INFORMATICA, 2016, Vol. 27, No. 2, 323–334

DOI: http://dx.doi.org/10.15388/Informatica.2016.87

Probability Models in Global Optimization

James M. CALVIN
Department of Computer Science, New Jersey Institute of Technology
Newark, NJ 07102-1982, USA
e-mail: calvin@njit.edu

Received: January 2016; accepted: April 2016

Abstract. This paper reviews the interplay between global optimization and probability models, concentrating on a class of deterministic optimization algorithms that are motivated by probability models for the objective function. Some complexity results are described for the univariate and multivariate cases.

Key words: optimization, statistical models, convergence.

1. Introduction

This paper is a survey of a class of algorithms and complexity results for global optimization based on probability models. Probability models have been used both to inspire the construction of new optimization algorithms, and also to analyse the average performance of algorithms.

The global optimization problem we consider starts with a class F of objective functions; a typical example is $C^r([0,1])$, the r-times continuously differentiable functions on the unit interval. To avoid the better-understood local optimization problem, we avoid classes that have properties such as convexity. For $f \in F$, we want to approximate the global minimum f^* using function values at sequentially chosen points. If the approximation is $A(f)$, then we want to make $A(f) - f^*$ small in some sense.

The global optimization problem is intractable in the worst case setting for the types of function classes that we have in mind. For example, if we take $F = C^r([0,1])$, then to ensure finite error we need a further assumption such as uniform bounds on the derivatives. For example, we can restrict to functions with $|f^{(r)}|_\infty \leqslant 1$. Then $\Theta\left(\epsilon^{-1/r}\right)$ function values are required in order to ensure that the approximation error is at most ϵ. If we take the domain to be the d-dimensional hypercube, then $\Theta\left(\epsilon^{-d/r}\right)$ function values are required, and so the problem is intractable in the multivariate case; see Wasilkowski (1992).

There is another reason to adopt an average-case analysis instead of a worst-case analysis. Adaptive algorithms choose each evaluation point as a function of previous function evaluations, while nonadaptive methods evaluate at the same points for each function. If the class F is convex and symmetric, that is if $f, g \in F$ and $\lambda \in [0,1]$ implies that $-f$ and $\lambda f + (1-\lambda)g \in F$, then in the worst case adaptive methods are essentially no more efficient than nonadaptive methods; see Ritter (2000). In practice adaptive methods are favoured over nonadaptive methods.

The pessimistic worst-case complexity of global optimization problems is a reflection of the fact that a worst case error bound is too much to expect with the weak assumptions that the optimizer starts with.

One alternative to a worst-case analysis is to consider randomized algorithms. Many randomized global optimization algorithms have been proposed; see, for example, Zhigljavsky and Žilinskas (2008) and Tikhomirov (2006). In this paper we will consider only deterministic algorithms.

Instead of seeking an error bound that holds for every $f \in F$, another approach is to construct algorithms that have small error on average. To carry out this approach we need a probability on F. Of course, the results we obtain may depend very much on the choice of probability. Gaussian probability measures are natural choices because they do not require much special structure of the class F and they are reasonably tractable. For these reasons Gaussian measure have been adopted for the average-case analysis of many numerical problems; see Ritter (2000).

Though we do not consider the problem in this paper, the random function model is well-suited to problems with noise-corrupted function evaluations. If independent Gaussian measures are adopted for the objective function and the noise in the observations, then the observed values have a Gaussian distribution. The case of Wiener measure for the objective function and independent Gaussian noise for the observations was treated in Calvin and Žilinskas (2005).

Several algorithms have been developed based on a probability model on a class F; for example, Kushner (1962), Mockus (1972). These are studied in a general setting in Žilinskas (1985). Given such an algorithm, the natural question is the average error of the algorithm under the assumed probability measure. This question will be addressed in Section 3. Another issue is the rate at which the error approaches zero asymptotically as the number of function evaluations grows for a fixed f in some class; this will be addressed in Section 4. General computational and algorithmic issues in global optimization are reviewed in Zhigljavsky and Žilinskas (2016).

