

On the minimum of a conditioned Brownian bridge

Aureli Alabert

Department of Mathematics
Universitat Autònoma de Barcelona
08193 Bellaterra, Catalonia
Aureli.Alabert@uab.cat

Ricard Caballero

Department of Mathematics
Universitat Autònoma de Barcelona
08193 Bellaterra, Catalonia
rcaballero@mat.uab.cat

March 15, 2017

Abstract

We study the law of the minimum of a Brownian bridge, conditioned to take specific values at specific points, and the law of the location of the minimum. They are used to compare some non-adaptive optimisation algorithms for black-box functions for which the Brownian bridge is an appropriate probabilistic model and only a few points can be sampled.

Keywords: Black-box optimisation, Brownian bridge, simulation.

Mathematics Subject Classification (2010): 90C26, 60J65, 65C05

1 Introduction

We study the law of the minimum of a Brownian bridge conditioned to pass through given points in the interval $[0, 1]$, and the location of this minimum. Our motivation is the investigation of the performance of algorithms based on probabilistic models in expensive black-box optimisation.

The probabilistic model point of view assumes the existence of a probability space from where the function at hand has been drawn. The choice of points to sample is guided by the probabilistic properties of this random function. Eventually, the values of the function at the points already sampled can be used to decide the next sampling point (*adaptive algorithms*) or can be neglected (*non-adaptive* or *passive algorithms*).

We assume here that the probabilistic model is completely specified, and given by the standard Brownian bridge on the interval $[0, 1]$; that means, the function to be optimised is a path of a standard Brownian motion process, conditioned to take certain values x_0 at $t = 0$ and x_1 at $t = 1$. More generally, one could set up a *statistical model* (a family of probabilistic models depending on some parameters) and improve sequentially the knowledge of the parameters using the values observed while sampling.

Probabilistic models try to account for heavy multimodality in the objective function. The irregularity and the independence of values over disjoint intervals of the Brownian bridge and other Markovian stochastic processes represent well this multimodality, although at a very local scale the functions found in practice are usually smooth.

Our main interest is in expensive black-box functions from which only a few points can be sampled, where it is more important to have an estimation of the absolute error incurred in approximating the true minimum than the convergence, the speed of convergence, or the complexity properties of the algorithm.

In this paper we establish some facts about the law of the minimum of a Brownian bridge on the interval $[0, 1]$, conditioned to hit some points $(t_i, x_i) \in (0, 1) \times \mathbb{R}$. The density function of the law can be computed exactly, but we argue that it is better to use simulation to obtain its features. We then use these simulations to evaluate empirically the performance of three simple non-adaptive algorithms when only

small samples are allowed. New adaptive algorithms in the same setting will be presented and compared elsewhere.

The Brownian bridge model in optimisation has been studied by several authors, from the point of view of the asymptotic properties of the algorithms (see, e.g. Locatelli [7], Ritter [9], Calvin [2, 3, 4]). We mention here just two facts:

1. *Long-run performance:* Sampling at n equidistant points and taking the value of the best sampled point as the approximation of the true minimum has an absolute error whose expectation is $O(1/\sqrt{n})$. The best adaptive algorithm is better than the best non-adaptive algorithm concerning improvement rates, but asymptotically both are $O(1/\sqrt{n})$. Thus, sampling at equidistant points is optimal in the long run.
2. *Complexity:* For algorithms using n function evaluations, the convergence to zero of the mean error cannot be $O(e^{-cn})$ for any constant c . (This convergence order is indeed attained in unimodal functions, for example by Fibonacci search.)

A general survey of probabilistic methods for optimisation can be found in Zhigljavsky and Žilinskas [12, ch. 4].

We establish some notations and preliminaries in Section 2. Section 3 is devoted to computing the probability that the minimum lies in a given interval determined by two of the conditioning points. In Section 4 we show how to simulate the law of the global minimum of the process. In Section 5 we test and compare three non-adaptive algorithms from the point of view of the expected difference between the best sampled point and the true minimum of the path, when the evaluation points are few. Finally, in Section 6, we compute the conditional distribution of the location of the minimum of a single Brownian bridge given the value of this minimum, and we show how to use it to simulate the location of the minimum of the whole process.

2 Preliminaries

In the sequel, for a given stochastic process $Z := \{Z_t, t \in I\}$, defined on a closed interval $I \subset \mathbb{R}$, we denote by

$$m(Z) := \min_{t \in I} Z_t \quad \text{and} \quad \theta(Z) := \arg \min_{t \in I} Z_t$$

the random variables giving the minimum value of Z and its location, respectively. In the cases we will treat here, the minimum exists and is unique with probability 1 but, to avoid any ambiguity, one can assume that $\theta(Z)$ is the first point where the minimum is achieved.

A standard Brownian motion W on the interval $[t_0, t_1]$, starting at (t_0, a) , $a \in \mathbb{R}$, is a Markov stochastic process with continuous paths, defined by the transition probability

$$p_{s,r}(x,y) = \frac{1}{\sqrt{2\pi(r-s)}} \exp \left\{ -\frac{(y-x)^2}{2(r-s)} \right\}, \quad t_0 \leq s < r \leq t_1,$$

and such that $W_{t_0} = a$ with probability 1.

A Brownian bridge B starting at (t_0, a) and ending at (t_1, b) has the law of a Brownian motion defined on the time interval $[t_0, t_1]$ starting at (t_0, a) and conditioned to take the value b at t_1 . The random variable B_t , $t_0 < t < t_1$, is Gaussian with mean $a + \frac{t-t_0}{t_1-t_0}(b-a)$ and variance $\frac{(t-t_0)(t_1-t)}{t_1-t_0}$.

The following results are known or easily deduced (see e.g. Karatzas and Shreve [6, Sec. 2.8]):

Proposition 2.1. *Let W be a Brownian motion starting at (t_0, a) , defined on the interval $[t_0, t_1]$. The density function of its minimum $m(W)$ is given by*

$$f_{m(W)}(y) = \sqrt{\frac{2}{\pi(t_1 - t_0)}} \exp\left\{\frac{-(a-y)^2}{2(t_1 - t_0)}\right\} \mathbf{1}_{\{y < a\}}. \quad (1)$$

Let B be a Brownian bridge from (t_0, a) to (t_1, b) . The density function of its minimum $m(B)$ is given by

$$f_{m(B)}(y) = \frac{2}{t_1 - t_0} (a + b - 2y) \exp\left\{\frac{-2(a-y)(b-y)}{t_1 - t_0}\right\} \mathbf{1}_{\{y < a, y < b\}}. \quad (2)$$

□

Given $0 = t_0 < t_1 < \dots < t_n < t_{n+1} = 1$, and real values x_0, \dots, x_{n+1} , we are interested in a stochastic process $X := \{X_t, t \in [0, 1]\}$ whose law is that of a Brownian bridge starting at (t_0, x_0) , ending at (t_{n+1}, x_{n+1}) , and conditioned to pass through all the intermediate points (t_i, x_i) , $i = 1, \dots, n$.

