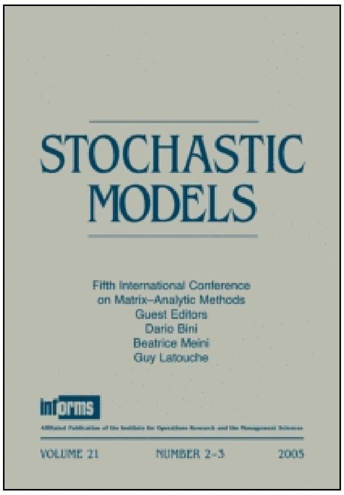

Stochastic Models
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597301

Complexity of non-adaptive optimization algorithms for a class of diffusions

James M. Calvin[a]; Peter W. Glynn[b]
[a] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA [b] Department of Operations Research, Stanford University, Stanford, CA

To cite this Article Calvin, James M. and Glynn, Peter W.(1996) 'Complexity of non-adaptive optimization algorithms for a class of diffusions', Stochastic Models, 12: 3, 343 — 365
To link to this Article: DOI: 10.1080/15326349608807389
URL: http://dx.doi.org/10.1080/15326349608807389

COMMUN. STATIST.—STOCHASTIC MODELS, 12(3), 343–365 (1996)

COMPLEXITY OF NON-ADAPTIVE OPTIMIZATION ALGORITHMS FOR A CLASS OF DIFFUSIONS

James M. CALVIN

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205

Peter W. GLYNN

Department of Operations Research
Stanford University
Stanford, CA 94305-4022

Abstract

This paper is concerned with the analysis of the average error in approximating the global minimum of a 1-dimensional, time-homogeneous diffusion by non-adaptive methods. We derive the limiting distribution of the suitably normalized approximation error for both random and deterministic non-adaptive approximation methods. We identify the form of the asymptotically optimal random non-adaptive approximation methods.

Key words and phrases. Global optimization, average-case complexity, diffusion processes.

1 Introduction

This paper is a study of the average error in approximating the global minimum of a one-dimensional diffusion using non-adaptive optimization algorithms. This topic is part of the general area of average-case performance analysis of global optimization algorithms. Our motivation for studying the problem in the one-dimensional diffusion setting is to generalize properties that have been established for the special case of Brownian motion, and to gain a detailed understanding of the relative performance of different non-adaptive methods in this special setting.

The probabilistic setting in which average performance of global optimization algorithms has usually been studied is the case of Brownian motion. Ritter [11] showed that any non-adaptive optimization algorithm has average convergence rate of order $n^{-1/2}$ in the number of observations n. Calvin [3] and Al-Mharmah and Calvin [1] compared the average performance of deterministic and random grids, and the latter reference derives the optimal random sampling density for Brownian motion. Asmussen et al [2] characterized the limiting normalized average error for a deterministic uniformly spaced grid of observations. Calvin and Glynn [4] derive the limiting distribution of the normalized error in the case of uniform sampling, and show that the error fails to converge for almost all Brownian paths (so for convergence, the error must be averaged over all paths). While Brownian motion has served as a model in many studies, it is not clear how the results and intuition obtained from the studies would change for more general Markov processes.

Our primary purpose is to show that the results obtained for Brownian motion carry over in some form to a large class of one-dimensional diffusions, namely the diffusions that appear "locally Brownian". Since the asymptotic study of the error depends on the sample path only in a neighborhood of the global minimum, processes that evolve over a short time period like Brownian motion should behave in a neighborhood of the global minimum the way that Brownian motion does. Previous results for Brownian motion have relied on a path decomposition and limit theorem that gives that the difference between the path and the global minimum after the minimum, with time scaled by n^{-1} and space scaled by $n^{-1/2}$, converges in distribution to a 3-dimensional Bessel process. Similarly, the same process going to the left of the minimizer converges in distribution to an independent 3-dimensional Bessel process. Using a path decomposition of Fitzsimmons [7], which generalizes an earlier path decomposition of Williams [14], we derive a similar result for general diffusions. The main difference is that the diffusion coefficient at the level of the global minimum comes into play in the normalization. This result allows us to characterize optimal random non-adaptive algorithms for specific diffusions.

In all cases we consider, the error multiplied by the square root of the number of observations converges in distribution. By "optimal" we mean that the limiting random variable has the smallest mean. Therefore, the algorithms we consider need not be optimal in any sense for a fixed number of observations, and we consider two algorithms to have equivalent asymptotic performance if their errors differ by $o(n^{-1/2})$ as the number of observations n goes to infinity.

In Section 2 we establish notation and assumptions and review some essential results on linear diffusions. Section 3 contains Fitzsimmons' path decomposition and the two-sided Bessel process limit result. The implications for non-adaptive algorithms are given in Section 4, where optimality results are given along with precise convergence rates.

2 Notation and Background

Let $(X(t) : t \geq 0)$ be a time homogeneous regular diffusion process on the real line with generator $\mathcal{A}$ acting on C^2 functions as

$$\mathcal{A}f(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x), \quad x \in \mathbb{R}, \tag{1}$$

where $\sigma(\cdot)$ does not vanish,

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K|x - y| \tag{2}$$

and

$$b^2(x) + \sigma^2(x) \leq K(1 + x^2) \tag{3}$$

for some constant $K > 0$. We assume that all paths of X are continuous and that X enjoys the strong Markov property. In terms of the functions σ and b, the scale function S can be expressed as

$$S(x) = \int_{x_0}^{x} \exp\left(- \int_{y_0}^{y} \frac{2b(z)}{\sigma^2(z)} dz \right) dy$$

and the speed measure has density m with respect to Lebesgue measure, where

$$m(x) = \frac{2}{S'(x)\sigma^2(x)}. \tag{4}$$

We recall some facts about diffusions; see for example Itô and McKean [8], Sections 4.6 and 4.11. For each $\alpha > 0$ there are strictly positive, linearly independent solutions g_1^α and g_2^α of

