JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 47, No. 1, SEPTEMBER 1985

Global Optimization and Stochastic Differential Equations[1,2]

F. ALUFFI-PENTINI,[3] V. PARISI,[4] AND F. ZIRILLI[5]

Communicated by R. A. Tapia

Abstract. Let $\mathbb{R}^n$ be the n-dimensional real Euclidean space, $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$, and let $f: \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. We consider the problem of finding the global minimizers of f. A new method to compute numerically the global minimizers by following the paths of a system of stochastic differential equations is proposed. This method is motivated by quantum mechanics. Some numerical experience on a set of test problems is presented. The method compares favorably with other existing methods for global optimization.

Key Words. Global optimization, stochastic differential equations.

1. Introduction

Let $\mathbb{R}^n$ be the n-dimensional real Euclidean space, $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$, and let $f: \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. In this paper, we consider the problem of finding the global minimizers of f, that is, the points $x^* \in \mathbb{R}^n$ such that

$$f(x^*) \leq f(x), \qquad \forall x \in \mathbb{R}^n. \tag{1}$$

A new method to compute numerically the global minimizers of f by following the paths of a system of stochastic differential equations is proposed. This method is motivated by quantum mechanics.

The importance of the global optimization problem is clear. For example, the root finding problem for the system $g(x) = 0$, where $g: \mathbb{R}^n \to \mathbb{R}^n$,

[1] This research has been supported by the European Research Office of the US Army under Contract No. DAJA-37-81-C-0740.

[2] The third author gratefully acknowledges Prof. A. Rinnooy Kan for bringing to his attention Ref. 4.

[3] Associate Professor, Dipartimento di Matematica, Università di Bari, Bari, Italy.

[4] Researcher, Dipartimento di Fisica, 2a Università di Roma "Tor Vergata," Roma, Italy.

[5] Professor, Istituto di Matematica, Università di Salerno, Salerno, Italy.

0022-3239/85/0900-0001$04.50/0

can be formulated as a global optimization problem by considering the function

$$F(x) = \|g(x)\|_2^2,$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^n$. Despite its importance and the contributions of many researchers, the situation with respect to algorithms for the global optimization problem is still unsatisfactory, and there is a need for methods with a solid mathematical foundation and good numerical performance. The situation for the problem of finding the local minimizers of f is much more satisfactory, and a large body of theoretical and numerical results has been established; see, for example, Ref. 1 and the references given therein.

Ordinary differential equations have been used in the study of the local optimization problem or the root finding problem by several authors; for a review, see Ref. 2. These methods usually approximate the local optimizers or roots by following the trajectories of suitable systems of ordinary differential equations. However, since property (1) is a global property (that is, it depends on the behavior of f on each point of $\mathbb{R}^n$) and since the methods that follow a trajectory of a system of ordinary differential equations are local (that is, they depend only on the behavior of f along the trajectory), there is no hope of building a completely satisfactory method for global optimization based on a system of ordinary differential equations. However, the situation is different if we consider a suitable stochastic perturbation of a system of ordinary differential equations as we now describe.

Let us consider the Ito stochastic differential equation

$$d\xi = -\nabla f(\xi)\, dt + \epsilon\, dw, \tag{2}$$

where ∇f is the gradient of f and $w(t)$ is a standard n-dimensional Wiener process. When $\epsilon = \epsilon_0$ is a constant, Eq. (2) is known as the Smoluchowski–Kramers equation (Ref. 3). This equation is a singular limit of the Langevin equation when the inertial terms are neglected. The Smoluchowski–Kramers equation has been used widely by solid state physicists and chemists to study physical phenomena such as atomic migration in crystals or chemical reactions. In these applications, $\epsilon_0 = \sqrt{(2kT/m)}$, where T is the absolute temperature, k the Boltzmann constant, m the reduced mass, and f the potential energy, so that (2) represents diffusion across potential barriers under the stochastic forces $\epsilon_0\, dw$.

It is well known that, if $\xi^{\epsilon_0}(t)$ is the solution process of (2) starting from an initial point x_0, then the probability density function of $\xi^{\epsilon_0}(t)$ approaches, as $t \to \infty$, the limit density $A_{\epsilon_0} \exp[-2f(x)/\epsilon_0^2]$, where A_{ϵ_0} is a normalization constant. The limit density is independent of x_0 and is peaked

(indicating concentration of particles) around the global minimizers of f, with narrower peaks if the constant ϵ_0 is smaller.