This process can be thought as the concatenation of $n + 1$ independent Brownian bridges $B^i := \{B_t^i, t \in [t_i, t_{i+1}]\}$, with respective end values x_i and x_{i+1} . In the optimisation application that we have in mind, the interior points t_1, \dots, t_n are the points sampled by the algorithm, and x_1, \dots, x_n are the observed values at those points.

The law of the minimum of the process X can be expressed in terms of the law of the minimum of its pieces, in the usual way. Despite the mutual independence of the Brownian bridges, this cannot be simplified further:

Proposition 2.2. *Let X be the conditioned Brownian bridge defined above, and $m(X)$ its minimum. Then, for all $y \in \mathbb{R}$,*

$$P\{m(X) > y\} = \prod_{i=0}^n \left(1 - \exp\left\{\frac{-2(x_{i+1} - y)(x_i - y)}{t_{i+1} - t_i}\right\}\right) \mathbf{1}_{\{y < \min(x_0, \dots, x_{n+1})\}}. \quad (3)$$

Proof. The formula comes from the standard computation of the law of the minimum of several independent random variables:

$$F_{m(X)}(y) = 1 - \prod_{i=0}^n (1 - F_{m(B^i)}(y)),$$

where $F_{m(X)}$ is the distribution function of $m(X)$, and $F_{m(B^i)}$ is the distribution function of the minimum of the Brownian bridge B^i , whose density is given by (2), adjusting the appropriate constants. □

Note that in the case when we do not condition to the end point (t_{n+1}, x_{n+1}) , we obtain a similar expression where, according to (1), the n -th factor in (3) is replaced by

$$1 - \int_{-\infty}^y \sqrt{\frac{2}{\pi(1 - t_n)}} \exp\left\{\frac{-(x_n - z)^2}{2(1 - t_n)}\right\} dz.$$

It would not be difficult to deal with this situation separately (a conditioned Brownian motion), but we will keep our assumptions for simplicity. Moreover, sampling at $t = 1$ reverts to our case.

It is natural to try to compute explicitly the density $f_{m(X)}$ of the minimum of X by conditioning to each of the intervals $[t_i, t_{i+1}]$:

$$f_{m(X)}(y) = \sum_{i=0}^n P\{\theta(X) \in [t_i, t_{i+1}]\} \cdot f_{m(X)|_{\theta(X) \in [t_i, t_{i+1}]}}(y).$$

Even though, as we will see, the probability of $\theta(X)$ lying in a given interval can be, in principle, computed exactly, the conditional densities in the second factors still depend on the process outside $[t_i, t_{i+1}]$; thus they are not simply densities of the minimum of a single Brownian bridge.

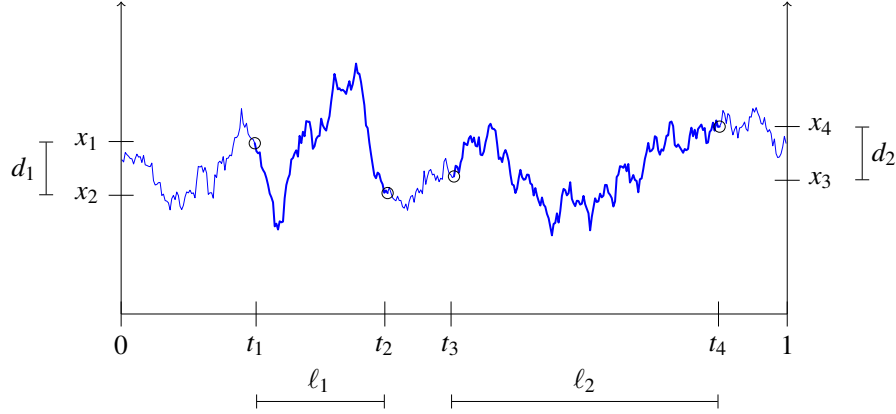


Figure 1: A path of Brownian motion conditioned to the circled points. Which of the intervals $[t_1, t_2]$ and $[t_3, t_4]$ is more likely to attain the lowest value?

3 Probability that $\theta(X)$ belong to $[t_i, t_{i+1}]$

3.1 Analytical formulae

The probability that the minimum of X is achieved in one of the intervals $[t_i, t_{i+1}]$ can be computed exactly:

Proposition 3.1. *The probability that the minimum of the process X is located in the interval $[t_i, t_{i+1}]$ is given by:*

$$P\{\theta(X) \in [t_i, t_{i+1}]\} = \int_{-\infty}^{\min(x_0, \dots, x_{n+1})} \frac{2}{t_{i+1} - t_i} (x_i + x_{i+1} - 2y) \exp\left\{\frac{-2(x_i - y)(x_{i+1} - y)}{t_{i+1} - t_i}\right\} \times \prod_{j \neq i} \left(1 - \exp\left\{\frac{-2(x_j - y)(x_{j+1} - y)}{t_{j+1} - t_j}\right\}\right) dy. \quad (4)$$

Proof. The random variables $m(B^0), \dots, m(B^n)$ are independent, because of the Markov property of Brownian motion. Therefore, their joint density is given by the product $\prod_{i=0}^n f_{m(B^i)}(y_i)$, where $f_{m(B^i)}$ is the density of the minimum of the i -th bridge. Denoting, for simplicity, $f_i := f_{m(B^i)}$ and F_i the corresponding distribution function,

$$\begin{aligned} P\{\theta(X) \in [t_i, t_{i+1}]\} &= \int_{\prod_{\substack{j=0 \\ j \neq i}}^n \{y_i < y_j\}} f_i(y_i) \times \prod_{\substack{j=0 \\ j \neq i}}^n f_j(y_j) dy_0 \cdots dy_n \\ &= \int_{-\infty}^{\infty} f_i(y_i) \times \left(\prod_{\substack{j=0 \\ j \neq i}}^n \int_{y_i}^{\infty} f_j(y_j) dy_j \right) dy_i = \int_{-\infty}^{\infty} f_i(y) \times \prod_{\substack{j=0 \\ j \neq i}}^n (1 - F_j(y)) dy. \end{aligned}$$

Now, the result is obtained using the densities (2) and their distribution functions. □

The integral in (4) can be obtained analytically using a computer algebra system. It is a long expression that we will not copy here. Let us compare, instead, the minimum on two different intervals:

Let $t_1 < t_2 \leq t_3 < t_4$ and consider the Brownian bridge B_1 from (t_1, x_1) to (t_2, x_2) and the Brownian bridge B_2 from (t_3, x_3) to (t_4, x_4) . Denote $\ell_1 := t_2 - t_1$, $d_1 := |x_2 - x_1|$, $\ell_2 := t_4 - t_3$, $d_2 := |x_4 - x_3|$, and $\xi := x_3 \wedge x_4 - x_1 \wedge x_2$. See Figure 1.