$$\mathcal{A}g(x) = \alpha g(x), \quad x \in \mathbb{R}$$

such that g_1^α is increasing and g_2^α is decreasing, g_1^α vanishes at $-\infty$ and g_2^α vanishes at ∞. (Because of our assumptions (2) and (3), $-\infty$ and $+\infty$ are natural boundaries, and the last conditions on g_1^α, g_2^α are the appropriate boundary conditions for the domain of the generator; see Ethier and Kurtz [6], p. 367.) Both functions are determined uniquely up to a positive multiple. Since g_1^α and g_2^α are linearly independent, the Wronskian $B^\alpha = g_1^{\alpha+} g_2^\alpha - g_1^\alpha g_2^{\alpha+}$ is constant, where g^+ is the right-hand scale derivative of g:

$$g^+(x) = \lim_{y \downarrow x} \frac{g(y) - g(x)}{S(y) - S(x)}.$$

In the following we will suppress α in the notation. The process has a symmetric transition density $p(t; x, y) = p(t; y, x)$ with respect to the speed measure, and p is jointly continuous in all three variables. The Green function g is given by

$$g(x,y) = g(\alpha; x, y) \triangleq \int_0^\infty e^{-\alpha t} p(t; x, y)\, dt = \begin{cases} \frac{1}{B} g_1^\alpha(x) g_2^\alpha(y) & \text{if } x \le y, \\ \frac{1}{B} g_2^\alpha(x) g_1^\alpha(y) & \text{if } x \ge y. \end{cases} \tag{5}$$

Let $f(t; x, y)$ be the density (with respect to Lebesgue measure) of the first hitting time from x to y (for the existence of the density, see

Csáki et al [5]). The Laplace transforms of hitting times are expressed by

$$E_x\, e^{-\alpha\tau_y} = \int_0^\infty e^{-\alpha t} f(t;x,y)dt = \begin{cases} g_1^\alpha(x)/g_1^\alpha(y) & \text{if } x \le y, \\ g_2^\alpha(x)/g_2^\alpha(y) & \text{if } x \ge y. \end{cases} \tag{6}$$

Let $M_t = \min_{0\le s\le t} X(s)$ and $T_t = \inf\{s \le t : X(s) = M_t\}$. We will have use for the joint distribution of M_t, T_t, and $X(t)$. This was derived by Fitzsimmons [7] and Csáki et al [5].

Theorem 1 *For $x \ge y, b \ge y$, and $0 \le s \le t$,*

$$P_b(M_t \in dy, X(t) \in dx, T_t \in ds) = \\ f(s;b,y)f(t-s;x,y)S(dy)m(x)\,dx\,ds.$$

With the exception of Lemma 10, we will be concerned only with the minimum over the unit interval, and will then write M, T instead of M_1, T_1.

Let Q_t^y be the semigroup of excursions above the level y; this semigroup has density $q^y(t;x,z)$ with respect to the speed measure:

$$q^y(t;x,z)m(z)dz = P_x(X(t) \in dz, \tau_y > t).$$

The density is jointly continuous, and symmetric in $x > y, z > y$. The following formula is derived in Csáki et al [5].

$$\hat{q}_\alpha^y(x,z) \triangleq \int_0^\infty e^{-\alpha t} q^y(t;x,z)\,dt = g(x,z) - \frac{g_2(x)}{g_2(y)} g(y,z). \tag{7}$$

Finally, we show that under our assumptions on b and σ, X is "locally Brownian".

Lemma 2 *For any $x_0 \in \mathbb{R}$, under P_{x_0}, the processes defined by*

$$Y_n(t) = \frac{\sqrt{n}}{\sigma(x_0)}\left(X\left(\frac{t}{n}\right) - x_0\right)$$

converge in distribution to standard Brownian motion.

Proof: Because of our assumptions (2) and (3) on b and σ, the stochastic integral equation

$$X(t) = x_0 + \int_{s=0}^{t} b(X(s))\,ds + \int_{s=0}^{t} \sigma(X(s))dB_s \tag{8}$$

has a weak solution, unique in the sense of probability law, for each $x_0 \in \mathbb{R}$ (see Karatzas and Shreve [9], Theorem 5.2.9); equivalently, the time homogeneous martingale problem associated with the coefficients is well posed. Therefore, there exists a Brownian motion B such that X is the solution of (8), and so

$$\begin{aligned} Y_n(t) &= \frac{\sqrt{n}}{\sigma(x_0)}\left\{\int_{s=0}^{t/n} b(X(s))\,ds + \int_{s=0}^{t/n} \sigma(X(s))dB_s\right\} \\ &= \frac{\sqrt{n}}{\sigma(x_0)}\left\{\int_{u=0}^{t} b(X(u/n))\,\frac{du}{n} + \int_{u=0}^{t} \sigma(X(u/n))dB_{u/n}\right\} \\ &= \int_{u=0}^{t} \frac{1}{\sqrt{n}\sigma(x_0)} b(n^{-1/2}Y_n(u) + x_0)\,du + \\ &\qquad \int_{u=0}^{t} \frac{1}{\sigma(x_0)}\sigma(n^{-1/2}Y_n(u) + x_0)\sqrt{n}\,dB_{u/n}. \end{aligned}$$

Since $\sqrt{n}B(u/n)$ is a Brownian motion, this shows that the generator of Y_n is given by

$$\mathcal{A}_n f(y) = \frac{1}{2}\frac{\sigma^2(x_0 + y/\sqrt{n})}{\sigma^2(x_0)} f''(y) + \frac{1}{\sqrt{n}\sigma(x_0)} b(x_0 + y/\sqrt{n}) f'(y). \tag{9}$$

The coefficients of f'' and f' in (9) converge to 1/2 and 0, respectively, both uniformly on compact intervals. Therefore, by Theorem 11.1.4 of Stroock and Varadhan [13], Y_n converges in distribution to standard Brownian motion. ∎

3 Path Decomposition

Our goal in this section is to prove a weak convergence result for X in a neighborhood of the global minimizer. This result will be similar to that obtained previously for Brownian motion, in that suitably scaled, the process to either side of the minimizer looks in the limit like a "two-sided Bessel process". We start with a path decomposition result. The following notation, and Theorem 3 below, are from Fitzsimmons [7].