The method that we propose attempts to obtain a global minimizer of f by looking at the asymptotic value, as $t \to \infty$, of a numerically computed sample trajectory of an equation like (2), where $\epsilon = \epsilon(t)$ is a function of time which tends to zero in a suitable way as $t \to \infty$. Similar ideas in the context of discrete optimization have been introduced by Kirkpatrick, Gelatt, and Vecchi (Ref. 4).

In Section 2, we describe our method; in Section 3, we consider the numerical integration problem; and, in Section 4, we present the results of numerical experiments on several test problems.

2. Method

Let us consider the Cauchy problem

$$d\xi = -\nabla f(\xi)\, dt + \epsilon(t)\, dw, \tag{3}$$

$$\xi(0) = x_0, \tag{4}$$

for the Ito stochastic differential equation (3), where $f: \mathbb{R}^n \to \mathbb{R}$ is the function to be globally minimized, ∇f is the gradient of f, $w(t)$ is an n-dimensional standardized Wiener process, and $\epsilon(t)$ is a given function. We assume that f and ϵ are sufficiently well behaved, so that our statements are meaningful; in particular, we assume that

$$\lim_{\|x\|_2 \to \infty} f(x) = +\infty,$$

$$\int_{\mathbb{R}^n} \exp[-\alpha^2 f(x)]\, dx < \infty, \qquad \forall \alpha \in R \backslash \{0\},$$

and that f has only a finite number of isolated global minimizers.

We propose to integrate numerically problem (3), (4) looking at the asymptotic value of a sample numerical trajectory solution to obtain a global minimizer of f. Let us start by considering problem (3), (4) when $\epsilon = \epsilon_0$ is a constant; that is,

$$d\xi = -\nabla f(\xi)\, dt + \epsilon_0\, dw(t), \tag{5}$$

$$\xi(0) = x_0. \tag{6}$$

Let $\xi^{\epsilon_0}(t)$ be the stochastic process solution of (5), (6); for any Borel set $A \subset \mathbb{R}^n$, we define

$$P^{\epsilon_0}(0, x_0, t, A) = \mathbb{P}\{\xi^{\epsilon_0}(t) \in A\}, \tag{7}$$

where $\mathbb{P}\{\cdot\}$ is the probability of $\{\cdot\}$ and $P^{\epsilon_0}(0, x_0, t, A)$ is the transition probability of $\xi^{\epsilon_0}(t)$. Under regularity assumptions for f, we have

$$P^{\epsilon_0}(0, x_0, t, A) = \int_A p^{\epsilon_0}(0, x_0, t, x)\, dx, \tag{8}$$

where the transition probability density $p = p^{\epsilon_0}(0, x_0, t, x)$ satisfies the following Fokker-Planck equation

$$\partial p/\partial t = (\epsilon_0^2/2)\Delta p + \operatorname{div}(\nabla f p), \tag{9}$$

with

$$\lim_{t\to 0} p^{\epsilon_0}(0, x_0, t, x) = \delta(x - x_0), \tag{10}$$

where Δ and div are the Laplacian and the divergence with respect to x and $\delta(\cdot)$ is the Dirac delta function.

Let A_{ϵ_0} be defined by

$$1/A_{\epsilon_0} \equiv \int_{\mathbb{R}^n} \exp[-2f(x)/\epsilon_0^2]\, dx < \infty. \tag{11}$$

Then, as $t \to \infty$, the transition probability density $p^{\epsilon_0}(0, x_0, t, x)$ approaches the function

$$p_\infty^{\epsilon_0}(0, x_0, x) = A_{\epsilon_0} \exp[-2f(x)/\epsilon_0^2]. \tag{12}$$

Clearly, $p_\infty^{\epsilon_0}$ is the probability density of a random variable $\xi_\infty^{\epsilon_0}$, so that $\xi^{\epsilon_0}(t) \to \xi_\infty^{\epsilon_0}$ in law when $t \to \infty$. Let us remark that $p_\infty^{\epsilon_0}$ does not depend on the initial condition x_0.

