

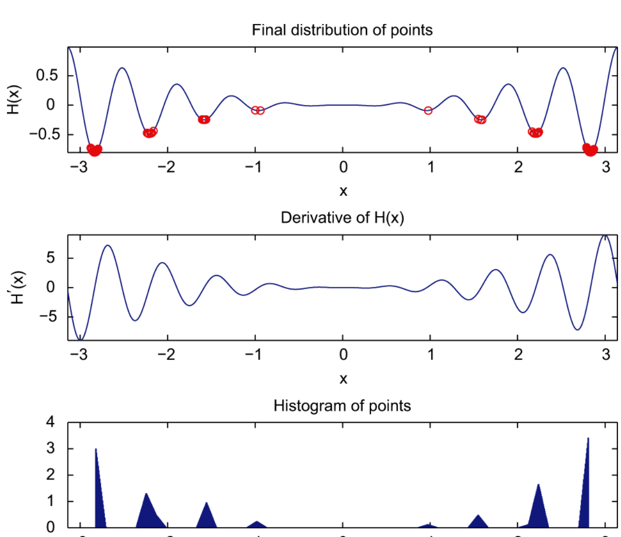


Fig. 6. Modified diffusion: $H(X) = H_2(X), X_0 = 0, a_0 = 0.5$.

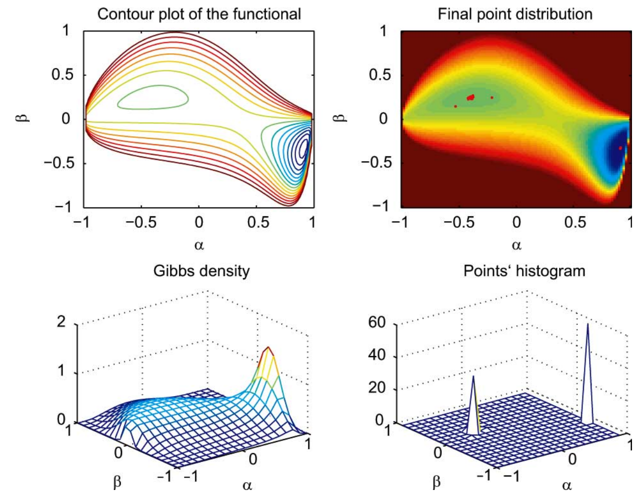


Fig. 7. Modified diffusion: $H(\alpha, \beta)$ is as in (19), $(\alpha_0, \beta_0) = (-0.4, 0.25)$, $a_0 = 1$.

global minimizer of the sort that is proposed is required to find the optimal solution. Assume for specificity that the observed response is generated by an (unknown) second-order system with the transfer function given by

$$H_*^0(z) = \frac{0.2 - 0.6z^{-1}}{1 - 1.1314z^{-1} + 0.25z^{-2}}. \tag{17}$$

We will search for a first-order filter

$$H_*(z) = \frac{\beta}{1 - \alpha z^{-1}} \tag{18}$$

whose response best matches the observed output of the system. The precise analytic form for the cost functional H is known [11] and written as

$$H(\alpha, \beta) = 1 - 2\beta\, H_*^0\left(\alpha^{-1}\right) + \frac{\beta^2}{1 - \alpha^2}. \tag{19}$$

Its contour plot is shown in Fig. 7. We note that this function has one wider local minimum and another steeper global minimum. We apply the standard and modified diffusion to the same function with $a_0 = 1.0$, $K = 3 \cdot 10^{-4}$, $N = 10^5$ and observe that as in previous examples, particles tend to cluster around the global minimum under the modified diffusion (Fig. 7), whereas they hover around a local minimum for a small discretization step in the case of the standard one (Fig. 8).

V. Conclusion

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 while it is impossible to improve the order of the speed of convergence, the speed of convergence may be improved by choosing the diffusion coefficient adapted to a particular functional H. The resulting dynamics contains the classical diffusion as a particular