Let Ω be the set of continuous functions $\omega : \mathbb{R}^+ \to \mathbb{R}$ and Ω_t the continuous functions $\omega : [0,t] \to \mathbb{R}$. Define the coordinate mappings $X(t) : \Omega \to \mathbb{R}$ by $X(t,\omega) = \omega(t)$. For each $t \geq 0$ let $\mathcal{F}_t = \sigma(X(s); 0 \leq s \leq t)$, and $\mathcal{F} = \vee_{t \geq 0} \mathcal{F}_t$. For $0 < l < 1$ and $y < x$, define the law $\widehat{K}$ on $(\Omega_l, \mathcal{F}_l)$ by the absolute probability and transition probabilities

$$\widehat{K}^{y,l,x}(X(t) \in dz) = \frac{q^y(l-t; z, x) f(t; z, y)}{f(l; x, y)} m(z)dz, \qquad t < l,\ z > y, \quad (10)$$

and

$$\widehat{K}^{y,l,x}(X(t+v) \in dw | X(t) = z) = \frac{q^y(v; z, w) q^y(l-t-v; w, x)}{q^y(l-t; z, x)} m(w)dw \quad (11)$$

for $t + v < l,\ w > y$. It can be shown that $\widehat{K}^{y,l,x}$ is the law of the diffusion X starting from y conditioned to always stay above y, and conditioned to be killed at time l at level x.

Let $K^{y,l,x}$ denote the image of $\widehat{K}^{y,l,x}$ under the time reversal mapping γ_l on Ω_l that takes $\omega \in \Omega_l$ into the path $\gamma_l \omega$ defined by $(\gamma_l \omega)(t) = \omega(l-t)$.

Theorem 3 *Under P_{x_0}, the path fragments $(X(t) : 0 \leq t < T)$ and $(X(T+u) : 0 \leq u < 1-T)$ are conditionally independent given $(M, T, X(1))$. Given $(M = y, T = u, X(1) = x)$ $(0 < u < 1, y < x)$, the above processes have the conditional laws $K^{x_0,u,y}$ and $\widehat{K}^{y,t-u,x}$, respectively.*

Let $P_{x_0,x_1,y,t}$ be a regular version of the conditional probability $P_{x_0}(\cdot \,|\, M = y, T = t, X(1) = x_1)$, and denote the corresponding expectation by $E_{x_0,x_1,y,t}$.

The following result will be key in showing convergence of the densities of the scaled path components.

Lemma 4 *For $x > 0$,*

$$\lim_{h \downarrow 0} \frac{1}{h} \frac{1}{2} \sigma^2(y) q^y(t; y+x, y+h) m(y) = f(t; y+x, y). \qquad (12)$$

Proof: Taking Laplace transforms gives

$$\begin{aligned}
&\int_0^\infty e^{-\alpha t} \frac{1}{h} \frac{1}{2} \sigma^2(y) q^y(t; y+x, y+h) m(y)\, dt \\
&= \frac{1}{h} \frac{1}{2} \sigma^2(y) m(y) \left\{ g(y+x, y+h) - \frac{g_2(y+x)}{g_2(y)} g(y, y+h) \right\} \\
&= \frac{1}{h} \frac{1}{2} \sigma^2(y) \frac{1}{B} m(y) \left\{ g_2(y+x) g_1(y+h) - \frac{g_2(y+x)}{g_2(y)} g_1(y) g_2(y+h) \right\} \\
&= \frac{1}{h} \frac{1}{2} \sigma^2(y) m(y) \frac{1}{B} \frac{g_2(y+x)}{g_2(y)} \{ g_2(y) g_1(y+h) - g_1(y) g_2(y+h) \} \\
&= \frac{1}{h} \frac{1}{2} \sigma^2(y) m(y) \frac{1}{B} \frac{g_2(y+x)}{g_2(y)} \times \\
&\qquad \{ g_2(y)[g_1(y+h) - g_1(y)] - g_1(y)[g_2(y+h) - g_2(y)] \} \\
&= \frac{1}{2} \sigma^2(y) m(y) \frac{S(y+h) - S(y)}{h} \frac{g_2(y+x)}{g_2(y)} \times \\
&\qquad \frac{1}{B} \left\{ g_2(y) \frac{g_1(y+h) - g_1(y)}{S(y+h) - S(y)} - g_1(y) \frac{g_2(y+h) - g_2(y)}{S(y+h) - S(y)} \right\} \\
&\qquad\qquad \to \frac{1}{2} \sigma^2(y) m(y) S'(y) \frac{g_2(y+x)}{g_2(y)} = \frac{g_2(y+x)}{g_2(y)}.
\end{aligned}$$

The last equality is by (4). Inverting the Laplace transform (using (6)) gives the result. ∎

Theorem 5 *For $A > 0$, under $\widehat{K}^{y,l,x}$,*

$$Y_n(t) = \frac{\sqrt{n}}{\sigma(y)} \left(X(t/n) - y \right), \quad 0 \le t \le l,$$

converges in distribution in $C([0,A])$ to a 3-dimensional Bessel process as $n \to \infty$.

Proof: By a suitable change of variables, the transition density of Y_n is given by

$$p_n(t,z;t+v,w) = \frac{\sigma}{\sqrt{n}} \frac{q^y(\frac{v}{n}; y + \frac{\sigma z}{\sqrt{n}}; y + \frac{\sigma w}{\sqrt{n}}) q^y(l - \frac{t+v}{n}; y + \frac{\sigma w}{\sqrt{n}}, x)}{q^y(l - \frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}, x)} m(w),$$

where we have written σ for $\sigma(y)$. By Lemma 4,

$$\lim_{n\to\infty} \frac{q^y(l - \frac{t+v}{n}; y + \frac{\sigma w}{\sqrt{n}}, x)}{q^y(l - \frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}, x)} = \frac{w}{z}.$$

By Lemma 2,

$$\frac{\sigma}{\sqrt{n}} q^y \left(\frac{v}{n}; y + \frac{\sigma z}{\sqrt{n}}; y + \frac{\sigma w}{\sqrt{n}}\right) m(w)$$

converges to the transition density from $z > 0$ to $w > 0$ in time v of a Brownian excursion above 0, which is

$$p_0(v;z,w) = \frac{1}{\sqrt{2\pi v}} \left(\exp\left(-(z-w)^2/2v\right) - \exp\left(-(z+w)^2/2v\right)\right).$$

This shows that $p_n(t,z;t+v,w)$ converges to $p_0(v;z,w)w/z$, which is the transition density of the 3-dimensional Bessel process.