We want to study the behavior of $p_\infty^{\epsilon_0}$ as $\epsilon_0 \to 0$ and the rate of approach of p^{ϵ_0} to $p_\infty^{\epsilon_0}$ as $t \to \infty$. We will consider for the sake of simplicity only the

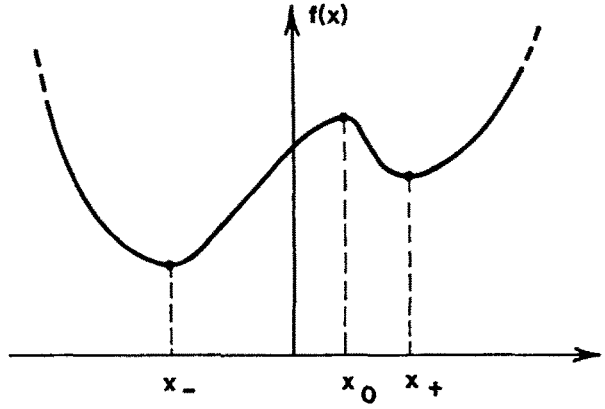

Fig. 1. The function $f(x)$.

one-dimensional case, when f is as in Fig. 1, i.e., with three extrema at the points $x_- < x_0 < x_+$, decreasing in $(-\infty, x_-)$ and (x_0, x_+), and increasing in (x_-, x_0) and $(x_+, +\infty)$, with $f(x) \to +\infty$ as $|x| \to \infty$, in such a way as to satisfy (11) for all $\epsilon_0 \neq 0$.

We have

$$\frac{df}{dx}(x_+) = \frac{df}{dx}(x_-) = \frac{df}{dx}(x_0) = 0.$$

Using the following notation:

$$f_+ = f(x_+), \qquad c_+ = \frac{d^2 f}{dx^2}(x_+),$$

$$f_- = f(x_-), \qquad c_- = \frac{d^2 f}{dx^2}(x_-),$$

$$f_0 = f(x_0), \qquad c_0 = -\frac{d^2 f}{dx^2}(x_0),$$

$$\Delta f_- = f_0 - f_- > 0, \qquad \Delta f_+ = f_0 - f_+ > 0,$$

it is easy to prove the following result.

Proposition 2.1. Let f be as above, and let c_0, c_+, c_- be greater than zero. The following results hold:

(i) if $\Delta f_- > \Delta f_+$ and $\exists \alpha > 0$ such that $f(x) \geq \alpha(x - x_-)^2 + f_-$, $\forall x \in \mathbb{R}$, then

$$\lim_{\epsilon_0 \to 0} p_\infty^{\epsilon_0}(0, x_0, x) = \delta(x - x_-); \tag{13}$$

(ii) if $\Delta f_- = \Delta f_+$ and $\exists \alpha > 0$ such that $f(x) \geq \alpha(x - x_-)^2 + f_-$, $\forall x \leq x_0$, and $f(x) \geq \alpha(x - x_+) + f_+$, $\forall x \geq x_0$, then

$$\lim_{\epsilon_0 \to 0} p_\infty^{\epsilon_0}(0, x_0, x) = \gamma\delta(x - x_-) + (1 - \gamma)\delta(x - x_+), \qquad \gamma = [1 + \sqrt{(c_-/c_+)}]^{-1}, \tag{14}$$

where the limits (13), (14) are taken in the distribution sense. Proposition 2.1 is easy to prove using the Taylor formula for f around x_-, x_+.

Remark 2.1. Proposition 2.1 shows that, as $\epsilon_0 \to 0$, the asymptotic probability density approaches a Dirac delta function concentrated on the global minimizer when there is a unique global minimizer ($\Delta f_- > \Delta f_+$) or approaches a linear combination of Dirac delta functions concentrated on

the global minimizers ($\Delta f_- = \Delta f_+$). The coefficients of the linear combination depend on the curvature of f at the global minimizers. These statements have a clear meaning in terms of $\xi_\infty^{\epsilon_0}$. Finally, Proposition 2.1 can be generalized easily to a wider class of functions f.