We ask ourselves which of the two intervals $[t_1, t_2]$ and $[t_3, t_4]$ is more likely to contain the lowest value. We have

$$\begin{aligned} P\{m(B_1) < m(B_2)\} &= \int_{\{y < \bar{y}\}} f_{m(B_1)}(y) f_{m(B_2)}(\bar{y}) dy d\bar{y} \\ &= \int_{-\infty}^{\infty} f_{m(B_1)}(y) \left(\int_y^{\infty} f_{m(B_2)}(\bar{y}) d\bar{y} \right) dy. \end{aligned}$$

Taking as new variables $y - x_0 \wedge x_1$ instead of y , and $\bar{y} - x_2 \wedge x_3$ instead of \bar{y} , we get

$$\int_{-\infty}^{\xi \wedge 0} \frac{2}{\ell_1} (d_1 - 2y) \exp\left\{\frac{2y(d_1 - y)}{\ell_1}\right\} \left(\int_{y-\xi}^0 \frac{2}{\ell_2} (d_2 - 2\bar{y}) \exp\left\{\frac{2\bar{y}(d_2 - \bar{y})}{\ell_2}\right\} d\bar{y} \right) dy,$$

which can be written

$$\int_{-\infty}^{\xi \wedge 0} \frac{2}{\ell_1} (d_1 - 2y) \exp\left\{\frac{2y(d_1 - y)}{\ell_1}\right\} \left(1 - \exp\left\{\frac{2(y - \xi)(d_2 - (y - \xi))}{\ell_2}\right\} \right) dy. \quad (5)$$

This integral is also computable analytically. Its value depends on five parameters $(\ell_1, d_1, \ell_2, d_2, \xi)$, which are independent from each other in a general setting. Therefore, there is no easy way to tell if it is more likely to find the minimum in one interval or the other. One observes, as the intuition suggests, that the above probability is an increasing function of ℓ_1 , d_2 and ξ , and that is decreasing in ℓ_2 and d_1 , when all the other parameters are fixed.

In the case when the intervals are $[0, t_1]$ and $[t_1, 1]$, then $\ell_2 = 1 - \ell_1$, and ξ can be expressed in terms of d_1 and d_2 , in different ways according to the relative positions $x_0 < x_1 < x_2$, $x_0 < x_2 < x_1$, or $x_1 < x_0 \wedge x_2$, so that the number of free parameters reduces to three.

Example 3.1. Let B_1 be the bridge from $(0, 0)$ to $(0.5, 0)$, and B_2 the bridge from $(0.5, 0)$ to $(1, d_2)$, for $d_2 \geq 0$, and set $p := P\{m(B_1) < m(B_2)\}$. The following table illustrates how p and d_2 are related.

p	d_2
0.5	0.0000
0.6	0.1837
0.7	0.4386
0.8	0.8384
0.9	1.6620
0.95	2.7302
0.99	6.8638

In fact, the explicit functional relationship is given by $p = \frac{1}{2} + \sqrt{\pi/8} d_2 \exp\{d_2^2/2\} (1 - \operatorname{erf}\{d_2/\sqrt{2}\})$, where $\operatorname{erf}()$ is the standard error function. If we keep the same first bridge, and make the second shorter and ending at zero, say from $(1 - \ell_2, 0)$ to $(1, 0)$, the dependence between p and the length ℓ_2 is even easier: $p = 1/(2\ell_2 + 1)$. Both are straightforward computations from expression (5).

By equating both expressions one obtains the variations in d_2 and ℓ_2 that give an equivalent raise of the probability that the first interval contain the lowest value.

3.2 Approximate computation

Despite the fact that the integrals (4) can be computed analytically, the time needed to solve them grows exponentially in the number n of intervals. Indeed, the exact computation involves decomposing the integrand in the sum of $O(2^n)$ terms. Each term has an elementary primitive, but in an optimisation

procedure in which more and more points are sampled, and consequently the Brownian bridge is conditioned to one more point each time, the computation becomes cumbersome very quickly. For example, with just 8 intervals, the computer algebra system *maxima* takes more than three hours to obtain the result, in an Intel i7 CPU with plenty of memory at its disposal (although *maxima* only uses one of its cores). It is therefore justified to resort to an approximate method.

We remark that adding one more point to the set of conditioning points (that means, splitting one of the intervals in two), forces to recompute from scratch the probabilities of all intervals. There seems to be no way to reuse previous computations.

As we have seen, the probabilities $P\{m(B^i) < m(B^j)\}$, for each pair of indices i, j , can be computed exactly and more easily than (4); nevertheless, they are not useful even to find the interval with the maximal probability. An interval $[t_i, t_{i+1}]$ may satisfy $P\{m(B^i) < m(B^j)\} > 1/2, \forall j \neq i$, and still not be the interval with the largest probability of containing $m(X)$. For instance, if we condition the Brownian motion to pass through the points

$$(0, 0), (0.144, 0.225), (0.610, 0.344), (1, 0.145),$$

we find that $P\{m(B^1) > m(B^2)\} = 0.5436$ and $P\{m(B^1) > m(B^3)\} = 0.5198$. However, the first interval is the least probable one to contain the minimum:

$$P_{\theta(X)}([t_0, t_1]) = 0.3124, \quad P_{\theta(X)}([t_1, t_2]) = 0.3374, \quad P_{\theta(X)}([t_2, t_3]) = 0.3502.$$

Even more, such a “winning” interval may not exist. For instance, conditioning to

$$(0, 0), (0.392, 0.031), (0.594, -0.157), (1, 0.435),$$

one gets the circular relation $P\{m(B^1) < m(B^2)\} = .5018$, $P\{m(B^2) < m(B^3)\} = .5032$, $P\{m(B^3) < m(B^1)\} = .5013$.