We apply a similar analysis to the entrance law of Y_n, which has density with respect to Lebesgue measure

$$\begin{aligned}\lambda_n(t,z) &= \frac{\sigma}{\sqrt{n}} \frac{q^y(l - \frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}; x) f(\frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}, y)}{f(l;x,y)} m(z) \\ &= \frac{[\frac{\sqrt{n}}{\sigma z}\frac{1}{2}\sigma^2 q^y(l - \frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}; x) m(z)] \cdot [\frac{2z}{n} f(\frac{t}{n}; y + \frac{\sigma z}{\sqrt{n}}, y)]}{f(l;x,y)}.\end{aligned}$$

By Lemma 4 the first term in the numerator converges to the denominator. Again by Lemma 2,

$$\frac{1}{n} f\left(\frac{t}{n}; y + \frac{z\sigma}{\sqrt{n}}, y\right) \to \frac{z}{t} \frac{1}{\sqrt{2\pi t}} \exp(-z^2/2t),$$

the first hitting time density for standard Brownian motion. Combining the results, we conclude that

$$\lambda_n(t,z) \to \frac{2z^2}{t}\frac{1}{\sqrt{2\pi t}}\exp(-z^2/2t),$$

which is the entrance law of the 3-dimensional Bessel process. By Scheffé's theorem, the entrance and transition laws converge to the corresponding laws for the 3-dimensional Bessel process, and so we have shown that the finite-dimensional distributions converge to the finite-dimensional distributions of the Bessel process.

It remains to establish tightness of the probabilities to conclude weak convergence. Applying Kolmogorov's criterion to our setting (Revuz and Yor [10], p. 474), it suffices to find strictly positive constants α, β, γ such that for every n and $s, t \geq 0$,

$$E\,|Y_n(t) - Y_n(s)|^\alpha \leq \beta|t-s|^{1+\gamma}. \tag{13}$$

In terms of the unscaled process X, (13) is equivalent to

$$E\,|X(t) - X(s)|^\alpha \leq \beta n^{\gamma+1-\alpha/2}|t-s|^{1+\gamma} \tag{14}$$

for all n. We must therefore have $\gamma + 1 \geq \alpha/2$, and it is sufficient for our needs if there exists an $\alpha > 2$ and a constant C such that

$$E\,|X(t) - X(s)|^\alpha \leq C|t-s|^{\alpha/2}. \tag{15}$$

This can be established using the transition densities directly. For example, taking $s = 0$ and using (10),

$$E\,X(t)^\alpha = \int_{z=y}^{\infty} (z-y)^\alpha \frac{q^y(l-t;z,x)f(t;z,y)}{f(l;x,y)} m(z)\,dz.$$

By (7),

$$\lim_{z\uparrow\infty} q^y(l-t;z,x) = \lim_{z\downarrow y} q^y(l-t;z,x) = 0,$$

so by the joint continuity of q^y,

$$\sup_{0\leq t\leq l,\ z\geq y} \frac{q^y(l-t;z,x)}{f(l;x,y)} = \kappa < \infty.$$

Therefore,

$$\begin{aligned} E\,X(t)^\alpha &\leq \kappa \int_{z=y}^{\infty} (z-y)^\alpha f(t;z,y) m(z)\,dz \\ &\leq \kappa \int_{z=y}^{\infty} (z-y)^\alpha p(t;y,z) m(z)\,dz \end{aligned}$$

where we have used the symmetry of $p(t;\cdot,\cdot)$. The result now follows from (3) and, for example, Problem 3.15, p. 306, of Karatzas and Shreve [9]. ∎

Let R_1 and R_2 be two independent 3-dimensional Bessel processes, and define a "two-sided Bessel process" $\widehat{R}$ by

$$\widehat{R}(t) = \begin{cases} R_1(t) & \text{if } t \geq 0, \\ R_2(-t) & \text{if } t \leq 0. \end{cases} \tag{16}$$

Theorem 6 *Conditioning on $T = t, M = y$, let*

$$Y_n^+(s) = \frac{\sqrt{n}}{\sigma(y)} \left(X\left(\min(t + s/n, 1)\right) - y \right),$$

and

$$Y_n^-(s) = \frac{\sqrt{n}}{\sigma(y)} \left(X\left(\max(t - s/n, 0)\right) - y \right),$$

and set

$$\widehat{Y}_n(s) = \begin{cases} Y_n^+(s) & \text{if } s \geq 0, \\ Y_n^-(-s) & \text{if } s < 0. \end{cases}$$

For fixed $A > 0$, under $P_{x_0,x_1,y,t}$,

$$\widehat{Y}_n \xrightarrow{\mathcal{D}} \widehat{R}$$

in $C\left([-A, A]\right)$ as $n \to \infty$.

Proof: This is an immediate consequence of Theorems 3 and 5. ∎

4 Approximation of the Minimum

In this section we will investigate asymptotic error associated with several non-adaptive methods of approximating the global minimum of a diffusion path. We begin by considering random methods; i.e., the observation points are chosen independently according to a fixed probability density.

Let h be a probability density on $[0, 1]$. We suppose that the minimum is approximated by

$$M_n = \min_{1 \le i \le n} X(t_i),$$

where the t_i are independent with density h. Set $\Delta_n \triangleq M_n - M$, the error after n observations. If the support of h is $[0, 1]$, then Δ_n converges to 0 almost surely, and we confine our attention to such h.

We first derive a limit result describing the observation process. We will use this with the result on the behavior of the process near the global minimum established in the last section. Since we are considering only non-adaptive algorithms, the two processes are independent.

Suppose that $t_1, t_2, \ldots, t_n$ are independent with density h, where h is a smooth density. Fix $t \in (0, 1)$, and define the counting processes

$$\begin{aligned} N_n^{t,+}(s) &= \#\{1 \le i \le n : t \le t_i \le t + \min(s, 1-t)\} \\ N_n^{t,-}(s) &= \#\{1 \le i \le n : t - \min(s, t) \le t_i \le t\} \end{aligned}$$

for $s \ge 0$. Note that $N_n^{t,+}(s) \sim Binomial(n, \int_{u=t}^{t+s} h(u)\, du)$ for $0 \le s \le 1 - t$.