Proposition 2.2. Under the previous hypotheses for f, Matkowsky and Schuss studied (Ref. 5) the rate of convergence of p^{ϵ_0} to $p_\infty^{\epsilon_0}$ as $t \to \infty$ by looking at the eigenvalues of the Fokker–Planck operator

$$L_{\epsilon_0}(\cdot) = (\epsilon_0^2/2)[\partial^2(\cdot)/\partial x^2] + (\partial/\partial x)[(df/dx)(\cdot)].$$

We note that $p_\infty^{\epsilon_0}$ is an eigenfunction with eigenvalue zero of L_{ϵ_0}, so that the rate of approach to $p_\infty^{\epsilon_0}$ is determined by the next eigenvalue $\lambda_1(\epsilon_0)$ of L_{ϵ_0}. Matkowsky and Shuss obtained for $\lambda_1(\epsilon_0)$ the following asymptotic expression as $\epsilon_0 \to 0$:

$$\lambda_1(\epsilon_0) \cong -[\sqrt{(c_+ c_0)}/2\pi] \exp[-(2/\epsilon_0^2)\Delta f_+], \tag{15}$$

so that roughly speaking we can imagine that

$$p^{\epsilon_0}(0, x, t, x) = p_\infty^{\epsilon_0} + \exp\left[\int_0^t \lambda_1(\epsilon_0)\, ds\right] \tilde{p}, \tag{16}$$

where $\tilde{p}$ is an eigenfunction corresponding to λ_1.

When $f(x)$ is a fourth-order polynomial with two minimizers, a complete analysis of the spectrum of L_{ϵ_0} in the limit $\epsilon_0 \to 0$ has been given by Angeletti, Castagnari, and Zirilli in Ref. 6.

Remark 2.2. Since $\lambda_1(\epsilon_0) \to 0$ as $\epsilon_0 \to 0$, from (16) we see that the rate of approach to $p_\infty^{\epsilon_0}$ becomes slower when ϵ_0 becomes smaller. On the other hand, from (12) we see that $p_\infty^{\epsilon_0}$ becomes more and more concentrated around the global optimizers as ϵ_0 goes to zero.

Let us go back now to (3), (4) when $\epsilon = \epsilon(t)$ is a given function of t, and let $\xi(t)$ be the solution of (3), (4). Let $P(0, x_0, t, A)$ be the transition probability of $\xi(t)$ and $p(0, x_0, t, x)$ the corresponding probability density. Under regularity assumptions for f, the probability density p satisfies the following Fokker–Planck equation:

$$\partial p/\partial t = (\epsilon^2(t)/2]\Delta p + \operatorname{div}(\nabla f p), \tag{17}$$

$$\lim_{t \to 0} p(0, x_0, t, x) = \delta(x - x_0). \tag{18}$$

In order to compute the global optimizers of f by following the paths of (3), (4), we would like to show that

$$\lim_{t \to \infty} p(0, x_0, t, x) = \sum_{i=1}^{m} \gamma_i\, \delta(x - x_i^*), \tag{19}$$

where γ_i are positive constants such that

$$\sum_{i=1}^{m} \gamma_i = 1$$

and where x_i^*, $i = 1, 2, 3, \ldots, m$, are the global minimizers of f.

The previous analysis of the corresponding problem with $\epsilon(t) = \epsilon_0$ suggests that, in order to have (19), we need

$$\lim_{t \to \infty} \epsilon(t) = 0; \tag{20}$$

and, as suggested by (16), we must require that

$$\int_0^{\infty} \exp\{-[2/\epsilon^2(t)]\Delta f_+\}\, dt = \infty, \tag{21}$$

where Δf_+ is the highest barrier to the global minimizers. We note that, in order to satisfy (21), $\epsilon(t)$ must go to zero very slowly.

The problem of giving a mathematically rigorous foundation to our method by proving (19) will be considered elsewhere. Based on the heuristic conditions (20), (21), we will consider now the problem of how to integrate numerically (3), (4) in order to obtain a global minimizer of f.

3. Numerical Integration

In the previous sections, we have proposed to obtain the global minimizers of f by following the paths defined by (3), (4) under suitable assumptions for $\epsilon(t)$ when $t \to \infty$. We want to consider here the problem of how to compute numerically these paths, keeping in mind that we are not really interested in the paths, but only in their asymptotic values.

The algorithm that we propose here is only preliminary, and further study is needed; however, as we will see in Section 4, even the present algorithm gives good numerical results on several test problems.

Let

$$\Delta t_k > 0, \quad t_k = \sum_{i=0}^{k-1} \Delta t_i (t_0 = 0), \qquad k = 0, 1, \ldots;$$

we discretize (3), (4) using the Euler–Cauchy method; that is, $\xi(t_k)$ is approximated by the ξ_k solution of the following finite difference equations:

$$\xi_{k+1} - \xi_k = -\Delta t_k \nabla f(\xi_k) + \epsilon(t_k)(w_{k+1} - w_k), \qquad k = 0, 1, \ldots, \tag{22}$$

$$\xi_0 = x_0. \tag{23}$$

Since for stability reasons Δt_k will be chosen rather small, and since condition (21) implies that $\epsilon(t)$ should go to zero very slowly in order to reach the asymptotic values of the paths of (3), (4), we expect that a large number of time integration steps (22) will be needed.