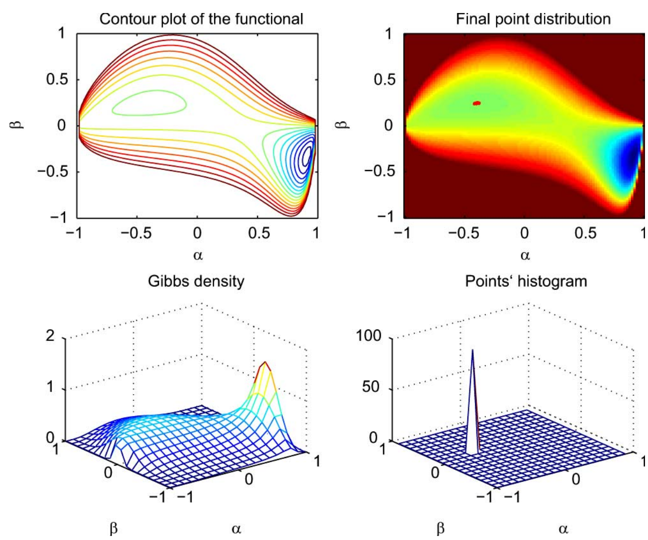


Fig. 8. Classical diffusion: $H(\alpha,\beta)$ is as in (19), $(\alpha_0,\beta_0)=(-0.4,0.25)$, $a_0=1$.

case, and outperforms the latter at finding the global minimum of an energy functional to which it is adapted.

We refer to our proposed approach as the *adapted diffusion based algorithm* since the choice of the function $\omega(x)$ depends on the energy function H. In this case, the diffusion coefficient is non-homogeneous and is determined by the remainder terms (37) and (38). We do not discuss here how to find an optimal $\omega(x)$ for an arbitrary energy function H as that remains to be a challenging problem. The main goal for us is to show that we can construct an important in applications 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 above 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. As the two diffusions are run for exactly the same number of iterations, the added computational cost amounts to evaluating the function ω, which is negligible.

APPENDIX
PROOF OF THEOREM 1

The main line of the proof follows reasoning from [10].

Fix $\varepsilon>0$ and consider two smooth functions ϕ_1 and ϕ_2 satisfying

$$\phi_1(x)=\begin{cases}1, & H(x)>3\varepsilon\\ \alpha_1(x), & H(x)\in[2\varepsilon,3\varepsilon]\\ 0, & H(x)<2\varepsilon\end{cases} \tag{20}$$

and

$$\phi_2(x)=\begin{cases}1, & H(x)<\varepsilon\\ \alpha_2(x), & H(x)\in[\varepsilon,3\varepsilon]\\ 0, & H(x)>3\varepsilon\end{cases} \tag{21}$$

where $0\le\alpha_1(x),\ \alpha_2(x)\le 1$. Construct a function

$$\phi_a(x)=\phi_1(x)-\eta(a)\phi_2(x) \tag{22}$$

such that

$$\int\phi_a(x)d\tilde{\gamma}^a(x)=0. \tag{23}$$

Since H is assumed to be smooth, one can choose the functions α_1 and α_2 such that ϕ_a is differentiable everywhere. Clearly,

$$\eta(a)=\frac{\int\phi_1(x)e^{-2H(x)/\delta(a)}\,dm(x)}{\int\phi_2(x)e^{-2H(x)/\delta(a)}\,dm(x)} \tag{24}$$

and because we can bound the numerator and denominator:

$$\begin{aligned}\int\phi_1(x)e^{-2H(x)/\delta(a)}\,dm(x)&\le e^{-4\varepsilon/\delta(a)}\\ \int\phi_2(x)e^{-2H(x)/\delta(a)}\,dm(x)&\ge m_\varepsilon e^{-2\varepsilon/\delta(a)}\end{aligned} \tag{25}$$

with $m_\varepsilon\equiv m\big(\{x: H(x)\le\varepsilon\}\big)$, we have

$$\eta(a)\le\frac{e^{-4\varepsilon/\delta(a)}}{m_\varepsilon e^{-2\varepsilon/\delta(a)}}=\frac{1}{m_\varepsilon}\,e^{-2\varepsilon/\delta(a)}\to 0 \tag{26}$$

when $\delta(a)\to 0$.

Consider the equation

$$\mathsf{L}_a\psi_a=\phi_a. \tag{27}$$

Because of (23) and (7), (27) has a solution. Expanding this equation, we get

$$\frac{1}{2}\delta(a)\omega(x)\psi_a''-\left(\omega(x)H'-\frac{1}{2}\delta(a)\omega'\right)\psi_a'=\phi_a(x) \tag{28}$$

which is solvable for ψ_a', and therefore

$$\psi_a'(x)=\frac{e^{2H(x)/\delta(a)}}{\omega(x)}\int_{-R}^{x}\frac{2}{\delta(a)}\phi_a(u)e^{-2H(u)/\delta(a)}\,du \tag{29}$$

and

$$\begin{aligned}\psi_a''(x)=&\frac{2}{\delta(a)}\frac{1}{\omega(x)}\phi_a(x)\\ &+\frac{\frac{2}{\delta(a)}\omega(x)H'(x)-\omega'(x)}{\omega^2(x)}e^{2H(x)/\delta(a)}\int_{-R}^{x}\frac{2}{\delta(a)}\phi_a(u)\\ &\times e^{-2H(u)/\delta(a)}\,du.\end{aligned} \tag{30}$$