Lemma 7 *As* $n \to \infty$,

$$\left(N_n^{t,-}\left(\frac{\cdot}{nh(t)}\right), N_n^{t,+}\left(\frac{\cdot}{nh(t)}\right) \right) \xrightarrow{\mathcal{D}} (N^-, N^+),$$

where N^- *and* N^+ *are independent Poisson processes with unit intensity.*

Proof: The finite-dimensional distributions are multinomial and are easily shown to converge to the appropriate limits. Since all the processes are step functions, the result follows. ∎

We will also need the following lemma to prove the main result.

Lemma 8 *Let $\{X(t) : t \geq 0\}$ be a 3-dimensional Bessel process, and define an independent Poisson process with intensity 1 and points of increase $\{T_1, T_2, T_3, \dots\}$. Set $Z = \min_{i\geq 1} X(T_i)$. Then*

$$P(Z \leq y) = \frac{\left(1 - e^{-\sqrt{2}y}\right)^2}{1 + e^{-2\sqrt{2}y}}. \tag{17}$$

Proof: Let $L_y = \sup\{t : X(t) = y\}$. Since X is transient, $L_y < \infty$ a.s. The process $Y(t) = X(L_y - t)$, $0 \leq t \leq L_y$ has the same law as $\{B(t) : 0 \leq t \leq T_0\}$, where B is a Brownian motion starting at y and run until it hits zero; see Williams [14]. The problem is therefore reduced to that of determining the law of the minimum of a Markov chain W_n that has the transition law of Brownian motion, sampled at exponentially distributed intervals and killed on hitting 0. Specifically, let r be the transition function of W. Then

$$\begin{aligned} r(y,z) &= \int_{t=0}^{\infty} e^{-t} P_y\left(B(t) \in dz,\, T_0 > t\right) dt \\ &= \int_{t=0}^{\infty} e^{-t} \frac{1}{\sqrt{2\pi t}} \left[\exp\left(-\frac{(y-z)^2}{2t}\right) - \exp\left(-\frac{(y+z)^2}{2t}\right)\right] dt \\ &= \begin{cases} \sqrt{2}\exp(-\sqrt{2}z)\sinh(\sqrt{2}y) & \text{if } 0 < y < z, \\ \sqrt{2}\exp(-\sqrt{2}y)\sinh(\sqrt{2}z) & \text{if } y > z > 0. \end{cases} \end{aligned}$$

Let $V(y) = P(Z \leq y)$, $y > 0$, and let $\tau \sim Exp(1)$. Then

$$V(y) = P_y\left(B(\tau) \in (0,y]\right) + P_y\left(B(\tau) > y\right) V(y), \tag{18}$$

and so

$$V(y) = \frac{\int_{z=0}^{y} r(y,z)dz}{1 - \int_{z=y}^{\infty} r(y,z)dz} = \frac{\left(1 - e^{-\sqrt{2}y}\right)^2}{1 + e^{-2\sqrt{2}y}}. \tag{19}$$

∎

Let Δ be a non-negative random variable that has the distribution

$$P(\Delta \leq z) = \tanh^2(\sqrt{2}z), \quad z \geq 0.$$

This random variable, which has mean $2^{-1/2}$, will appear as the limit of the normalized approximation errors under random observations.

Denote by $g_{x_0}(y,t)$ the joint density of (M,T) under P_{x_0}, which is given by Theorem 1.

Theorem 9 *If observations are chosen independently according to the smooth density h, then under $P_{x_0,x_1,y,t}$,*

$$\sqrt{n}\frac{\sqrt{h(t)}}{\sigma(y)}\Delta_n \xrightarrow{\mathcal{D}} \Delta \tag{20}$$

and

$$\sqrt{n}E_{x_0,x_1,y,t}(\Delta_n) \to \frac{\sigma(y)}{\sqrt{2h(t)}} \tag{21}$$

as $n \to \infty$. For $z \geq 0$,

$$P_{x_0}(\sqrt{n}\Delta_n \leq z) \to \int_{y \leq x_0} \int_{t \in [0,1]} \tanh^2\left(\sqrt{2}\frac{\sqrt{h(t)}}{\sigma(y)}z\right) g_{x_0}(t,y)\, dt\, dy. \tag{22}$$

In particular, if $\sigma(y) = \sigma$ and $h(t) = 1$ are constant, then under P_{x_0},

$$\sqrt{n}\frac{1}{\sigma}\Delta_n \xrightarrow{\mathcal{D}} \Delta. \tag{23}$$

Combining (20) and (21), we observe that under $P_{x_0,x_1,y,t}$,

$$\frac{\Delta_n}{2E_{x_0,x_1,y,t}(\Delta_n)} \xrightarrow{\mathcal{D}} \Delta \tag{24}$$

as $n \to \infty$. The unconditional version of (24) holds in the case of constant σ and h.

The proof of Theorem 9 will make use of the following two lemmas.

Lemma 10 *For* $0 < a < b$ *and* $y < x$,

$$\lim_{n\to\infty} \sqrt{n} P_{y+b/\sqrt{n}} \left(M_{t-A/n} \in (y, y + a/\sqrt{n}), X(t - A/n) \in dx \right) 1_{(t>A/n)} = a\, m(dx) S'(y) f(t; x, y). \quad (25)$$