Let r be an n-dimensional random vector of length 1 uniformly distributed on the $(n-1)$-dimensional sphere; then, for any given nonrandom vector $v \in \mathbb{R}^n$, its projection $\langle v, r\rangle\, r$ along r is such that

$$n \cdot E(\langle v, r\rangle r) = v,$$

where $E(\cdot)$ is the expected value and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product in $\mathbb{R}^n$. This suggests that, in order to save numerical work (i.e., function evaluations), we may replace $\nabla f(\xi_k)$ in Eq. (22) by the expression

$$n\langle \nabla f(\xi_k), r\rangle r, \tag{24}$$

where (24), the directional derivative in the direction r, may be further approximated by finite differences with some mesh size Δx_k.

When forward differences are used, $n+1$ function evaluations are needed to approximate ∇f, while only two function evaluations are needed to approximate the directional derivative. Finally, some heuristic algorithms are used to choose Δt_k and Δx_k to avoid instabilities. Condition (21) suggests that $\epsilon(t)$ should go to zero very slowly as t goes to infinity, so that computing a single path of (3), (4), choosing $\epsilon(t)$ as required by (21), and following this path for a long enough period of time to obtain a global minimizer does not seem very efficient.

We consider the following alternative strategy.

(i) N paths of (3), (4) are computed with $N > 1$ ($N = 7$ in the numerical experience shown in Section 4) using the algorithm described before, and $\epsilon(t)$ is kept constant;

(ii) f is computed along the paths and used as a merit function. After a number of steps of numerical integration, the N computed paths are compared. The worst path is discarded, and the numerical integration is continued after splitting one of the remaining $N-1$ paths into two paths.

The new path has a different value of $\epsilon(t) = \text{const}$; $\epsilon(t)$ is usually decreased; occasionally, it can be increased if the paths are stuck in a local minimizer as detected by looking at the previously computed values of f.

(iii) Repeat from step (ii).

4. Test Problems and Numerical Experience

The algorithm described in Sections 2 and 3 has been tested on a set of test problems. The first 18 test problems have been taken from the

literature; they were proposed as a set of problems to test global optimization methods by Levy and Montalvo (Ref. 7).

We shall make use of the penalization function

$$u(x, a, k, m) = \begin{cases} k(x-a)^m, & x > a, \\ 0, & -a \leq x \leq a, \\ k(-x-a)^m, & x < -a. \end{cases}$$

The test problems are given below.

Problem 1. *Goldstein's Function.* Let

$$f(x) = x^6 - 15x^4 + 27x^2 + 250;$$

the function f has three minima:

$$x = -3, \quad f(x) = 7,$$

$$x = 0, \quad f(x) = 250,$$

$$x = 3, \quad f(x) = 7.$$

The minimizers $x = \pm 3$ are the global minimizers of f.

Problem 2. *Penalized Shubert Function.* Let

$$g_1(x) = \sum_{i=1}^{5} i \cos((i+1)x+1);$$

the function g_1 is the Shubert function. We define the penalized Shubert function $f(x)$ as follows:

$$f(x) = g_1(x) + u(x, 10, 100, 2).$$

This function has 19 minima in the region $\{x \mid |x| < 10\}$; three of these minima are global ones and are located at

$$x = -7.70831, \quad x = -1.42513, \quad x = 4.85805.$$

Problem 3. *Two-Dimensional Penalized Shubert Function.* Let

$$f(x, y) = \left\{ \sum_{i=1}^{5} i \cos[(i+1)x+1] \right\} \left\{ \sum_{i=1}^{5} i \cos[(i+1)y+1] \right\} + u(x, 10, 100, 2) + u(y, 10, 100, 2).$$

The function f has 760 minima (18 of them are global minima) in the region $\{(x, y) \mid |x| < 10, |y| < 10\}$.