All these arguments support the need to compute (4) numerically. It is easy to do it with a rigorous error bound: For some $\hat{x} < \min(x_0, \dots, x_{n+1})$, split the integral into the two intervals $(-\infty, \hat{x}]$ and $[\hat{x}, \min\{x_0, \dots, x_{n+1}\}]$. On the first one, the integral is bounded by

$$\begin{aligned} & \int_{-\infty}^{\hat{x}} \frac{2}{t_{i+1} - t_i} (x_i + x_{i+1} - 2y) \exp\left\{\frac{-2(x_i - y)(x_{i+1} - y)}{t_{i+1} - t_i}\right\} dy = \\ & \exp\left\{\frac{-2(x_i - \hat{x})(x_{i+1} - \hat{x})}{t_{i+1} - t_i}\right\} \leq \exp\left\{\frac{-2(x_i \wedge x_{i+1} - \hat{x})^2}{t_{i+1} - t_i}\right\}. \end{aligned}$$

To make this quantity less than a fixed small ε , we can take $\hat{x} < x_i \wedge x_{i+1} - \left(\frac{t_{i+1} - t_i}{2} \log \frac{1}{\varepsilon}\right)^{1/2}$.

For the second interval, denoting the integrand by f and using for instance the standard rectangle rule with step size h , the error is bounded by $\frac{1}{2} \|f'\|_{\infty} \cdot h \cdot L$, where $L := \min(x_0, \dots, x_{n+1}) - \hat{x}$.

Differentiating f and taking into account that all the exponentials take values less than 1, one obtains $\|f'\|_{\infty} \leq C$ with

$$C := \frac{4}{t_{i+1} - t_i} \left[1 + (x_i + x_{i+1} - 2\hat{x}) \sum_{j=0}^n \frac{1}{t_{j+1} - t_j} (x_j + x_{j+1} - 2\hat{x}) \right],$$

and the integration step size to ensure an error less than ε must be

$$h \leq \frac{2\varepsilon}{C \cdot L}.$$

A much more efficient method but with a not completely rigorous error bound is given by the quadpack functions present in the C Gnu Scientific Library and the Fortran SLATEC Library, which apply a Gauss-Kronrod rule [8]. With $n = 50$, the computation is completed in less than one-tenth of second, in an Intel

	Set 1. Result: 0.05722062072176488			Set 2. Result: 0.3539550244743264		
	error	time	memory	error	time	memory
1) exact		1	1.26		57.2	199
2) quadpack	$< 10^{-16}$	1.07	1	$< 10^{-13}$	1	1
3) Romberg	$< 10^{-11}$	1.07	1.54	$< 10^{-11}$	1.07	1.63
4) Riemann left	$< 10^{-16}$	122	141	4.146×10^{-6}	121	137
5) Riemann random	1.008×10^{-6}	103	81.8	4.283×10^{-6}	100	79.0
6) simulation	4.179×10^{-3}	249	191	6.045×10^{-3}	252	188
	Set 3. Result: 0.003053658531871728			Set 4. Result: 0.3498434691309963		
	error	time	memory	error	time	memory
2) quadpack		1	1		1	1
3) Romberg	$< 10^{-12}$	3.41	2.89	$< 10^{-11}$	3.52	2.91
4) Riemann left	$< 10^{-18}$	197	294	$< 10^{-7}$	199	158
5) Riemann random	1.366×10^{-7}	143	157	8.140×10^{-6}	144	84.3
6) simulation	1.446×10^{-3}	459	462	1.06×10^{-2}	456	254

Table 1: See Example 3.2

i7 CPU at 2.40GHz with 20GB RAM, using the quadpack routines implemented in the computer algebra system *maxima*, with an estimated absolute error rarely bigger than 10^{-9} .

The integral of (4) can also be transformed into an integral on $[0, 1]$ setting $y = \min_i x_i - (1 - x)/x$ (this is in fact what quadpack does), and the new integrand does not present any singularity.

Example 3.2. In Table 1, we show the effective computation of the probability that the minimum fall in the first interval, in several situations and with different methods. Sets 1 and 2 comprise four intervals, with end-points at $t = (0, .1, .2, .5, 1)$, and values $x = (0, 0, 0, 0, 0)$ and $x = (0, .1, .2, .3, .4)$ respectively. Sets 3 and 4 comprise sixteen intervals, with end-points

$$t = (0, .025, .050, .075, .100, .125, .150, .175, .200, .275, .350, .425, .500, .625, .750, .875, 1) ,$$

and all images set to zero in set 3 and to $x = i/40$, $i = 0, \dots, 16$, in set 4.

The methods are: 1) the analytical computation of the integral (4), only in the case of fewest intervals (“exact”); 2) the quadpack functions through *maxima*; 3) the romberg routine built-in in *maxima*; 4) the Riemann approximations with 10 000 subintervals, taking always their left points; 5) the Riemann approximations with the same number of subintervals, taking a random point in each one; and 6) the simulation method explained in the next section. In 3), 4), 5), the computations are also made after the mentioned explicit transformation to the interval $[0, 1]$. In 6) a sample of size 10 000 is taken. For the methods including randomness, 5) and 6), we show the highest error observed after 20 realizations.

All computations were programmed in *maxima*. Time and memory are relative to the fastest and the more economic method in each case; we used the figures reported by *maxima* itself in a single run. They give therefore just a rough idea of the computational cost. In the case of 16 intervals, the “exact” computation is infeasible and we have taken the result of quadpack as the base for the figures of the other methods.

4 Simulating the law of the minimum

We are interested in approximating in an effective way the law of the minimum of the Brownian motion conditioned to the points $(t_0, x_0), \dots, (t_{n+1}, x_{n+1})$, with $t_0 = 0$ and $t_{n+1} = 1$, so that particular parameters such as its moments can also be easily estimated. To this end, taking into account the difficulty and length of the analytical computations implied by (3), we resort to simulation.

A minimum value for each bridge from (t_i, x_i) to (t_{i+1}, x_{i+1}) can be easily simulated from its distribution function $F_{m(B_i)}$, which is explicitly invertible:

$$F_{m(B_i)}^{-1}(z) = \frac{1}{2}(x_i + x_{i+1} - ((x_{i+1} - x_i)^2 - 2(t_{i+1} - t_i) \log z)^{1/2}), \quad z \in (0, 1).$$

Since $m(X) = \min\{m(B^0), \dots, m(B^n)\}$, we can simulate a minimum value of X as the minimum of the simulated minima of each bridge. The computational cost is linear in n . At the same time, the relative frequency with which each interval contributes to the global minimum constitutes another way to approximate the probabilities $P\{\theta(X) \in [t_i, t_{i+1}]\}$ of Section 3.2. This is what is done in row 6 of Table 1 for the interval $[0, t_1]$.