Proof: Using Theorem 1 and the expression for the Laplace transforms of the hitting time densities (6), the Laplace transform of the left hand side of (25) can be expressed as

$$\begin{aligned}
&\int_{t=A/n}^{\infty} e^{-\alpha t} \sqrt{n} P_{y+b/\sqrt{n}} \left(M_{t-A/n} \in (y, y + a/\sqrt{n}), X(t - A/n) \in dx \right) dt \\
&= \int_{t=A/n}^{\infty} e^{-\alpha t} \sqrt{n} \int_{z=y}^{y+a/\sqrt{n}} \int_{u=0}^{t-A/n} f(u; y + b/\sqrt{n}, z) \\
&\qquad \cdot f(t - A/n - u; x, z) S(dz)\, du\, m(dx)\, dt \\
&= \sqrt{n}\, m(dx) \int_{z=y}^{y+a/\sqrt{n}} \int_{t=A/n}^{\infty} e^{-\alpha t} \int_{u=0}^{t-A/n} f(u; y + b/\sqrt{n}, z) \\
&\qquad \cdot f(t - A/n - u; x, z) S(dz)\, du\, dt \\
&= \sqrt{n}\, m(dx) \int_{z=y}^{y+a/\sqrt{n}} \int_{s=0}^{\infty} e^{-\alpha(s+A/n)} \int_{u=0}^{s} f(u; y + b/\sqrt{n}, z) \\
&\qquad \cdot f(s - u; x, z) S(dz)\, du\, ds \\
&= \sqrt{n}\, m(dx)\, e^{-\alpha A/n} \int_{z=y}^{y+a/\sqrt{n}} \frac{g_2^\alpha(y + b/\sqrt{n})}{g_2^\alpha(z)} \frac{g_2^\alpha(x)}{g_2^\alpha(z)} S(dz) \\
&= \sqrt{n}\, m(dx)\, e^{-\alpha A/n} g_2^\alpha(x) g_2^\alpha(y + b/\sqrt{n}) \int_{z=y}^{y+a/\sqrt{n}} \frac{S(dz)}{(g_2^\alpha(z))^2} \\
&= a\; m(dx)\, e^{-\alpha A/n} g_2^\alpha(x) g_2^\alpha(y + b/\sqrt{n}) \frac{1}{a/\sqrt{n}} \int_{z=y}^{y+a/\sqrt{n}} \frac{S(dz)}{(g_2^\alpha(z))^2}.
\end{aligned}$$

Taking the limit as $n \to \infty$ gives

$$a\, m(dx) S'(y) \frac{g_2^\alpha(x)}{g_2^\alpha(y)},$$

which is the Laplace transform of the right hand side of (25). ∎

Let

$$\delta_n^A = \min_{\frac{A}{n} \le t \le l} X(t) - y.$$

Lemma 11 *For* $0 < a < \infty$,

$$\lim_{A \uparrow \infty} \limsup_{n \to \infty} \hat{K}^{y,l,x} \left(\sqrt{n} \delta_n^A \le a \right) = 0.$$

Proof: For $l > 0$ and $x, z \in \mathbb{R}$, let $P_z^{l,x}$ be the law of the diffusion bridge from z at time 0 to x at time l (see Salminen [12]); i.e., $P_z^{l,x}$ is a regular conditional P_z distribution of $(X(s) : 0 \le s \le l)$ given $(X(l) = x)$. Then

$$\begin{aligned}
&\hat{K}^{y,l,x} \left(\sqrt{n} \delta_n^A \le a \right) \\
&= \int_{z=y}^{\infty} \hat{K}^{y,l,x} \left(\sqrt{n} \delta_n^A \le a \,|\, X(A/n) = z \right) \hat{K}^{y,l,x} \left(X(A/n) \in dz \right) \\
&= \int_{z=0}^{\infty} \hat{K}^{y,l,x} \left(\delta_n^A \le y + a/\sqrt{n} \,|\, X(A/n) = y + z/\sqrt{n} \right) \\
&\qquad \cdot \hat{K}^{y,l,x} \left(\sqrt{n} \left(X(A/n) - y \right) \in dz \right) \\
&= \int_{z=0}^{\infty} P_{y+z/\sqrt{n}}^{l-A/n,x} \left(M_{l-A/n} \le y + a/\sqrt{n} \,|\, M_{l-A/n} > y \right) \\
&\qquad \hat{K}^{y,l,x} \left(\sqrt{n} \left(X(A/n) - y \right) \in dz \right) \\
&= \int_{z=0}^{\infty} \frac{P_{y+z/\sqrt{n}}^{l-A/n,x} \left(M_{l-A/n} \in (y, y + a/\sqrt{n}] \right)}{P_{y+z/\sqrt{n}}^{l-A/n,x} \left(M_{l-A/n} \in (y, y + z/\sqrt{n}] \right)} \\
&\qquad \cdot \hat{K}^{y,l,x} \left(\sqrt{n} \left(X(A/n) - y \right) \in dz \right) \\
&= \int_{z=0}^{\infty} \frac{P_{y+z/\sqrt{n}} \left(M_{l-A/n} \in (y, y + a/\sqrt{n}],\, X(l - A/n) \in dx \right)}{P_{y+z/\sqrt{n}} \left(M_{l-A/n} \in (y, y + z/\sqrt{n}],\, X(l - A/n) \in dx \right)} \\
&\qquad \cdot \hat{K}^{y,l,x} \left(\sqrt{n} \left(X(A/n) - y \right) \in dz \right) \\
&= \int_{z=0}^{a} \hat{K}^{y,l,x} \left(\sqrt{n} \left(X(A/n) - y \right) \in dz \right)
\end{aligned}$$

$$+\int_{z=a}^{\infty} \frac{P_{y+z/\sqrt{n}}\left(M_{l-A/n} \in (y, y+a/\sqrt{n}],\ X(l-A/n) \in dx\right)}{P_{y+z/\sqrt{n}}\left(M_{l-A/n} \in (y, y+z/\sqrt{n}],\ X(l-A/n) \in dx\right)} \hat{K}^{y,l,x}\left(\sqrt{n}\,(X(A/n)-y) \in dz\right). \tag{26}$$

The second to last equation follows from the fact that for $A \in \sigma\{X(s) : 0 \le s \le l\}$,

$$P_z(A, X(l) \in dx) = P_z^{l,x}(A) P_z(X(l) \in dx).$$

By Theorem 5, under $\hat{K}^{y,l,x}$,

$$\sqrt{n}\,(X(A/n) - y) \xrightarrow{\mathcal{D}} \sigma(y) R(A)$$

as $n \to \infty$, where R is a 3-dimensional Bessel process. By Lemma 10, the ratio in the integrand in the last line of (26) converges to a/z, and so

$$\hat{K}^{y,l,x}\left(\sqrt{n}\delta_n^A \le a\right) \to P(\sigma(y)R(A) \le a) + E\left(\frac{a}{\sigma(y)R(A)}; \sigma(y)R(A) \ge a\right).$$