Problem 4. *Two-Dimensional Penalized Shubert Function,* $\beta = 0.5$.

$$f(x, y) = \left\{ \sum_{i=1}^{5} i \cos[(i+1)x+1] \right\} \left\{ \sum_{i=1}^{5} i \cos[(i+1)y+1] \right\}$$
$$+ \beta[(x+1.42513)^2 + (y+0.80032)^2]$$
$$+ u(x, 10, 100, 2) + u(y, 10, 100, 2),$$

where $\beta = 0.5$ and $(-1.42513, -0.80032)$ is a point where the function f with $\beta = 0$ has a global minimizer. This function has roughly the same behavior as the function considered in Problem 3, but has a unique global minimizer at $(-1.42513, -0.80032)$, where the function f is equal to -186.7309.

Problem 5. *Two-Dimensional Penalized Shubert Function,* $\beta = 1$. The function f is the one given in Problem 4 with $\beta = 1$.

Problem 6. *Camel Function.* Let f be given by

$$f(x, y) = (4 - 2.1x^2 + x^4/3)x^2 + xy + (-4 + 4y^2)y^2.$$

This function has six minima; two of them are global minima and are located at $(-0.0898, 0.7126)$, $(0.0898, -0.7126)$.

Problems 7–9 are obtained from the following formula:

$$g_2(x) = (\pi/n) \left\{ k_2 \sin^2(\pi y_1) \right.$$
$$\left. + \sum_{i=1}^{n-1} (y_i - A_2)^2 [1 + k_2 \sin^2(\pi y_{i+1})] + (y_n - A_2)^2 \right\}, \tag{25}$$

where

$$x = (x_1, x_2, \ldots, x_n)^T,$$
$$y_i = 1 + (x_i - 1)/4, \qquad i = 1, 2, \ldots, n,$$
$$k_2 = 10, \qquad A_2 = 1.$$

In the region

$$\Omega = \{x \in R \mid -10 \leq x_i \leq 10, i = 1, 2, \ldots, n\},$$

the function (25) has roughly 5^n local minimizers and a unique global minimizer located at

$$x_i = 1, \qquad i = 1, 2, \ldots, n.$$

We penalize the function (25) as follows:

$$f(x) = g_2(x) + \sum_{i=1}^{n} u(x_i, 10, 100, 4). \tag{26}$$

Problem 7. The function $f(x)$ is given by (26) with $n = 2$.

Problem 8. The function $f(x)$ is given by (26) with $n = 3$.

Problem 9. The function $f(x)$ is given by (26) with $n = 4$.

Problems 10–12 are obtained from the following formula:

$$g_3(x) = (\pi/n)\left\{k_3 \sin^2(\pi x_1) + \sum_{i=1}^{n-1} (x_i - A_3)^2[1 + k_3 \sin^2(\pi x_{i+1})] + (x_n - A_3)^2\right\}, \tag{27}$$

where

$$k_3 = 10, \qquad A_3 = 1,$$

$$x = (x_1, x_2, \ldots, x_n)^T.$$

In the region

$$\Omega = \{x \in \mathbb{R}^n \mid -10 \leq x_i \leq 10,\ i = 1, 2, \ldots, n\},$$

the function (27) has roughly 10^n local minimizers and a unique global minimizer at $x_i = 1$, $i = 1, 2, \ldots, n$. We penalize the function (27) as follows:

$$f(x) = g_3(x) + \sum_{i=1}^{n} u(x_i, 10, 100, 4). \tag{28}$$

Problem 10. The function $f(x)$ is given by (28) with $n = 5$.

Problem 11. The function $f(x)$ is given by (28) with $n = 8$.

Problem 12. The function $f(x)$ is given by (28) with $n = 10$.

Problems 13–18 are obtained from the following formula:

$$g_4(x) = k_4\left\{\sin^2(\pi l_0 x_1) + \sum_{i=1}^{n-1} (x_i - A_4)^2[1 + k_5 \sin^2(\pi l_0 x_{i+1})] + (x_n - A_4)^2[1 + k_5 \sin^2(\pi l_1 x_n)]\right\}, \tag{29}$$

where

$$k_4 = 0.1, \qquad k_5 = 1, \qquad A_4 = 1, \qquad l_0 = 3, \qquad l_1 = 2.$$

In the region

$$\Omega = \{x \in \mathbb{R}^n \mid -10 \leq x_i \leq 10,\ i = 1, 2, \ldots, n\},$$

the function (29) has roughly 30^n local minimizers and a unique global minizer at $x_i = 1$, $i = 1, 2, \ldots, n$. In the region