2. Notation

Let f be a real-valued function defined on the unit interval $[0, 1]$, and denote the global minimum by $f^* = \min_{0 \leqslant s \leqslant 1} f(s)$. The optimizer adaptively chooses points $t_1, t_2, \ldots \in [0, 1]$,

$$t_1,\ t_2\big(t_1, f(t_1)\big),\ t_3\big(t_1, f(t_1), t_2, f(t_2)\big),\ \ldots$$

and, after n observations, forms an approximation to the global minimum M based on $\{t_i, f(t_i)\colon i = 1, 2, \ldots, n\}$. We allow each t_{n+1} to depend on the previous observations; i.e. the $(n+1)$st observation point is given as a function

$$t_{n+1} = h_{n+1}\big(t_1, f(t_1), \ldots, t_n, f(t_n)\big).$$

We are mainly interested in the error in approximating the function's global minimum, which we denote by $\Delta_n = M_n - f^*$, where $M_n = \min_{i \leqslant n} f(t_i)$.

Let $\{z_n\}$ be a sequence of positive numbers, where z_n may depend on the first $n-1$ observations $\{t_i, f(t_i);\ i < n\}$. We consider the optimization method based on the following procedure: given $t_1, t_2, \ldots, t_n$ and $f(t_1) = x_1, f(t_2) = x_2, \ldots, f(t_n) = x_n$, choose the next point $t_{n+1} \in [0, 1]$ to maximize

$$P\big(f(t_{n+1}) < M_n - z_n \,\big|\, f(t_1) = x_1, \ldots, f(t_n) = x_n\big). \tag{1}$$

3. Univariate Models

In addition to the sequences of observation sites $\{t_0, t_1, t_2, \ldots\}$, it will be necessary to refer to the ordered observations for each fixed n. Therefore, for $n \geqslant 2$ let

$$0 \equiv t_0^n < t_1^n < t_2^n < \cdots < t_{n-1}^n < t_n^n \leqslant 1$$

be the ordered observations, so that $\{t_i^n : i \leqslant n\} = \{t_i : i \leqslant n\}$.

3.1. *Brownian Motion*

Denote the linear interpolation between observed values by

$$L_n(s) = \frac{t_i^n - s}{t_i^n - t_{i-1}^n} f(t_{i-1}^n) + \frac{s - t_{i-1}^n}{t_i^n - t_{i-1}^n} f\big(t_i^n\big), \quad t_{i-1}^n \leqslant s \leqslant t_i^n,\ 0 \leqslant i \leqslant n. \tag{2}$$

Then $L_n(s)$ is the conditional expected value of $f(s)$ given the observations $\{t_i^n, f(t_i^n);\ i \leqslant n\}$. For $1 \leqslant i \leqslant n$ define

$$\rho_i^n \equiv \int_{t_{i-1}^n}^{t_i^n} \frac{ds}{(L_n(s) - M_n + z_n)^2}. \tag{3}$$

In the case of Brownian motion, these quantities are related to the probabilities at (1) by

$$P\big(f(t_{n+1}) < M_n - z_n \,\big|\, f(t_1) = x_1, \ldots, f(t_n) = x_n\big) = \exp\big(-2/\rho_i^n\big).$$

Given the first $n-1$ steps, the algorithm chooses the next observation in the subinterval with the largest value of ρ_i^{n-1} and evaluates the function within that interval. Variations are to choose the midpoint of the interval or to choose the point that maximizes the probability (1).