Differentiating once again we get to the leading order as $\delta(a) \to 0$:

$$\psi_a'''(x) = \left(\frac{2}{\delta(a)}\right)^3 \frac{(H'(x))^2}{\omega(x)} e^{2H(x)/\delta(a)} \times \int_{-R}^{x} \phi_a(u) e^{-2H(u)/\delta(a)} du + O\left(\frac{1}{\delta^2(a)} e^{2H(x)/\delta(a)}\right). \quad (31)$$

Let

$$\psi_a(x) = \int_{-R}^{x} \psi_a'(u) du.$$

Now with the help of (8) we construct a Taylor–Lagrange expansion of $\psi_a(X_{n+1})$:

$$\begin{aligned}\psi_a(X_{n+1}) = \psi_a(X_n) &- \left(\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a - \sqrt{a\delta(a)\omega(X_n)}\, W_n\right) \psi_a'(X_n) \\ &+ \frac{1}{2}\left(\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a - \sqrt{a\delta(a)\omega(X_n)}\, W_n\right)^2 \psi_a''(X_n) \\ &- \frac{1}{6}\left(\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a - \sqrt{a\delta(a)\omega(X_n)}\, W_n\right)^3 \psi_a'''(\tilde{X}_n) \end{aligned} \quad (32)$$

where $\tilde{X}_n$ is a point in the interval between X_n and X_{n+1}. Taking the expectation of both sides of the above equality, and using the stationarity of X_n and equalities $\mathbb{E}W_n = \mathbb{E}W_n^3 = 0$, $\mathbb{E}W_n^2 = 1$, we have

$$\begin{aligned} 0 = &- \mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)\psi_a'(X_n)\right] a \\ &+ \frac{1}{2}\mathbb{E}_{\gamma^a}\left[\left(\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a - \sqrt{a\delta(a)\omega(X_n)}\, W_n\right)^2 \psi_a''(X_n)\right] \\ &- \frac{1}{6}\mathbb{E}_{\gamma^a}\left[\left(\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) a - \sqrt{a\delta(a)\omega(X_n)}\, W_n\right)^3 \psi_a'''(\tilde{X}_n)\right] \\ = &- a\,\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)\psi_a'(X_n)\right] \\ &+ \frac{1}{2}a\,\delta(a)\mathbb{E}_{\gamma^a}\left[\omega(X_n)\psi_a''(X_n)\right] \\ &+ \frac{1}{2}a^2\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)^2 \psi_a''(X_n)\right] \\ &- \frac{1}{2}a^2\delta(a)\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n)\frac{1}{2}\delta(a)\omega'(X_n)\right) - \omega(X_n)\psi_a'''(\tilde{X}_n)\right] + O(a^3). \end{aligned} \quad (33)$$

Dividing by a, from (28) and (33), we obtain:

$$\begin{aligned} &- \mathbb{E}_{\gamma^a}\left[\phi_a(X_n)\right] \\ &= \frac{1}{2}a\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)^2 \psi_a''(X_n)\right] \\ &- \frac{1}{2}a\delta(a)\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right) \omega(X_n)\psi_a'''(\tilde{X}_n)\right] + O(a^2). \end{aligned} \quad (34)$$

Recall that

$$\mathbb{E}_{\gamma^a}\phi_a(X_n) = \int \phi_1(x)\,\gamma^a(dx) - \eta(a)\int \phi_2(x)\,\gamma^a(dx). \quad (35)$$

Since $\phi_2(x)$ is a bounded function, and $\eta(a) \to 0$, the entire second term in the right-hand side vanishes as $a \to 0$. Also

$$\int \phi_1(x)\,\gamma^a(dx) \geq \gamma^a\left(\{H(x) > 3\varepsilon\}\right) > 0 \quad (36)$$

so in order to prove the theorem, we need to show that the right-hand side of (34) tends to 0 as $a \to 0$. From (30), we have for the first term in (34) as $a \to 0$:

$$a\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)^2 \psi_a''(X_n)\right] = O\left(\frac{a}{\delta^2(a)} e^{C/\delta(a)}\right) \mathbb{E}_{\gamma^a}\left[\omega(X_n)\left(H'(X_n)\right)^3\right]. \quad (37)$$