$$\Omega_1 = \{x \in \mathbb{R}^n \mid -5 \leq x_i \leq 5,\ i = 1, 2, \ldots, n\},$$

the function (29) has roughly 15^n local minimizers and a unique global minimizer at $x_i = 1$, $i = 1, 2, \ldots, n$. We penalize the function (29) as follows:

$$f(x) = g_4(x) + \sum_{i=1}^{n} u(x_i, 10, 100, 4), \tag{30}$$

or

$$f(x) = g_4(x) + \sum_{i=1}^{n} u(x_i, 5, 100, 4). \tag{31}$$

Problem 13. The function $f(x)$ is given by (30) with $n = 2$.

Problem 14. The function $f(x)$ is given by (30) with $n = 3$.

Problem 15. The function $f(x)$ is given by (30) with $n = 4$.

Problem 16. The function $f(x)$ is given by (31) with $n = 5$.

Problem 17. The function $f(x)$ is given by (31) with $n = 6$.

Problem 18. The function $f(x)$ is given by (31) with $n = 7$.

Problems 19–22 have been created by the third author.

Problem 19. Let

$$f(x) = x^4/4 - x^2/2 + 0.1x;$$

the function f has two minima, one for positive x and one for negative x. The one for negative x is the global one.

Problem 20. Let

$$f(x, y) = x^4/4 - x^2/2 + 0.1x + y^2/2;$$

the function f has two minima, $(x_1, 0)$ and $(x_2, 0)$, where x_1, x_2 are the minimizers of the function of Problem 19. The minimizer with the negative x corresponds to the global minimizer.

Problem 21. Let

$$f(x, y) = 0.5x^2 + 0.5[1 - \cos(2x)] + y^2;$$

the function f has several local minima, and the global minimizer is the origin.

Problem 22. Let $n > 0$, $m < 0$ and

$$f(x, y) = 10^n x^2 + y^2 - (x^2 + y^2)^2 + 10^m (x^2 + y^2)^4;$$

the function f has a local minimum at the origin and two global minimizers on the y-axis.

Problem 23. Let

$$f(x) = \left[\sum_{i=1}^{5} i x_i^2 \right]^{1/4},$$

where

$$x = (x_1, \ldots, x_5)^T;$$

the function $f(x)$ has a unique minimizer at $x = 0$, where the function is not differentiable; moreover, the Hessian of $f(x)$ is not defined at $x = 0$ and is not positive definite in a neighborhood of $x = 0$.

Problem 24 has been suggested by Wolff (Ref. 8).

Problem 24. Let

$$f(x, y) = -F(x, y) + u(x, 10^4, 100, 2) + u(y, 10^4, 100, 2),$$

where

$$F(x, y) = \prod_{i=1}^{14} [\Phi(x_i - x)/y]^{1-\delta_i} [1 - \Phi(x_i - x)/y]^{\delta_i}.$$

The function $\Phi(x)$ is given by

$$\Phi(x) = \int_{-\infty}^{x} [1/\sqrt{(2\pi)}] \exp(-t^2/2)\, dt.$$

Table 1. Data points for Problem 24.

x_i	1219	1371	1377	1144	1201	1225	1244
δ_i	0	0	0	1	1	1	1
x_i	1254	1304	1328	1351	1356	1370	1390
δ_i	1	1	1	1	1	1	1

The data points x_i, δ_i are given in Table 1. The function $f(x, y)$ has an absolute minimizer at (1523.2, 277.5) and a spurious relative minimizer, due to the penalization, at $(-6607.3, -10^4)$.

The numerical results obtained are shown in Table 2.

The program is run twice on each problem, the first time with a given stopping criterion. NF1 is the number of function evaluations (including the ones needed to evaluate the gradient) used in this first run; the result obtained is shown in Column 3. In the second run, a more stringent stopping criterion is used; Columns 4, 5 have the same meaning as Columns 2, 3,

Table 2. Numerical results.