As $A \uparrow \infty$ the last two expressions each converge to 0, which completes the proof. ∎

Proof of Theorem 9: Conditioning on $M = y, T = t$, and $X(1) = x$, let Δ_n^+ represent the error to the right of t; the analysis of the error to the left is similar. If N_n^+ is the number of the first n observations that fall to the right of t, then

$$\sqrt{n}\frac{\sqrt{h(t)}}{\sigma(y)}\Delta_n^+ = \sqrt{n}\frac{\sqrt{h(t)}}{\sigma(y)} \min\{X(t+\tau_1^n) - y, \ldots, X(t+\tau_{N_n^+}^n) - y\},$$

where the $t + \tau_i^n$'s represent the ordered observations to the right of t. Using the scaled process Y_n defined in Theorem 5 (and noting that the definition of Y_n makes sense for positive n not necessarily integer), the above becomes

$$\sqrt{n}\frac{\sqrt{h(t)}}{\sigma(y)}\Delta_n^+ = \min\left\{Y_{nh(t)}\left(\tau_1^n\, nh(t)\right), \ldots, Y_{nh(t)}\left(\tau_{N_n^+}^n\, nh(t)\right)\right\}.$$

Fix $A > 0$ and let

$$Z_n = \min\left\{Y_{nh(t)}\left(\tau_1^n\, nh(t)\right), \ldots, Y_{nh(t)}\left(\tau_{N_n^+}^n\, nh(t)\right)\right\},$$

$$Z_n^A = \min_{\tau_i^n \le A(nh(t))^{-1}} \left\{Y_{nh(t)}\left(\tau_i^n\, nh(t)\right)\right\},$$

and

$$\delta_n^A = \min\{X(s) - y : t + \frac{A}{nh(t)} \le s \le 1\}.$$

Let Y be a 3-dimensional Bessel process, and $\{T_i : i \ge 1\}$ be the jump times of an independent Poisson process with rate 1. Set

$$Z^A = \min_{T_i \le A} \{Y(T_i)\}.$$

Then for $z > 0$, with Z as defined in Lemma 8,

$$\begin{aligned} |P(Z_n \le z) - P(Z \le z)| \;\le\; & |P(Z_n \le z) - P(Z_n^A \le z)| \\ + \; & |P(Z_n^A \le z) - P(Z^A \le z)| \\ + \; & |P(Z^A \le z) - P(Z \le z)|. \end{aligned}$$

Now,

$$|P(Z_n \le z) - P(Z_n^A \le z)| \le P(\sqrt{n}\delta_n^A \le z) \to 0$$

as n, and then $A \uparrow \infty$,by Lemma 11. Since $Y_{nh(t)}$ converges to a 3-dimensional Bessel process, the arguments of $Y_{nh(t)}$ converge to the jump times of a Poisson process by Lemma 7, and $N_n^+ \to \infty$ a.s., we conclude from Lemma 8 that

$$|P(Z_n^A \le z) - P(Z^A \le z)| \to 0$$

as $n \to \infty$. Since $Z^A \xrightarrow{\mathcal{D}} Z$ as $A \uparrow \infty$,

$$|P(Z^A \le z) - P(Z \le z)| \to 0.$$

Therefore, $Z_n \xrightarrow{\mathcal{D}} Z$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} P\left(\sqrt{n}\frac{\sqrt{h(t)}}{\sigma(y)}\Delta_n^+ \le z\right) = \frac{\left(1 - e^{-\sqrt{2}z}\right)^2}{1 + e^{-2\sqrt{2}z}}.$$

Since $\Delta_n = \min\{\Delta_n^+, \Delta_n^-\}$, where Δ_n^+ and Δ_n^- have the same distribution and are independent, (20) follows.

The limit in (21) follows from the uniform integrability of $\{\sqrt{n}\Delta_n : n \geq 1\}$ under $P_{x_0,x_1,y,t}$. To show uniform integrability, it is enough to establish that

$$\sup_n E_{x_0,x_1,y,t}\left(n\Delta_n^2\right) < \infty. \tag{27}$$

But $\Delta_n \leq X(t_*^n) - X(t)$, where t_*^n is the closest observation to the right of t. Let γ_n be the length of the observation gap straddling t. Then

$$\sup_n E_{x_0,x_1,y,t}\left(n\Delta_n^2\right) \leq \sup_n E_{x_0,x_1,y,t}\left(X^2(\gamma_n)\right), \tag{28}$$

where X has law $\hat{K}^{y,l,x_1}$. By (14), there is a constant C such that

$$\sup_n E_{x_0,x_1,y,t}\left(X^2(\gamma_n)\right) \leq \sup_n nCE(\gamma_n) < \infty, \tag{29}$$

thus confirming uniform integrability of the family $\{\sqrt{n}\Delta_n : n \geq 1\}$ under $P_{x_0,x_1,y,t}$.

Finally, (22) is obtained from

$$\begin{aligned} &P_{x_0}(\sqrt{n}\Delta_n \leq z) \\ &\quad = \int_{y,t,x_1} P_{x_0,x_1,y,t}\left(\frac{\sqrt{nh(t)}}{\sigma(y)}\Delta_n \leq \frac{\sqrt{h(t)}}{\sigma(y)}z\right) \times \\ &\qquad\qquad P_{x_0}(M \in dy, T \in dt, X(1) \in dx_1) \\ &\quad \to \int_{y\in\mathbb{R}}\int_{t\in[0,1]} \tanh^2\left(\sqrt{2}\frac{\sqrt{h(t)}}{\sigma(y)}z\right) g_{x_0}(t,y)\,dt\,dy. \end{aligned}$$