For example, with set 1 of Example 3.2, and a sample of size 10 000, we have obtained the following confidence intervals for the probabilities of each interval to host the minimum:

interval	95% C.I.
[0, 0.1]	[0.3501, 0.3715]
[0.1, 0.2]	[0.0955, 0.1169]
[0.2, 0.5]	[0.2362, 0.2576]
[0.5, 1]	[0.2758, 0.2972]

The computations have been done in R with the MultinomialCI package, based on the algorithm of Sison and Glaz [10].

Figure 2 shows the result of simulating the minimum of the process in the way described above, conditioned to equispaced points with images equal to zero. The figure includes an histogram and an estimation of the density using the polynomial splines algorithm described in [11], as implemented in the logspline package in R.

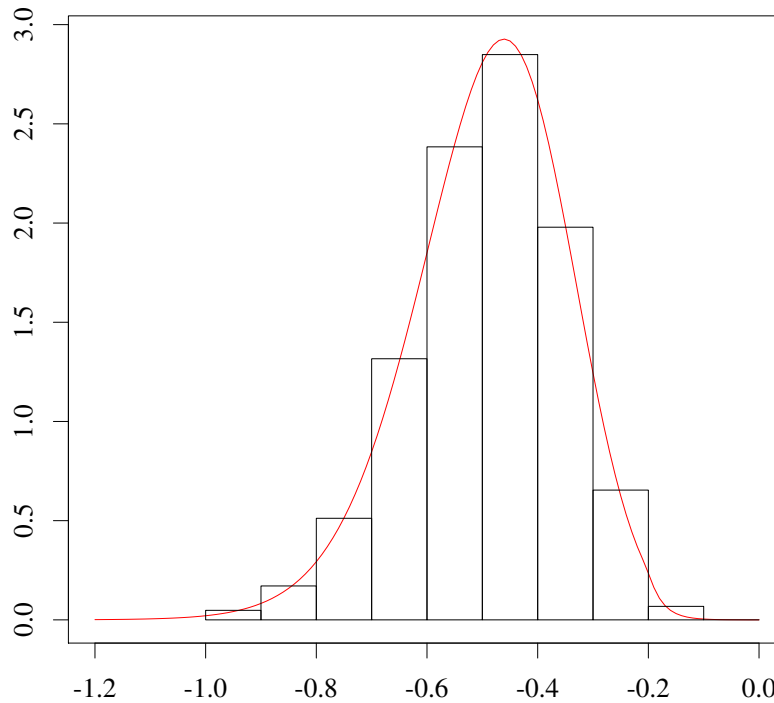


Figure 2: Histogram and density estimation of the minimum of a Brownian motion conditioned to pass through the points $\{(k/4, 0), k = 0, \dots, 4\}$, with a sample size of 10000 observations. An asymptotic 95% symmetric confidence interval for the mean yielded $[-0.4939, -0.4884]$. The sample median was -0.4814.

5 Non-adaptive optimisation

Suppose now that we have a black-box optimisation problem in which we can assume that the Brownian bridge is a good probabilistic model for the function at hand. That means, suppose that we are trying to find a point in the interval $[0, 1]$ with an image as close as possible to the true minimum of a given but unknown path, drawn at random from the law of a Brownian bridge. The value of the bridge at the end-points is assumed to be given, or that they have already been sampled. It is also assumed that we are allowed to sample the path at a fixed small quantity n of points in $[0, 1]$.

A non-adaptive algorithm for this optimisation problem consists in deciding beforehand the n points where we are going to sample the path. They have the same convergence order as the best adaptive algorithm as $n \rightarrow \infty$, namely $O(n^{-1/2})$, are much simpler to implement, and offer parallelisation opportunities. Therefore it is worth comparing non-adaptive algorithms in terms of the size of the error incurred for small n . In a forthcoming paper we will discuss and compare some adaptive heuristics.

We will first consider and compare two strategies: Sampling at equidistant points $\frac{k}{n+1}$, $k = 1, \dots, n$, and sampling at random uniformly distributed points. We apply both to a bridge with values 0 and 1 at the end-points and to a symmetric bridge (same value at the end-points). We obtain approximate 95% confidence intervals for the difference between the minimal sampled value and the true minimum of the path. The results are summarised in Table 2. Formally, the confidence intervals estimate the expectation

$$\mathbb{E} \left[\min_{0 \leq i \leq n+1} B_{t_i} - \min_{t \in [0,1]} B_t \right],$$

where B is the initial bridge joining $(0, x_0)$ and $(1, x_{n+1})$. In one case the t_i are fixed; in the other, they are themselves random.

The procedure for the computations is as follows:

1. Fix the number of points to sample. We have used $n = 2, 4, 8, 16, 32, 64$ to see the evolution of the intervals when the number of points increases.
2. Sample a path of the Brownian bridge at point $t_1 = 1/(n+1)$; this is done by simulating a value of B_{t_1} , which is easy because its law is Gaussian. Sample at point $t_2 = 2/(n+1)$ the bridge from (t_1, x_1) to $(1, x_{n+1})$. Proceed similarly to get the values of the path at all equidistant points.
3. For the simulation at the n random points in $[0, 1]$, determine first at which subinterval of all previously sampled points the new one belongs to, and sample from the corresponding bridge. The equidistant points of step 2 and their evaluations are included here so that both methods are in fact applied to the same path.
4. From all the $2n$ sampled points of steps 2 and 3, estimate the expectation of the minimum of the path to which they belong, with the method described in Section 4. We have used a simulation of size 1000 in this case, taking the mean of the values obtained.
5. For each sampling strategy, compute the difference between the best sampled point and the estimated minimum of the path.
6. Repeat steps 2–5 a number of times (we used 1000), and construct the asymptotic confidence intervals from the sets of differences obtained, for both strategies.