Problem	NF1	Whether a global minimizer has been found	NF2	Whether a global minimizer has been found	Remarks
1	3,184	Yes	7,168	Yes	
2	26,893	Yes	77,699	Yes	
3	3,218	No	241,215	Yes	
4	8,755	Yes	76,894	Yes	
5	97,761	Yes	183,819	Yes	
6	5,393	Yes	10,822	Yes	
7	84,782	Yes	159,549	Yes	
8	19,041	Yes	72,851	Yes	
9	18,942	Yes	49,690	Yes	
10	18,433	Yes	72,226	Yes	
11	4,322	No	136,061	Yes	
12	49,701	Yes	98,985	Yes	
13	9,492	Yes	23,770	Yes	
14	19,114	Yes	66,010	Yes	
15	35,139	Yes	122,166	Yes	
16	53,398	Yes	66,365	Yes	
17	15,534	Yes	98,974	Yes	
18	16,542	Yes	109,886	Yes	
19	6,751	Yes	16,487	Yes	
20	3,402	Yes	12,249	Yes	
21	10,286	Yes	19,940	Yes	
22	4,791	Yes	7,390	Yes	$n=-m=1$
22	3,037	Yes	4,853	Yes	$n=-m=2$
22	5,028	Yes	8,235	Yes	$n=-m=3$
22	14,710	Yes	27,859	Yes	$n=-m=4$
22	51,285	Yes	74,194	Yes	$n=-m=5$
22	17,610	Yes	4,042,861	Yes	$n=-m=6$
23	15,102	Yes	34,110	Yes	
24	48,802	Yes	69,512	Yes	

respectively. All the remaining parameters [initial value for $\epsilon(t)$, etc] are fixed once and for all during the runs.

The initial point x_0 has been chosen as follows:

$x_0 = 0$, for Problems 1-18;

$x_0 = 1/50$, for Problem 19;

$x_0 = (1, 0)$, for Problem 20;

$x_0 = (-3, 0)$, for Problem 21;

$x_0 = (0, 1)$, for Problem 22;

$x_0 = (10^3, 10^3, \ldots, 10^3)$, for Problem 23;

$x_0 = (-1250, -1000)$, for Problem 24.

For Problems 19-22 and 24, initial point x_0 has been chosen close to a local minimizer.

The condition number of the Hessian at the solution of Problem 22 increases with $n, -m$; the Hessian at the solution of Problem 23 is not defined; the Hessian at the solution of Problem 24 is ill conditioned; the remaining problems have well-conditioned Hessians at the solutions.

The numerical experience contained in Table 2 shows that the present implementation of our method is much more sensitive to ill conditioning than to the total number of local minimizers. This seems to be due to the method used to integrate numerically the stochastic differential equations. However, we should remark that, on Problems 10, 11, 12, 16, 17, 18 (which have very large number of local minimizers), the global one is obtained by using a number of function evaluations which is much smaller than the number of local minimizers. Our method gives satisfactory results on all the test problems, including Problem 23, that is not differentiable at the solution. Finally, we note that, given the stochastic nature of the method, the amount of work needed to solve a problem depends on the problem and on the sequence of random numbers generated during the numerical integration.

We feel that further work of both mathematical and numerical character must be spent on the ideas presented in this paper.

References

1. POWELL, M. J. D., Editor, *Nonlinear Optimization 1981*, Academic Press, New York, New York, 1982.
2. ZIRILLI, F., *The Use of Ordinary Differential Equations in the Solution of Nonlinear Systems of Equations*, Nonlinear Optimization 1981, Edited by M. J. D. Powell, Academic Press, New York, New York, 1982.
3. SCHUSS, Z., *Theory and Applications of Stochastic Differential Equations*, John Wiley and Sons, New York, New York, Chapter 8, 1980.

4. KIRKPATRICK, S., GELATT, C. D., JR., and VECCHI, M. P., *Optimization by Simulated Annealing*, Science, Vol. 220, pp. 671–680, 1983.
5. MATKOWSKY, B. J., and SCHUSS, Z., *Eigenvalues of the Fokker–Planck Operator and the Approach to Equilibrium for Diffusions in Potential Fields*, SIAM Journal on Applied Mathematics, Vol. 40, pp. 242–254, 1981.
6. ANGELETTI, A., CASTAGNARI, C., and ZIRILLI, F., *Asymptotic Eigenvalue Degeneracy for a Class of One-Dimensional Fokker–Planck Operators*, Journal of Mathematical Physics, Vol. 26, pp. 678–690, 1985.
7. LEVY, A. V., and MONTALVO, A., *Algoritmo de Tunelización para la Optimización Global de Funciones*, Universidad Nacional Autónoma de México, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Report No. 204, 1979.
8. WOLFF, S., Private Communication, 1983.