Global Optimization Using Space-Filling Curves and Measure-Preserving Transformations

Hime A. e Oliveira Jr. and Antonio Petraglia

Abstract. This work proposes a multi-start global optimization algorithm that uses dimensional reduction techniques based upon approximations of space-filling curves and simulated annealing, aiming to find global minima of real-valued (possibly multimodal) functions that are not necessarily well behaved, that is, are not required to be differentiable or continuous. Given a real-valued function with a multidimensional and compact domain, the method builds an equivalent, one-dimensional problem by composing it with a space-filling curve (SFC), searches for a small group of candidates and returns to the original higher-dimensional domain, this time with a small set of "promising" starting points. Finally, these points serve as seeds to the algorithm known as Fuzzy Adaptive Simulated Annealing, aiming to find the global optima of the original cost functions. New SFCs are built with basis on the well-known Sierpiński SFC, a subtle modification of a theorem by Hugo Steinhaus and several results of ergodic theory.

1 Introduction

A significant number of techniques for global optimization of numerical functions based on space-filling curves or its approximations have been proposed [2, 6]. One common characteristic shared by these contributions is that the functions under study have certain regularity properties, such as being of Lipschitz type or differentiable. If those properties are not satisfied, or we cannot prove whenever they are, the problem is outside the scope of the corresponding method. It is thus of interest

Hime A. e Oliveira Jr
National Cinema Agency, Rio de Janeiro, Brazil
e-mail: hime@engineer.com

Antonio Petraglia
Department of Electrical Engineering, Federal University of Rio de Janeiro,
Rio de Janeiro, Brazil
e-mail: antonio@pads.ufrj.br

A. Gaspar-Cunha et al. (Eds.): Soft Computing in Industrial Applications, AISC 96, pp. 121–130.
springerlink.com

to devise a way of handling such difficulties, as well. Another situation that occurs frequently is related to the poor precision attained by some methods whenever the domain dimension gets higher - such a fact is related to the difficulty in minimizing the resulting one-dimensional auxiliary functions that exhibit a large number of local minima and very noisy graphs, fractal-like indeed. Considering that the great majority, if not all, of global minimization methods fail in such extreme situations, such an approach is in some cases considered to be ineffective. The reasons for this are related to the way the multidimensional domain is filled up and to the quality of the one-dimensional minimization process applied to the resulting auxiliary function.

To be successful in practice, the global optimization approach must devise a good approximation of a space-filling curve (SFC) whose image is, or contains, the compact domain of the function under study, and to use a one-dimensional global minimization algorithm that can find precise approximations to optimizing points of the composed map, whose domain is a closed interval, say [0,1], and that assumes real values, possessing the same extremes as the original function, among which the desired global optimum is included. Unfortunately, such conditions are not easily satisfied and past efforts were only partially successful in finding good results. In [2], for example, an interesting and promising paradigm is presented, but the results focus on low dimensional spaces. Another issue is related to the adequacy of the chosen way for filling up the original multidimensional domain - by projecting the image of an approximation of a particular SFC onto the space generated by vectors corresponding to directions that have larger variance (by means of principal component analysis), we can show whether there are poorly visited regions, inside which extreme points could be located.

To find SFCs capable of overcoming such obstacles, measure-preserving transformations, key theorems of general topology and ergodic theory were taken as inestimable tools [1, 3, 4]. In this paper, we assume, without loss of generality, that all optimization problems are related to unconstrained global minimization of real functions and all SFCs have the unitary interval [0,1] as their domain.

It's worth to highlight that an important qualitative property of SFCs is their ability to "sweep" deterministically high-dimensional domains, so as to improve the likelihood of finding good "seeds" for complementary, posterior global optimization stages, taking into account the existence of several methods whose final results depend strongly upon their starting points [7]. In this fashion, it is of interest to investigate new ways of finding adequate starters, particularly those located in attraction basins of global optima.

Despite of the existence of many good global optimization methods that could be used in such a posterior stage, we have chosen the fuzzy adaptive simulated annealing algorithm, taking into account its excellent performance in difficult optimizations tasks [9] and the maturity of the adaptive simulated annealing paradigm itself [7, 10]. Nevertheless, it's possible to replace it by any other method, especially those ones depending on good starting points.