We now describe a particular version following this general approach. Let $\tau_n = \min_{1 \leqslant i \leqslant n} t_i^n - t_{i-1}^n$ denote the smallest distance between function evaluation points. The first two observation points of the algorithm are fixed: $t_1 = 1$ and $t_2 = 1/2$. Define

$$g(x) = 4\sqrt{x \log(1/x)}$$

for $0 < x \leqslant 1/2$. The function g is increasing and $g(x) \downarrow 0$ as $x \downarrow 0$. We define the ρ_i^n at Eq. (3) by setting $z_n = g(\tau_n)$. Then

$$\rho_i^n = \int_{t_{i-1}^n}^{t_i^n} \frac{ds}{(L_n(s) - M_n + g(\tau_n))^2} = \frac{t_i^n - t_{i-1}^n}{(f(t_{i-1}^n) - M_n + g(\tau_n))(f(t_i^n) - M_n + g(\tau_n))}.$$

The algorithm operates as follows: at each step, split the interval with the largest value of $\rho_i^n \equiv \rho^n$ at the midpoint. More precisely, suppose we have made k evaluations. Compute ρ_i^k, $1 \leqslant i \leqslant k$, and let i be an index such that $\rho_i^k \geqslant \rho_j^k$ for all $1 \leqslant j \leqslant k$. The next function evaluation is made at the midpoint

$$t_{k+1} = \frac{t_{i-1}^k + t_i^k}{2}.$$

The following theorem is proved in Calvin (2011a).

Theorem 1. *For the algorithm described above, there is a positive constant c such that as the number of observations $n \to \infty$,*

$$P\big(\Delta_n \leqslant n^{1/4} \exp(-c\sqrt{n})\big) \to 1.$$

The complexity analysis of optimization of the Brownian motion was initiated in Ritter (1990), where it was shown that for any nonadaptive algorithm using n function evaluations the average error is at least of order $n^{-1/2}$. Equi-spaced points gives an error of order $n^{-1/2}$, and thus achieves the lower bound for nonadaptive algorithms. We see from Theorem 1 that adaptive algorithms can be exponentially more powerful than nonadaptive algorithms.

A natural question is to what extent the convergence rate in Theorem 1 can be improved. In Calvin (2007), it was shown that for any (adaptive) algorithm using n function evaluations the average error is at least of order $\exp(-c\, n/\log(n))$ for some positive constant c.

The convergence rate described in Theorem 1 is sensitive to the probability model. If instead of the Wiener process the objective function were smooth, then the convergence rate would be much slower.

In some applications a parametric family of probabilities might be assumed for the objective function. Then as the optimization algorithm proceeds it would be reasonable to estimate the parameters of the probability measure based on the observations.

3.2. *Smooth Classes*

The Brownian motion paths are nowhere differentiable with probability one. A natural way to obtain a model for smooth functions is by integrating the Brownian paths.

Denote by W_r the *r-fold integrated Wiener measure* on $F = C^r([0,1])$; this is the Gaussian measure with covariance function

$$K_r(s,t) = \int_F f(s) f(t) P_r(df) = \int_{u=0}^{1} \frac{(s-u)_+^r (t-u)_+^r}{(r!)^2}\, du,$$

where $x_+ = \max(0, x)$. P_0 is the classical Wiener measure (corresponding to Brownian motion), and if $f \in C^{r-1}([0,1])$ is distributed according to W_{r-1}, then

$$g(t) = \int_{s=0}^{t} f(s)\, ds, \quad 0 \leqslant t \leqslant 1,$$

is distributed according to W_r.

In order to ensure that a sample path has a global minimizer in the interior of the unit interval, we construct a conditional r-fold integrated Wiener measure P_r by translating each path by a suitable polynomial so that prescribed boundary conditions are satisfied (Novak *et al.*, 1995). Let $r \geqslant 1$ denote the smoothness, and $\{a_i, 0 \leqslant i \leqslant r\}$ and $\{b_i, 0 \leqslant i \leqslant r\}$ the boundary values at 0 and 1, respectively. We consider a class of functions

$$F = \left\{ f \in C^r\left([0,1]\right) : f^{(i)}(0) = a_i,\ f^{(i)}(1) = b_i,\ \text{for } i = 0, 1, \ldots, r \right\}.$$

In order to ensure that the global minimum occurs in the interior of the interval, we assume that $a_1 < 0$ and $b_1 > 0$.