The results are summarised in Table 2 for two different initial bridges. We observe that the equidistant sampling performs better for both. One also observes that the errors are smaller for the non-symmetric bridge; this can be explained by a smaller variance of its minimum value (these variances can be computed analytically from the density (2)), despite the fact that many evaluations are possibly wasted in a non-promising region.

no. of points	Bridge from (0,0) to (1,1)		Bridge from (0,0) to (1,0)	
	95% C.I. eqd	95% C.I. rnd	95% C.I. eqd	95% C.I. rnd
2	[0.2390, 0.2517]	[0.2547, 0.2729]	[0.3417, 0.3549]	[0.3791, 0.4012]
4	[0.2025, 0.2132]	[0.2163, 0.2320]	[0.2552, 0.2649]	[0.3002, 0.3194]
8	[0.1659, 0.1745]	[0.1759, 0.1891]	[0.1944, 0.2023]	[0.2183, 0.2317]
16	[0.1280, 0.1341]	[0.1376, 0.1475]	[0.1390, 0.1447]	[0.1651, 0.1760]
32	[0.0920, 0.0963]	[0.1040, 0.1111]	[0.0987, 0.1028]	[0.1198, 0.1283]
64	[0.0663, 0.0694]	[0.0778, 0.0838]	[0.0712, 0.0741]	[0.0851, 0.0910]

Table 2: Approximate confidence intervals for the expectation of the error when estimating the minimum of a Brownian bridge by the minimum of the sampled values, for equidistant ('eqd') and random ('rnd') sampling.

no. of points	95% C.I. eqp
2	[0.1975, 0.2144]
4	[0.1637, 0.1780]
8	[0.1275, 0.1387]
16	[0.0909, 0.0991]
32	[0.0677, 0.0741]
64	[0.0491, 0.0538]

Table 3: The third method, dividing $[0, 1]$ in intervals of equal probability to host the minimum, for a bridge from (0,0) to (1,1).

The quotient between the estimated errors with equidistant and with random sampling decreases when the number of points increases. Calvin [2] showed that when $n \rightarrow \infty$, this quotient approaches ≈ 0.8239 . Equidistant sampling is therefore a better choice, if n is really fixed in advance.

A third logical non-adaptive strategy for the non-symmetric bridge is to sample at points that divide $[0, 1]$ in intervals that have the same probability to contain the minimum. This is what happens with the equidistant points in the case of a symmetric bridge. We need first to find the points $0 = t_0 < t_1 < \dots < t_n < t_{n+1} = 1$ such that

$$P\{\theta(B) \in [t_i, t_{i+1}]\} = 1/(n+1), \quad \text{for all } i = 0, \dots, n,$$

where B is the bridge from (0,0) to (1,1). This can be done numerically without difficulty with the general formula (7) for the density of $\theta(B)$ that we are going to prove in the next section, and that in this case reduces to

$$f_{\theta(B)}(s) = \sqrt{\frac{2(1-s)}{\pi s}} \exp\left\{\frac{-s}{2(1-s)}\right\} \mathbf{1}_{\{0 \leq s \leq 1\}}.$$

With the same number of simulations used in Table 2, we get the results of Table 3.

Comparing both tables, we see that the expectation of the error for the equiprobable ('eqp') strategy is the lowest of the three; however, the confidence intervals for the error are longer than with 'eqd', because the variance turns out to be larger. The variances for 'rnd' are the largest.

6 Simulating the law of the location of the minimum

The location of the minimum of a continuous function is an ill-posed problem: small changes in the function may result in big changes in the location of the minimum. Therefore, the information about the location of the minimum given by the sampled values is limited, and possibly of less practical importance

than the information about the minimum value. Anyway, we can try to visualise this information through the law of $\theta(X) := \arg \min_{[0,1]} X_t$.

This law can be simulated with the auxiliary use of the minima of all bridges B^0, \dots, B^n , which in turn can be easily simulated as we have seen in Section 4. We prove first that conditioned to all these minima, the variables $\theta(X)$ and $\theta(B^j)$, where j is the index of the interval where the global minimum is attained, have the same law. This is the contents of the next proposition:

Proposition 6.1. *Denote $x^* := \min_{0 \leq i \leq n+1} x_i$ and $\Psi_j := \{y = (y_0, \dots, y_n) \in [-\infty, x^*]^{n+1} : y_j = \min_i y_i\}$. For any Borel set $A \subset [0, 1]$,*

$$P\{\theta(X) \in A \mid m(B^0) = y_0, \dots, m(B^n) = y_n\} = P\{\theta(B^j) \in A \mid m(B^j) = y_j\}$$

on Ψ_j almost everywhere with respect to Lebesgue measure.

Proof. First, we prove the equality

$$P\{\theta(X) \in A \mid m(B^0) = y_0, \dots, m(B^n) = y_n\} = P\{\theta(B^j) \in A \mid m(B^0) = y_0, \dots, m(B^n) = y_n\}.$$

on Ψ_j almost everywhere with respect to Lebesgue measure (y-a.e. on Ψ_j for short).

It is clear that the support of $\theta(X)$ and of $\theta(B^j)$ with respect to the conditional law is the interval $I_j := [t_j, t_{j+1}]$, y-a.e. on Ψ_j ; therefore, we can assume $A \subset I_j$.

Denoting $s^* := \theta(X)$, since $X_s = B_s^j$, $\forall s \in I_j$, and $s^* \in I_j$ almost surely with respect to the conditional law, y-a.e. on Ψ_j , we have $m(X) = X_{s^*} = B_{s^*}^j = m(B^j)$. This implies $\theta(B^j) = s^*$ almost surely, y-a.e. on Ψ_j , due to the almost sure uniqueness of the location of the minimum of a Brownian bridge.

Finally, the equality

$$P\{\theta(B^j) \in A \mid m(B^1) = y_1, \dots, m(B^n) = y_n\} = P\{\theta(B^j) \in A \mid m(B^j) = y_j\}.$$

comes from the independence of B^j from all the other bridges. \square

From Proposition 6.1, we see that to simulate the location of the minimum of X it is enough to simulate the minima of all bridges, select the lowest of them y_j , and then simulate the location $\theta(B^j)$ of the minimum of the bridge B^j conditioned only to $m(B^j) = y_j$. We need first the law of the vector $(m(B^j), \theta(B^j))$. This is stated in the next proposition. We also give the marginal law of $\theta(B^j)$, since we have not been able to find it in the literature in this generality, even though it will not be used directly in the simulation.

Proposition 6.2. *Writing $\ell := t_{j+1} - t_j$ and $d := |x_{j+1} - x_j|$, the minimum of the bridge B^j from (t_j, x_j) to (t_{j+1}, x_{j+1}) and its location have the joint density*

$$f_{(m(B^j), \theta(B^j))}(y, s) = \frac{(x_j - y)(x_{j+1} - y)\sqrt{2\ell}}{\sqrt{\pi(s - t_j)^3(t_{j+1} - s)^3}} \exp\left\{\frac{d^2}{2\ell} - \frac{(x_j - y)^2}{2(s - t_j)} - \frac{(x_{j+1} - y)^2}{2(t_{j+1} - s)}\right\} \mathbf{1}_{\{y < x_j, y < x_{j+1}, t_j \leq s \leq t_{j+1}\}}, \quad (6)$$

and the density of the location is

$$f_{\theta(B^j)}(s) = \left[\frac{d}{\ell^{3/2}} \sqrt{\frac{2}{\pi h(s)}} \exp\left\{\frac{-d^2}{2\ell} h(s)\right\} + \frac{\ell - d^2}{\ell^2} \operatorname{erfc}\left\{\left(\frac{d^2}{2\ell} h(s)\right)^{1/2}\right\} \right] \mathbf{1}_{\{t_j \leq s \leq t_{j+1}\}}, \quad (7)$$

where

$$h(s) = \begin{cases} (t_{j+1} - s)/(s - t_j), & \text{if } x_{j+1} \leq x_j \\ (s - t_j)/(t_{j+1} - s), & \text{if } x_j \leq x_{j+1} \end{cases}$$

and $\operatorname{erfc}()$ is the complementary error function $1 - \operatorname{erf}()$.