2 Auxiliary Theoretical Results

In this work, we define a SFC as a surjective and continuous function from a real interval, say [0,1], to a compact subset of a finite-dimensional vector space, which can be identified to $\mathbb{R}^n$, the n-dimensional Euclidean space. The SFCs were well studied in the past and there are many theoretical results stating necessary conditions for their existence [11]. Besides, long before the invention of digital computers, a number of great mathematicians have proposed constructive examples and established several interesting properties [11] of SFCs. More recently, researchers have found ways to apply previous knowledge about SFCs to various relevant areas, including global optimization of numerical functions. In this paper, the fundamental idea is to compose a given objective function with a SFC corresponding to a compact superset of the respective domain. In such a way we can always reduce a multivariate problem to a univariate one. Hence, at least in theory, it would be possible, by solving the auxiliary one-dimensional problem, to go back to the n-dimensional domain and find the desired optimum point. Unfortunately, when such ideas are brought to the digital computer realm, some complications arise, particularly in high-dimensional domains.

The main drawback concerning implementation issues is that virtually all curves idealized in the far past did not take into account the finite word length of digital computers (one good reason for that is that digital computers were invented long after their synthesis). Typically, the first proposed SFCs were based on infinite expansions and used, for instance, the property that elements in [0,1] can be represented as $0.t_1t_2t_3t_4t_5...$ in a given basis B, where each t_i is an integer between 0 and $B-1$ (extremes included). It is thus necessary to find adequate approximations of SFCs if we want to pursue this kind of approach. Initially, a reasonable alternative seems to be the use of (approximations of) Sierpiński SFCs, taking into account the availability of their precise defining formulas, as follows [11]:

$$x(t) = f(t), \;\; y(t) = f(t - 1/4), \;\; t \in [0,1] \tag{1}$$

where f is a bounded, even and continuous real function whose expression is given by

$$f(t) = \frac{\Theta(t)}{2} - \frac{\Theta(t)\Theta(\tau_1(t))}{4} + \frac{\Theta(t)\Theta(\tau_1(t))\Theta(\tau_2(t))}{8} - ... \tag{2}$$

The 1-periodic functions $\Theta(t)$ and $\tau_k(t)$ are defined in [11], so that the resulting curve $(x(t), y(t))$ is a 2-dimensional SFC and shows to be well-suited to numerical calculations. To build higher-dimensional SFCs starting from this one, some results were crucial, as stated by the following theorems, whose proofs can be found in [11]. First, however, we need to present some definitions.

Definition 1. A function $\varphi : [0,1] \to \mathbb{R}$ is uniformly distributed with respect to the Lebesgue measure if, for any (Lebesgue) measurable set $A \subset \mathbb{R}$, we have

$$\Lambda_1(\varphi^{-1}(A)) = \Lambda_1(A) \tag{3}$$

where Λ_1 is the Lebesgue measure in the real line.

Definition 2. n measurable functions $\varphi_1, \varphi_2, ..., \varphi_n : [0,1] \to \mathbb{R}$ are stochastically independent with respect to Lebesgue measure if, for any n measurable sets $A_1, A_2, ..., A_n \subset \mathbb{R}$,

$$\Lambda_1(\bigcap_{j=1}^{n} \varphi_j^{-1}(A_j)) = \prod_{j=1}^{n} \Lambda_1(\varphi_j^{-1}(A_j)) \tag{4}$$

Theorem 1. *(H. Steinhaus) If $\varphi_1, \varphi_2, ..., \varphi_n : [0,1] \to \mathbb{R}$ are continuous, non-constant and stochastically independent with respect to the Lebesgue measure, then*

$$f = (\varphi_1, \varphi_2, ..., \varphi_n) : [0,1] \to \varphi_1([0,1]) \times \varphi_2([0,1]) \times ... \times \varphi_n([0,1]) \tag{5}$$

is a SFC.