Under P_r, f has a continuous r-th derivative, but is nowhere $(r+1)$-times differentiable with probability one. This class of probability models has been used extensively in the average-case analysis of problems such as zero finding, integration, and function approximation; see Ritter (2000) and references therein, and Wasilkowski (1992). The class of models is of particular value for studying optimization since it provides a hierarchy of models with increasing smoothness, allowing us to study the impact of smoothness on the complexity of optimization. Furthermore, if $f \sim P_r$, then the vector process $(f(t), f'(t), \ldots, f^{(r)}(t))$ is a Markov process as well as a Gaussian process, and the combined techniques of the two classes aid in the analysis.

Calvin and Žilinskas (2005) construct an algorithm for the once-integrated Wiener process. The authors established that using the algorithm defined at Eq. (1) with constant $z_n = c > 0$, the error is of order $n^{-3/2}$. By contrast, Calvin and Žilinskas (2000a, 2000b) considered stationary Gaussian processes with twice continuously differentiable paths, for which the convergence rate (for constant z_n) is n^{-2}.

3.3. *Lower Bound*

In this section we consider lower bounds on the complexity of the global optimization problem for the probabilities P_r, $r \geqslant 1$. We are interested in the smallest error that can be attained with general algorithms that can evaluate the function or its derivatives at adaptively chosen points, and furthermore can stop adaptively.

The approximation to the minimizer is based on evaluations at a finite number of points, which we indicate by

$$A(f) = \phi\big(N_n(f)\big),$$

where the information $N_n(f)$ is given by

$$N_n(f) = \big[f^{(j_1)}(t_1),\, f^{(j_2)}(t_2), \ldots, f^{(j_n)}(t_n)\big].$$

The initial point t_1 and the order of derivative j_1 are fixed, and subsequent points and derivative orders are chosen adaptively. The total number of evaluations $n = n(f)$ is also chosen adaptively. For any $k > 1$, after the information

$$N_k(f) = \big[f^{(j_1)}(t_1),\, f^{(j_2)}(t_2), \ldots, f^{(j_k)}(t_k)\big]$$

has been computed, a termination rule determines if $n(f) > k$, and if so then t_{k+1} and j_{k+1} are computed from $N_k(f)$.

We call $n(f)$ the cardinality of information, and define the cost of an algorithm applied to the function f to be $n(f)$. Since we are interested in a lower bound, we ignore the computational cost of determining the evaluation points and constructing the approximation. The error, when applied to f, is

$$\Delta(f) \equiv f\big(A(f)\big) - \min_{0\leqslant t\leqslant 1} f(t).$$

For our probability P_r defined on F, we define the average cost and average error

$$\operatorname{cost}(\phi, N) = \int_F n(f)\, P_r(df),$$

and

$$\operatorname{error}(\phi, N) = \int_F \big[f(A(f)) - \min_{0\leqslant t\leqslant 1} f(t)\big] P_r(df) = \int_F \Delta(f)\, P_r(df).$$

The local error

$$E\big(\Delta(f)\big|N_n(f)\big) = E\Big(f\big(A(f)\big) - \min_{0\leqslant t\leqslant 1} f(t)\big|N_n(f)\Big)$$

is minimized by choosing $A(f)$ to be a minimizer of the conditional mean; then

$$f\big(A(f)\big) = \min_{0\leqslant t\leqslant 1} E\big(f(t)\big|N_n(f)\big).$$

With this optimal choice the local error is

$$E\big(\Delta(f)\,|\,N_n(f)\big) = E\Big(\min_{0\leqslant t\leqslant 1} E\big(f(t)\big|N_n(f)\big) - \min_{0\leqslant t\leqslant 1} f(t)\big|N_n(f)\Big). \tag{4}$$

The following result is proved in Calvin (2011b).

Theorem 2. *Let $\epsilon > 0$ and $r \geqslant 1$ and consider an arbitrary algorithm that has average error at most ϵ. There exists a positive number C, depending only on the boundary values $\{a_i, b_i,\ 0 \leqslant i \leqslant r\}$ and r, such that the average cost of the algorithm is at least*

$$C \cdot \log(1/\epsilon)^{\frac{1}{(2r-1)(2r+1)}}.$$

It is interesting to compare results for the global optimization problem with results for the zero-finding problem; both have been studied for the case of the conditional r-fold integrated Wiener measures P_r. For zero finding, any algorithm that uses a fixed number of function evaluations requires on order $\log(1/\epsilon)$ evaluations to obtain an ϵ-approximation. In contrast, an algorithm that uses an adaptive stopping rule obtains an ϵ approximation with on average order $\log\log(1/\epsilon)$ function evaluations. The algorithm that achieves this bound uses only function values, and not derivatives; see Novak *et al*. (1995).