Analogously, using representation in (31), the second term in (34) could be written as

$$a\delta(a)\mathbb{E}_{\gamma^a}\left[\left(\omega(X_n)H'(X_n) - \frac{1}{2}\delta(a)\omega'(X_n)\right)\omega(X_n)\psi_a'''(\tilde{X}_n)\right] = O\left(\frac{a}{\delta^2(a)} e^{C/\delta(a)}\right)\mathbb{E}_{\gamma^a}\left[\omega^2(X_n)H'(X_n)\frac{\left(H'(\tilde{X}_n)\right)^2}{\omega(\tilde{X}_n)}\right] \quad (38)$$

as $a \to 0$. Expressions (37) and (38) go to zero if

$$\ln a - 2\ln\delta(a) + \frac{C}{\delta(a)} \to -\infty$$

which is satisfied if

$$\delta(a) \to 0 \quad \text{under} \quad \delta(a) \geq -\frac{C_0}{\ln a} \quad (39)$$

with a constant C_0, $C_0 > C$. Finally, for arbitrary $\varepsilon > 0$:

$$\gamma^a\left(\{H(x) > 3\varepsilon\}\right) \to 0$$

as $a \to 0$. The theorem is proved.

ACKNOWLEDGMENT

The authors would like to thank the Earth Resources Laboratory Founding Members Consortium for the support of this work.

REFERENCES

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[2] S. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, 1992.

[3] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[4] S. Geman and G. Hwang, "Diffusion for global optimization," *SIAM J. Control Optim.*, 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 J. Control Optim.*, vol. 25, pp. 737–753, 1987.

[6] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods, A Mathematical Introduction*. New York: Springer, 2003.

[7] X. Descombes and E. Zhizhina, "Applications of Gibbs fields methods to image processing problems," *Problems Inf. Transmiss.*, vol. 40, no. 3, pp. 279–295, Sep. 2004.

[8] A. D. Wentzell, *A Course in the Theory of Stochastic Processes*. New York: McGraw-Hill Int., 1981.

[9] T. M. Liggett, *Interacting Particle Systems*. New York: Springer, 2005.

[10] G. Pflug, "Applications of stochastic approximation methods," in *Stochastic Approximation and Optimization of Random Systems*, L. Ljung, G. Pflug, and H. Walk, Eds. Basel, Switzerland: Birkhäuser Verlag, 1992.

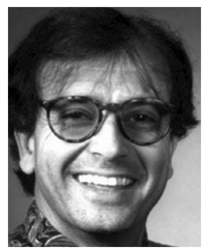
[11] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, vol. 6, no. 2, pp. 4–21, Apr. 1989.

Oleg V. Poliannikov received the M.S. degree in mathematics from Moscow State University, Russia, in 1998 and the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, in 2003.

He is currently a Postdoctoral Associate at the Earth Resources Laboratory at the Massachusetts Institute of Technology, Cambridge. His research interests include wave propagation, random media, and stochastic optimization.

Elena Zhizhina received the Ph.D. degree in mathematics and the Doctor of Science degree, both from Moscow State University, Russia, in 1990 and 2006, respectively.

She was an Assistant Professor at the Moscow Power-Engineering Institute from 1989 to 1994. Since 1994, she has been a Senior Scientific Researcher in the Dobrushin Mathematical Laboratory at the Institute for Information Transmission Problems of Russian Academy of Sciences, Moscow, Russia. She was involved in long-term joint projects with the Technical University (Muenchen), Bielefeld University, INRIA (Sophia-Antipolis). Her research interests are spin models of statistical physics, Gibbsian fields, spectral analysis of the generators of stochastic dynamics, birth-and-death spacial processes, disordered stochastic systems on the lattice, and applications of stochastic dynamics models to image processing.



Hamid Krim (S'80–M'85–SM'98–F'08) received the Ph.D. degree in electrical engineering from Northeastern University, Boston, MA, in 1991.

As a Member of Technical Staff at AT&T Bell Labs, he has worked in the area of telephony and digital communication systems/subsystems. In 1991, he became an NSF Postdoctoral Scholar at the Foreign Centers of Excellence (LSS Supelec/ University of Orsay, Paris, France). He subsequently joined the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (MIT), Cambridge, MA, as a Research Scientist performing/supervising research in his area of interest, and later as a faculty member in the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, in 1998. He is an original contributor and now an affiliate of the Center for Imaging Science sponsored by the U.S. Army. His research interests are in statistical signal processing and mathematical modeling with a keen emphasis on applications.

Dr. Krim is a recipient of the NSF Career Young Investigator award. He was on the editorial board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and regularly contributes to the society in various capacities. He is a member of SIAM and of Sigma Xi.