Theorem 2. *If $f = (\varphi, \psi) : [0,1] \to [0,1] \times [0,1]$ is (Lebesgue) measure-preserving and onto, then its coordinate functions φ, ψ are uniformly distributed and stochastically independent.*

Taking into account that the Sierpiński SFC is measure-preserving ([11], page 111), we conclude that its coordinate functions are uniformly distributed and stochastically independent, and can be used to synthesize higher-dimensional SFCs with coordinates

$$\begin{aligned} x_1(t) &= \varphi(t) \\ x_2(t) &= \varphi \circ \psi(t) \\ &\dots\dots\dots\dots \\ x_n(t) &= \varphi \circ \psi \circ \psi \circ \dots \circ \psi(t) \\ t &\in [0,1] \end{aligned} \tag{6}$$

that are non-constant, continuous and stochastically independent. Unfortunately, after a few experiments it was clear that for higher-dimensional domains (around 8 dimensions), approximations of such "pure" Sierpiński based SFCs failed to fill up adequately the compact domains of interest, as will be shown in the sequel (Fig. 1(a)). This fact is due to distortions caused by numerical approximations, despite the theoretical curve being really a SFC one. Aiming to find a better curve, we composed the original Sierpiński function with an invertible (Lebesgue) measure-preserving transformation that is a natural extension of a particular Generalized Lüroth Series transformation [4], mapping [0,1)x[0,1) onto itself. That new mapping was found through a new partition, that we called the bisection partition, as defined in (7). To generate a new SFC mapping [0,1] onto [-1,1]x[-1,1], it was necessary to use linear homeomorphisms from [-1,1]x[-1,1] to [0,1]x[0,1] and vice-versa. Let us denote such a transformation as τ, and derive its expression as shown below

$$D = \{1,2,3,...\} = \mathbb{N}, \;\; I_k = [l_k, r_k) = [\frac{1}{2^k}, \frac{1}{2^{k-1}}), \; k \in D$$

$$
\begin{aligned}
s(x) &= \frac{1}{r_k - l_k} = \frac{1}{\frac{1}{2^{(k-1)}} - \frac{1}{2^k}} = 2^k,\ x \in I_k \\
h(x) &= \frac{l_k}{r_k - l_k} = \frac{1}{2^k} 2^k = 1,\ \ x \in I_k \\
&\left\{ \begin{array}{l} s_1(x) = s(x) = 2^k \\ h_1(x) = h(x) = 1 \end{array} \right\},\ x \in I_k \\
T(x) &= x.s(x) - h(x) = 2^k x - 1,\ x \in I_k \\
\tau(x,y) &= (T(x), \frac{h_1(x) + y}{s_1(x)}) = \\
(2^k x - 1, \frac{1}{2^k} + \frac{y}{2^k}) &= (2^k x - 1, 2^{-k}(1+y)),\ x \in I_k
\end{aligned}
\tag{7}
$$

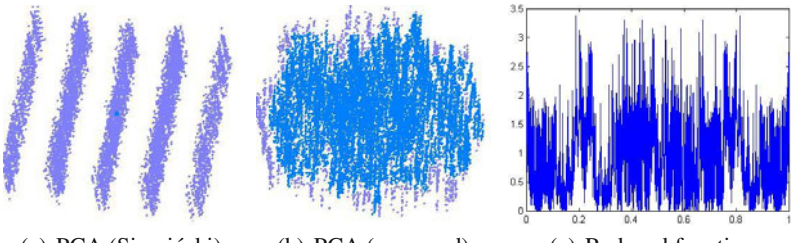
Here, the $\{I_k : k \in \mathbb{N}\}$ form the bisection partition.