4. Asymptotic Results for Multivariate Optimization

The idea of the optimization algorithm based on (1) extends naturally to higher dimensions, but the probability computations are much more difficult. One approach is to subdivide the domain into polyhedral subsets and define for each subset an analog of the quantity defined at Eq. (3). Computing the distribution of the error for the resulting algorithms is difficult, but an asymptotic analysis of the error is feasible for some subdivision strategies. We outline the results for two such strategies in this section.

4.1. *Rectangular Subdivision*

The first algorithm operates by decomposing $[0, 1]^d$ into (hyper)-rectangles as follows. Given a current decomposition, choose one of the rectangles (according to the maximal value of a criterion to be defined below) and bisect it along the longest axis by evaluating the function at up to 2^{d-1} midpoints of the longest rectangle edges.

Let v_n denote the smallest volume of a hyperrectangle after n iterations. Define

$$q \equiv \frac{3 \cdot 2^{2/3} e^{-1}}{2\log(2)} \approx 1.27$$

and

$$g(x) = q \cdot d\big(x \log(1/x)\big)^{2/d} \tag{5}$$

for $0 < x \leqslant 1/2$ and $g(1) = q \cdot d$. Let L_n denote the multilinear function that has the same values as f at the vertices of the smallest enclosing rectangle. Note that M_n is equal to the global minimum of L_n. For $1 \leqslant i \leqslant n$ set

$$\rho_i^n \equiv \frac{|R_i|}{(L_n(c_i) - M_n + g(v_n))^{d/2}}, \tag{6}$$

where we denote the volume of set A by $|A|$ and c_i is the centre of rectangle i.

The idea is that the criterion ρ_i^n tends to be large for large (unexplored) rectangles, and also large when the local approximation $L_n(c_i)$ is near the smallest observed value M_n. The additional term $g_n(v_n)$ regulates the tradeoff so that the error converges to zero for any continuous f, and the desired asymptotic error bound can be obtained.

The error bound will depend on the regularity of the function as measured by a certain seminorm. For a compact set K and $f \in C^2(K)$, define the seminorm

$$\left|D^2 f\right|_{\infty,K} \equiv \sup_{x\in K} \sup_{\substack{u_1,u_2\in\mathcal{R}^d \\ |u_i|=1}} \left|D_{u_1} D_{u_2} f(x)\right|,$$

where $D_y f$ is the directional derivative of f in the direction y.

The following theorem is proved in Calvin *et al*. (2015).

Theorem 3. *Suppose that $f \in F$ has a unique global minimizer x^* in the interior of the domain, and denote the matrix of second-order partial derivatives at the minimizer by $D^2 f(x^*)$. There is a number $n_0(f)$ such that for $n \geqslant n_0(f)$,*

$$\Delta_n \leqslant \frac{1}{8}\left|D^2 f\right|_{\infty,[0,1]^d} (q\cdot d) \exp\left(-\sqrt{n}\,\beta(f,d)\right),$$

where

$$\beta(f,d) = \left(\frac{2\Gamma(1+d/2)\sqrt{\det(D^2 f(x^*))}}{(2\pi)^{d/2}(d(d+1))2^{d-1}(2(q\cdot d))^{d/2}}\right)^{1/2}.$$

The limiting error is smaller the larger the determinant of the second derivative at the minimizer; a larger second derivative allows the search effort to concentrate more around the minimizer. While the convergence rate in terms of the number of function evaluations n is quite fast, the term $\beta(f,d)$ decreases exponentially fast as d increases.