∎

It is natural to consider the problem of finding the optimal observation density h in the sense of minimizing the asymptotic normalized expected error. Recall that $g_{x_0}(y,t)$ is the joint P_{x_0} density of (M,T). If the $\{\sqrt{n}\Delta_n : n \geq 1\}$ are uniformly integrable under P_{x_0}, then

$$E_{x_0}\left(\sqrt{n}\Delta_n\right) \to \frac{1}{\sqrt{2}}\int_y\int_{t=0}^1 \frac{\sigma(y)}{\sqrt{h(t)}} g_{x_0}(y,t)\,dt\,dy, \tag{30}$$

the mean of the limiting distribution given by (22). This holds for Brownian motion, for example. Assuming that (30) holds, then the optimization problem is the following calculus of variations problem: Choose a smooth density h to minimize

$$\int_y \int_{t=0}^1 \frac{\sigma(y)}{\sqrt{h(t)}} g_{x_0}(y,t)\, dt\, dy = \int_{t=0}^1 \frac{1}{\sqrt{h(t)}} \int_y \sigma(y) g_{x_0}(y,t) dy\, dt.$$

The optimal density is given by

$$h(t) = \frac{\left\{\int_y \sigma(y) g_{x_0}(y,t) dy\right\}^{2/3}}{\int_{s=0}^1 \left\{\int_y \sigma(y) g_{x_0}(y,s) dy\right\}^{2/3} ds}.$$

In particular, if $\sigma(\cdot)$ is constant, the optimal observation density is proportional to the density of the location of the minimizer raised to the power 2/3.

Next, we consider deterministic approximation schemes. The simplest is given by a uniform grid:

$$t_i = \frac{i-1}{n-1} \text{ for } i = 1, 2, \ldots, n. \tag{31}$$

The following is a straightforward adaptation of Theorem 1 of Asmussen et al [2] to the present situation.

Theorem 12 *If $\sigma(y) = \sigma$ is constant and if observations are chosen according to the uniform grid (31), then*

$$\frac{\sqrt{n}}{\sigma} \Delta_n \xrightarrow{\mathcal{D}} W \tag{32}$$

as $n \to \infty$, where

$$W = \min_{k \in \mathbb{Z}} \{\widehat{R}(U + k)\},$$

U is uniform$(0,1)$, independent of $\widehat{R}$, the two-sided Bessel process defined at (16).

While an explicit formula for the law of W is not available, the mean is given by

$$E(W) = \frac{1}{\sqrt{2\pi}}\zeta(1/2) \approx 0.5826, \tag{33}$$

where ζ is Riemann's zeta function.

We conclude by noting the implications of our results for some particular diffusions.

For Brownian motion, the error has been analyzed by several authors, including Ritter [11] and Al-Mharmah and Calvin [1]. The distribution of the location of the minimizer is arcsine, and the optimal sampling density is Beta(2/3, 2/3) (see Al-Mharmah and Calvin [1]). The corresponding limiting normalized mean error is

$$\sqrt{n}E\left(\Delta_n\right) \to \frac{1}{\pi\sqrt{2}}\mathcal{B}(2/3,2/3)^{3/2} \approx 0.662281,$$

where $\mathcal{B}$ is the beta function.

If P_0 is the law of Brownian motion starting from 0, then under $P_{0,0,y,t}$, the coordinate process is a Brownian bridge with minimum $X(t) = y$. From Theorem 9, under $P_{0,0,y,t}$

$$\sqrt{n}\sqrt{h(t)}\,\Delta_n \xrightarrow{\mathcal{D}} \Delta,$$

and

$$\sqrt{n}E_{0,0,y,t}(\Delta_n) \to \frac{1}{\sqrt{2h(t)}}.$$

Since the minimizer is uniformly distributed over the unit interval, this shows that the optimal sampling density is $h = 1$, i.e., the uniform distribution. The corresponding limiting normalized mean error for the Brownian bridge is

$$\sqrt{n}E\left(\Delta_n\right) \to \frac{1}{\sqrt{2}} \approx 0.707107.$$

For a diffusion with constant diffusion coefficient σ, the optimal sampling density is proportional to the density of the minimizer raised to the power 2/3. However, manageable expressions for the distribution of the minimizer are available for few diffusions.

5 Acknowledgements

The first author was supported by the National Science Foundation under grant DDM-9010770. The second author was supported by the National Science Foundation under grant DDM-9101580 and the Army Research Office under Contract No. DAAL03-91-G-0319.

References

[1] Al-Mharmah, H. and Calvin, J., Optimal random non-adaptive algorithm for optimization of Brownian motion. To appear in *J. Global Optim.*, 1994.

[2] Asmussen, S., Glynn, P. and Pitman, J., Discretization error in simulation of one-dimensional reflecting Brownian motion. Technical Report 94–1, Dept. of Operations Research, Stanford University, 1994.

[3] Calvin, J., Average performance of passive algorithms for global optimization. *J. Math. Anal. Appl.* **191** (1995), 608-617.

[4] Calvin, J. and Glynn, P. W., Average behavior of random search for the maximum. Submitted.

[5] Csáki, E., Földes, A. and Salminen, P., On the joint distribution of the maximum and its location for a linear diffusion. *Ann. Inst. Henri Poincarè* **23-2** (1987), 179–194.

[6] Ethier, S. and Kurtz, T., *Markov Processes: Characterization and Convergence.* Wiley, New York, 1986.

[7] Fitzsimmons, P. J., Excursions above the minimum for diffusions. Unpublished manuscript (1985).

[8] Itô, K. and McKean, H., *Diffusion Processes and Their Sample Paths.* Springer, Berlin, 1965.

[9] I. Karatzas and Shreve, S., *Brownian Motion and Stochastic Calculus*, Second Edition. Springer-Verlag, New York, 1991.

[10] Revuz, D. and Yor, M., *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin, 1991.

[11] Ritter, K., Approximation and optimization on the Wiener space. *J. Complexity* **6** (1990), 337-364.

[12] Salminen, P., Brownian excursions revisited. *Sem. Stochastic Processes 1983* (eds. E. Cinlar, K. L. Chung, R. K. Getoor). Birkhäuser, Boston, 1983.

[13] Stroock, D. W. and Varadhan, S. R. S., *Multidimensional Diffusion Processes.* Springer-Verlag, New York, 1979.

[14] Williams, D., Path decomposition and continuity of local time for one-dimensional diffusions, I. *Proc. London Math. Soc.* Ser. 3 **28** (1974), 738–68.

Received: 12/21/1994
Revised: 11/10/1995
Accepted: 12/11/1995