The proposed SFC is given by the following sequence of operations:

$$
[0,1] \rightarrow [-1,1] \times [-1,1] \rightarrow [0,1] \times [0,1] \rightarrow [0,1] \times [0,1] \rightarrow [-1,1] \times [-1,1] \tag{8}
$$

Sierpiński Homeomorphism τ Homeomorphism

It should also be observed that although τ was initially defined only in [0,1)x[0,1), we extended it to [0,1]x[0,1] in an obvious way, so that the composite path can reach all regions of the desired domain. To assess the filling degree of the curves relatively to the compact set $[-1,1]^n$, we present in Figs. 1(a) and 1(b) the plots produced by PCA projections of the corresponding multidimensional curves onto the 2 maximum variance directions. Qualitatively, it can be stated that the more filled the graph area is, the more adequate the filling curve will be. The generated paths consisted of 20,000 8-dimensional points of each kind of curve, and conventional PCA was carried out through a customized computer program. The parameterizing domain was chosen as the closed interval [0,1] in all cases. As can be seen from the plot in Fig. 1(a), the projected points of "pure" Sierpiński curves did not fill adequately the maximum variance region. The composed transformation, in Fig. 1(b), presented substantially better performance, showing denser and more uniform covering. Let



(a) PCA (Sierpiński) (b) PCA (proposed) (c) Reduced function

Fig. 1 Domain projections and example of dimensional reduction ($2 \rightarrow 1$)

us thus analyze what happens to the graph of the very well behaved bi-dimensional function $f(x_1,x_2) = (x_1 - 0.3)^2 + (x_2 - 0.3)^2$, restricted to the square [-1,1]x[-1,1], when it is composed with the proposed SFC, yielding a univariate, real-valued function with compact domain [0,1]. The result of uniformly sampling the interval [0,1] by 10000 points is displayed in Fig. 1(c), from which we conclude that a simple dimensional reduction from 2 to 1 transforms a smooth surface into the very rough line $g(t) = f(\varphi_1(t), \varphi_2(t))$. This phenomenon worsens as the original number of dimensions increases. On the other side, by carefully observing Fig. 1(c), we notice that there are regions where the maximum (3.38) and minimum (0) values of f in [-1,1]x[-1,1] are visibly approximated, and could serve as hints to start more effectively preexisting and efficient global optimization algorithms, be them stochastic or deterministic.

Returning to the above example function we find that for t = 0.2908 the one-dimensional function attains the value 0.000119, which is the minimum of the 10000 points used in the illustrative discretization and an approximation for the actual 0 value, corresponding to (0.3, 0.3). Going back to the two-dimensional domain, we find (0.299, 0.289) as the associated bi-dimensional point, which is reasonably close to the actual minimizer. It is worth noting that although Steinhaus theorem is stated only for continuous functions, it is also true for surjective (over [0,1]), piecewise continuous coordinate functions, as well, as is the case for the components of the proposed SFC. In fact, we could state the following theorem, whose proof follows from the corresponding one of Steinhaus theorem in [11].

Theorem 3. *(modified Steinhaus) If* $\varphi_1, \varphi_2, \ldots, \varphi_n : [0,1] \to [0,1]$ *are piecewise continuous, surjective, non-constant and stochastically independent with respect to the Lebesgue measure, then f (defined below) is a SFC.*

$$f \equiv (\varphi_1, \varphi_2, ..., \varphi_n) : [0,1] \to \varphi_1([0,1]) \times \varphi_2([0,1]) \times \ldots \times \varphi_n([0,1]) \qquad (9)$$

3 Fuzzy Adaptive Simulated Annealing

Adaptive simulated annealing (ASA) [7] is a sophisticated and rather effective global optimization method. The ASA technique is particularly well suited to applications involving neuro-fuzzy systems and neural network training [10], thanks to its superior performance and simplicity. Unfortunately, stochastic global optimization algorithms typically present certain periods of poor improvement in their way to a global optimum. In simulated annealing implementations, that behavior is mainly due to the cooling schedule, whose speed is limited by the characteristics of probability density functions, which are employed with the purpose of generating new candidate points. In this fashion, if we choose to employ the so-called Boltzmann annealing, the temperature has to be lowered at a maximum rate of $T(k) = T(0)/ln(k)$. In the case of fast annealing, the schedule becomes $T(k) = T(0)/k$, if assurance of convergence with probability 1 is to be maintained, resulting in a faster schedule. The approach based on ASA has an even better default scheme, given by