Proof. The joint law of W_t , with W a standard Brownian motion $W = \{W_s, s \geq 0\}$ starting at $W_0 = a$, its minimum m_t up to time t and the location θ_t of this minimum, is known to have the density

$$P_a\{W_t \in db, m_t \in dy, \theta_t \in ds\} = \frac{(a-y)(b-y)}{\pi\sqrt{s^3(t-s)^3}} \exp\left\{-\frac{(a-y)^2}{2s} - \frac{(b-y)^2}{2(t-s)}\right\} \mathbf{1}_{\{y < a, y < b, 0 \leq s \leq t\}} \quad (8)$$

(see Karatzas-Shreve [6, Prop. 2.8.15], or Csáki et al. [5], where it is extended to general diffusions); the formula is usually stated for the maximum, but (8) is easily deduced by symmetry, taking into account that $-W$ is a Brownian motion with starting value $-a$ at time 0.

Consequently, if W starts instead at time $u < t$, we have

$$P_{(u,a)}\{W_t \in db, m_t \in dy, \theta_t \in ds\} = \frac{(a-y)(b-y)}{\pi\sqrt{(s-u)^3(t-s)^3}} \exp\left\{-\frac{(a-y)^2}{2(s-u)} - \frac{(b-y)^2}{2(t-s)}\right\} \mathbf{1}_{\{y < a, y < b, u \leq s \leq t\}}.$$

Conditioning to $\{W_t = b\}$, one finds the joint density of the minimum m and its location θ for a Brownian bridge B joining the points (u, a) and (t, b) :

$$P_{(u,a),(t,b)}\{m \in dy, \theta \in ds\} = \frac{(a-y)(b-y)\sqrt{2(t-u)}}{\sqrt{\pi}(s-u)^3(t-s)^3} \exp\left\{\frac{(b-a)^2}{2(t-u)} - \frac{(a-y)^2}{2(s-u)} - \frac{(b-y)^2}{2(t-s)}\right\} \mathbf{1}_{\{y < a, y < b, u \leq s \leq t\}},$$

which is equivalent to (6).

Integrating out y , we get, if $a < b$,

$$P_{(u,a),(t,b)}\{\theta \in ds\} = \left[\frac{b-a}{(t-u)^2} \sqrt{\frac{2(t-u)(t-s)}{\pi(s-u)}} \exp\left\{\frac{-(b-a)^2(t-s)}{2(t-u)(s-u)}\right\} + \frac{(t-u)-(b-a)^2}{(t-u)^2} \operatorname{erfc}\left\{(b-a)\sqrt{\frac{t-s}{2(t-u)(s-u)}}\right\} \right] \mathbf{1}_{\{u \leq s \leq t\}},$$

and, in case $a > b$,

$$P_{(u,a),(t,b)}\{\theta \in ds\} = \left[\frac{a-b}{(t-u)^2} \sqrt{\frac{2(t-u)(s-u)}{\pi(t-s)}} \exp\left\{\frac{-(b-a)^2(s-u)}{2(t-u)(t-s)}\right\} + \frac{(t-u)-(b-a)^2}{(t-u)^2} \operatorname{erfc}\left\{(a-b)\sqrt{\frac{s-u}{2(t-u)(t-s)}}\right\} \right] \mathbf{1}_{\{u \leq s \leq t\}},$$

from where we get (7). □

Proposition 6.3. *The location of the minimum $\theta(B^j)$ conditioned to $m(B^j)$ has a density of the form*

$$f_{\theta(B^j)|_{m(B^j)=y}}(s) = C(y)(s-t_j)^{-3/2}(t_{j+1}-s)^{-3/2} \exp\left\{-\frac{A(y)}{2(s-t_j)} - \frac{B(y)}{2(t_{j+1}-s)}\right\} \cdot \mathbf{1}_{\{t_j \leq s \leq t_{j+1}\}} \quad (9)$$

for $y < x_j, y < x_{j+1}$, where A, B and C are positive constants depending only on y .

Proof. This is an immediate computation from the joint density (6) and the marginal (2), yielding (9) with

$$A(y) = (x_j - y)^2, \quad B(y) = (x_{j+1} - y)^2, \\ C(y) = \frac{(t_{j+1} - t_j)^{3/2}(x_j - y)(x_{j+1} - y)}{\sqrt{2\pi}(x_j + x_{j+1} - 2y)} \exp\left\{\frac{(x_{j+1} + x_j - 2y)^2}{2(t_{j+1} - t_j)}\right\}.$$

□

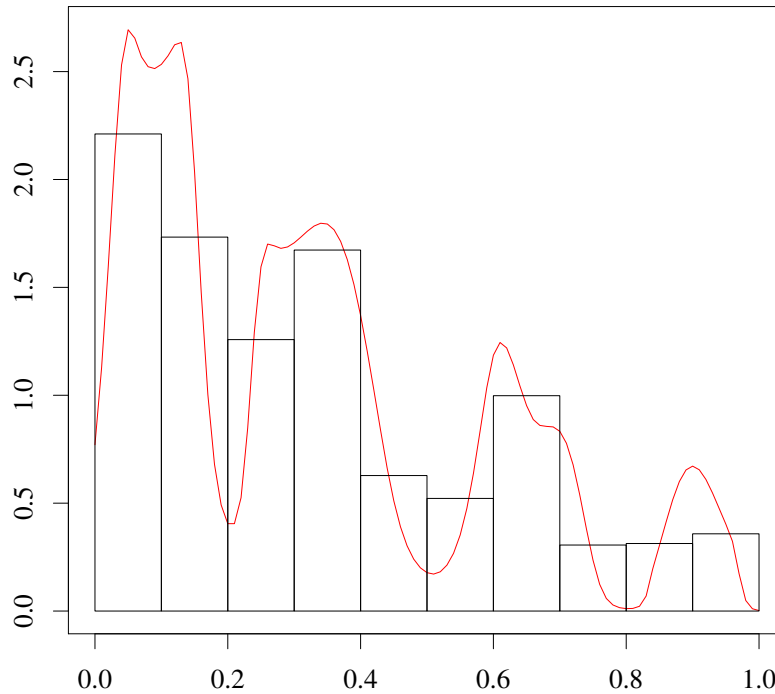


Figure 3: Density estimation for the location of the minimum of the Brownian bridge conditioned to the points $(0,0), (0.2,0.06), (0.5,0.16), (0.8,0.26), (1,0.20)$.