4.2. *Delaunay Subdivision*

The rectangular decomposition suffers from some drawbacks. At each iteration, as many as 2^{d-1} function evaluations are required before making another decision. Also, in high dimensions rectangles are not too efficient in terms of interpolation accuracy. An approach based on Delaunay triangulations has been successful in the bivariate case. A main advantage compared with the rectangular subdivisions is that there is only one function evaluation on each iteration (at the centre of a simplex). The main obstacle is to choose points so that the simplexes of the Delaunay triangulation have a certain quality (avoiding so-called "slivers").

As before, assume that f has been evaluated at the vertices of the hypercube $[0,1]^d$. A triangulation is a partition of the cube into simplexes with the set of vertices equal to the set of observations. Suppose that the algorithm has evaluated the function at n points $P = \{x_1, x_2, \ldots, x_n\}$. The Delaunay triangulation has the property that no point of P lies in the interior of the circumsphere of any of its constituent simplexes. In the planar case,

this triangulation has the desirable property that it maximizes the minimum angle of any simplex over all triangulations of P. For a given point set, some simplexes of the Delaunay triangulation may have small interior angles, which leads to large error bounds in our analysis. We require a Delaunay mesh with a certain quality which is defined in terms of the quantity

$$Q(T) \equiv \frac{rad(T)^2}{|T|^{2/d}}, \tag{7}$$

where $rad(T)$ is the radius of the sphere circumscribing T. For our method we require that this quantity be bounded above over all simplexes:

$$\max_i Q(T_i) \leqslant q(d),$$

where the bound can depend on the dimension. By adding extra (so-called Steiner) points, a mesh with certain quality guarantees can be constructed in the 2-dimensional case. For example, with $d = 2$, we can guarantee a bound of

$$Q(T) \leqslant \frac{16}{\sqrt{7}} \equiv q(2) \approx 6.0474. \tag{8}$$

The algorithm that we now describe works for any dimension d, though for $d > 2$ we have no quality bound like (8). Let v_n denote the smallest volume of a simplex after n iterations and modify the function g define at (5) by replacing the factor $q \cdot d$ by a function $q(d)$:

$$g(x) = q(d)\big(x \log(1/x)\big)^{2/d}$$

for $0 < x \leqslant 1/2$ and $g(x) = q$ for $x > 1/2$. Let $t_{i,j}$, $j = 1, 2, \ldots, d+1$, denote the vertices of simplex T_i and $\hat{f}_i$ the average of the function values at the vertices:

$$\hat{f}_i = \frac{1}{d+1} \sum_{j=1}^{3} f(t_{i,j}).$$

Define

$$\rho_i^n \equiv \frac{|T_i|}{(\hat{f}_i - M_n + g(v_n))^{d/2}}. \tag{9}$$

Suppose that f has been evaluated at n points, and the Delaunay triangulation has been constructed. For each simplex, compute ρ_i^n. Choose the next point x_{n+1} as the centroid of the simplex with the largest value of ρ_i^n. Evaluate $f(x_{n+1})$ and construct the new Delaunay triangulation including the new point. Set $M_{n+1} = \min(M_n, f(x_{n+1}))$. Note that at each step we make a single new evaluation, unlike in the rectangular decomposition case.

Assume that f has a unique minimizer $x^* \in (0,1)^d$ and let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d > 0$ be the eigenvalues of $D^2 f(x^*)$, so $\det D^2 f(x^*) = \lambda_1 \lambda_2 \ldots \lambda_d$. It is shown in Calvin and Žilinskas (2014) that the error after n evaluations is of order $\exp(-c\sqrt{n})$ for a constant c depending on the objective function f. More precisely,

Theorem 4. *Let dimension $d = 2$ and assume that each triangle T in the Delaunay triangulation has a quality bound $Q(T) \leqslant q$. Then the error Δ_n after n iterations satisfies*

$$\liminf_{n\to\infty} n^{-1/2} \log\left(\frac{1}{\Delta_n}\right) \geqslant \frac{(\lambda_1\lambda_2)^{1/4}}{2\sqrt{6q\pi}}. \tag{10}$$

The assumption of unique global minimizer is only for convenience. In the case of multiple global minimizers the constant on the right-hand size is replaced by a function of the eigenvalues at all of the global minimizers. However, for the convergence rate in (10) to hold it is essential that the global minimizers be isolated.

This result is a generalization to two dimensions of the main result in Calvin *et al.* (2012). That result had the same $n^{-1/2}$ normalizing rate that appears in (10), with a different constant on the right-hand side. In particular, if the univariate smooth function f has a unique global minimizer at $t^* \in (0,1)$, then for the one-dimensional algorithm,

$$\liminf_{n\to\infty} n^{-1/2} \log\left(\frac{1}{\Delta_n}\right) \geqslant \left(\frac{f''(t^*)}{12}\right)^{1/4}.$$

Acknowledgement. Many of the results described in this paper are based on joint work and conversations with Antanas Žilinskas.

References

Calvin, J.M. (2007). A lower bound on complexity of optimization on the Wiener space. *Theoretical Computer Science*, 383, 132–139.

Calvin, J.M. (2011a). An adaptive univariate global optimization algorithm and its convergence rate under the Wiener measure. *Informatica*, 22, 471–488.

Calvin, J.M. (2011b). A lower bound on complexity of optimization under the r-fold integrated Wiener measure. *Journal of Complexity*, 27, 404–416.

Calvin, J., Žilinskas, A. (2005). One-dimensional global optimization for observations with noise. *Computers and Mathematics with Applications*, 50, 157–169.

Calvin, J., Žilinskas, A. (2014). On a global optimization algorithm for bivariate smooth functions. *Journal of Optimization Theory and Applications*, 163, 528–547.

Calvin, J., Chen, Y., Žilinskas, A. (2012). An adaptive univariate global optimization algorithm and its convergence rate for twice continuously differentiable functions. *Journal of Optimization Theory and Applications*, 155, 628–636.

Calvin, J., Gimbutiene, G., Phillips, W., Žilinskas, A. (2015). On a global optimization algorithm for multivariate smooth functions. Submitted for publication.

Kushner, H. (1962). A versatile stochastic model of a function of unknown and time-varying form. *Journal of Mathematical Analysis and Applications*, 5, 150–167.

Mockus, J. (1972). On Bayesian methods of search for the extremum. *Automatics and Computers*, 3, 53–62.

Novak, E., Ritter, K., Woźniakowski, H. (1995). Average case optimality of a hybrid secant-bisection method. *Mathematics of Computation*, 64, 1517–1539.
Ritter, K. (1990). Approximation and optimization on the Wiener space. *Journal of Complexity*, 6, 337–364.
Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Mathematics*, Vol. 1733. Springer, Berlin.
Tikhomirov, A.S. (2006). On the Markov homogeneous optimization method. *Computational Mathematics and Mathematical Physics*, 3, 361–375.
Wasilkowski, G. (1992). On average complexity of global optimization problems. *Mathematical Programming*, 57, 313–324.
Zhigljavsky, A., Žilinskas, A. (2008). *Stochastic Global Optimization*. Springer-Verlag, New York.
Zhigljavsky, A., Žilinskas, A. (2016). Stochastic global optimization: a review on the occasion of 25 years of *Informatica*. *Informatica*, 27, 229–256.
Žilinskas, A. (1985). Axiomatic characterization of a global optimization algorithm and investigation of its search strategy. *OR Letters*, 4, 35–39.

J. Calvin is a professor in the department of computer science at the New Jersey Institute of Technology. His research interests are in global optimization and probabilistic analysis of algorithms.

Tikimybiniai globaliosios optimizacijos modeliai

James M. CALVIN

Straipsnyje pateikta globaliosios optimizacijos algoritmų, pagrįstų tikimybiniais tikslo funkcijų modeliais, apžvalga, akcentuojanti žinomus apžvelgtų algoritmų sudėtingumo įverčius. Tikimybinių modelių reikšmė čia dvejopa: pirma, jais grindžiama algoritmo idėja; antra, jų atžvilgiu vertinamas algoritmų efektyvumas.