Notice that the density (7) is not bounded as $h(s)$ goes to zero at one of the end-points of the interval $[t_j, t_{j+1}]$. Therefore, it is not easy to sample exactly from it. However, the conditional density (9) given a value $y < \min\{x_j, x_{j+1}\}$ it is bounded, which makes it more amenable to the acceptance/rejection method (see e.g. Asmussen and Glynn [1]). The result is a sample from the joint density of the minimum and its location, from where one obtains a sample of the marginal law of the location. This trick, together with Proposition 6.1, will allow us to simulate the location of the minimum of the whole process X , conditioned to pass through the given set of points.

To apply acceptance/rejection by comparison with a uniform distribution, the global maximum of the function (9) should be easily calculated or approximated from above. This is indeed the case: There are two obvious minima at the end-points t_j and t_{j+1} ; the remaining extremal points are the roots of the 3-degree polynomial

$$3C(y)[(s-t_j)^2(t_{j+1}-s) - (s-t_j)(t_{j+1}-s)^2] + A(y)(t_{j+1}-s)^2 - B(y)(s-t_j)^2,$$

as can be seen by differentiating in s and multiplying by $2(s-t_j)^{7/2}(t_{j+1}-s)^{7/2}$. This polynomial may have one or three real roots, corresponding to a unique maximum, or to two maxima, with a minimum in between. In any case, the global maximum can be computed exactly and the acceptance/rejection method can be implemented for this density.

In Figure 3, we see the result of a simulation of size 10 000, with a density estimation using the `logsplines` method in R. As it was remarked in Section 2, once we are considering the minimum of the whole process, the different bridges are no longer independent; in particular, the shape of the density in each subinterval is not the one to be expected from formula (7), and in fact it is quite difficult to predict from the conditioning values. Hence the interest to have an exact simulation method.

As a more clear example of the last remark, consider the concatenation of two symmetric bridges, from $(0,0)$ to $(0.5,0)$, and from $(0.5,0)$ to $(1,0)$. Separately, the location of their minima follows a uniform distribution; however, the location of the global minimum follows the density simulated in Figure 4. The fact that the minimum of the two minima tends to take a lower value than a single minimum drifts away its location from the end-points of the subintervals.

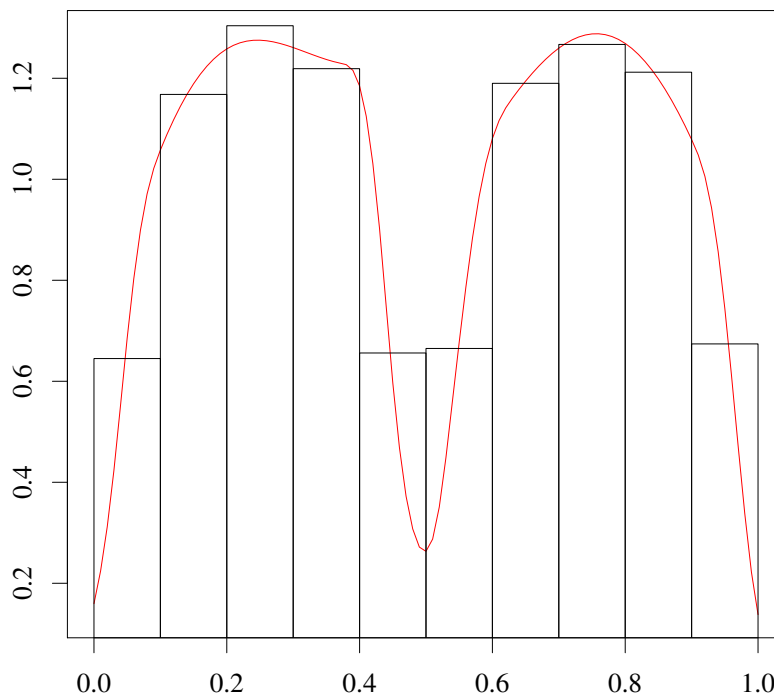


Figure 4: Density estimation for the location of the global minimum of two concatenated symmetric identical Brownian bridges.

7 Acknowledgements

This work has been supported by grants numbers MTM2011-29064-C03-01 from the Ministry of Economy and Competitiveness of Spain; and UNAB10-4E-378, co-funded by the European Regional Development Fund (ERDF).

References

- [1] Søren Asmussen and Peter W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.
- [2] James M. Calvin. Average performance of passive algorithms for global optimization. *J. Math. Anal. Appl.*, 191(3):608–617, 1995.
- [3] James M. Calvin. Average performance of a class of adaptive algorithms for global optimization. *Ann. Appl. Probab.*, 7(3):711–730, 1997.
- [4] James M. Calvin. Lower bound on complexity of optimization of continuous functions. *J. Complexity*, 20(5):773–795, 2004.
- [5] Endre Csáki, Antónia Földes, and Paavo Salminen. On the joint distribution of the maximum and its location for a linear diffusion. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(2):179–194, 1987.
- [6] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [7] M. Locatelli. Bayesian algorithms for one-dimensional global optimization. *J. Global Optim.*, 10(1):57–76, 1997.

- [8] Robert Piessens, Elise de Doncker-Kapenga, Christoph W. Überhuber, and David K. Kahaner. *QUADPACK*, volume 1 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1983. A subroutine package for automatic integration.
- [9] Klaus Ritter. Approximation and optimization on the Wiener space. *J. Complexity*, 6(4):337–364, 1990.
- [10] Cristina P. Sison and Joseph Glaz. Simultaneous confidence intervals and sample size determination for multinomial proportions. *J. Amer. Statist. Assoc.*, 90(429):366–369, 1995.
- [11] Charles J. Stone, Mark H. Hansen, Charles Kooperberg, and Young K. Truong. Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.*, 25(4):1371–1470, 1997. With discussion and a rejoinder by the authors and Jianhua Z. Huang.
- [12] Anatoly Zhigljavsky and Antanas Žilinskas. *Stochastic global optimization*, volume 9 of *Springer Optimization and Its Applications*. Springer, New York, 2008.