$$T_i(k) = T_i(0)\exp(-C_i k^{1/D}) \tag{10}$$

because of its improved generating distribution. The constant C_i is a user-defined parameter, and D is the number of independent variables of the function under minimization (dimension of the domain). Notice that subscripts indicate independent evolution of temperatures for each parameter dimension. In addition, it is possible to take advantage of simulated quenching, that is,

$$T_i(k) = T_i(0)\exp(-C_i k^{Q_i/D}) \tag{11}$$

where Q_i is termed the quenching parameter. By attributing to Q_i values greater than 1 we obtain a gain in speed, but the convergence to a global optimum is no longer assured [7]. Such a procedure could be used for higher-dimensional parameter spaces, whenever computational resources are scarce. The internal structure of a successful approach to accelerate the ASA algorithm, using a simple fuzzy controller that dynamically adjusts certain user's parameters related to quenching, is described in [9] - the so-called fuzzy ASA algorithm. As in any other method aiming at global optimization of arbitrary numerical functions, ASA and fuzzy ASA techniques could benefit from the choice of good starting points. Accordingly, the main point of the present work is to find a small set of good seeds able to avoid convergence to suboptimal regions.

4 Proposed Algorithm

We propose the following algorithm to find a global minimum of a given function $f : C \to \mathbb{R}$, where C is a compact subset of some n-dimensional Euclidean space $\mathbb{R}^n$. No condition of regularity, such as differentiability or even continuity, is imposed on f, and C is usually a hyper-rectangle. If it is not, we can always find one hyper-rectangle that contains it, taking into account its compactness. So, from this point on we assume that $C = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$. The algorithm is:

(i) Using (6) and the sequence of operations shown in (8), find an SFC $\varphi = (\varphi_1, \varphi_2, ..., \varphi_n) : [0,1] \to [-1,1] \times [-1,1] \times ... \times [-1,1]$;
(ii) Transform the original, multidimensional minimization problem into a unidimensional one by composing φ and a linear isomorphism $\Psi : [-1,1] \times [-1,1] \times \ldots \times [-1,1] \to [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$, defining g as the composition of φ, Ψ and f, from [0,1] onto $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$, sharing with f the same extreme values;
(iii) Submit g to a one-dimensional global minimization process and find a finite subset of best candidates to global minimizers of f, say $\{t_1, t_2, t_3, ..., t_N\}$, contained in [0,1]. In this work, this set has N=3 elements, but such a number could be easily reconfigured, if necessary;

(iv) Compute the set $\phi = \{\Psi(\varphi(t_1)), \Psi(\varphi(t_2)), \ldots, \Psi(\varphi(t_N))\}$, jumping back to the original domain;
(v) Use the elements of ϕ as starting points for the fuzzy ASA algorithm;
(vi) Choose the best point (corresponding to the minimum value of f) as the final output of the algorithm.

The unidimensional minimization process in (iii) is to be chosen by the implementer. In this work we used intentionally a simple scheme in our experiments (uniform sequential search), aiming to highlight the filling ability of the proposed SFC.

5 Experiments

As noticed in [8], it is usual in the literature to employ certain sets of test functions for evaluation of optimization methods. However, the chosen problems may not be the best ones for testing global optimization algorithms, as the functions belonging to them are relatively simple and regions of attraction of the global minimizers could be easily caught, despite their complicated appearances. Consequently, it is argued that they do not present sufficient difficulty to stress the minimization ability of new optimization approaches. Hence, it is necessary to idealize more sophisticated and systematic tests to verify their performances. To assess the effectiveness of the proposed method, a scheme similar to the one used in [8] was adopted, by employing a certain class of test functions, specified in Table 1, produced by the GKLS generator [5], which allows us to evaluate algorithms in a more complete way. It generates classes of 100 test functions having the same number of local minima plus one global minimum, supplying complete parametric information about each of the functions, such as their dimensions, the values of local minimizers and respective coordinates, placement and sizes of attraction regions of the global minimizer, which are described by parameters r_g (radius of the approximate attraction region of the global minimizer) and d (distance from the global minimizer to the paraboloid vertex). We refer the reader to [5] for more details. In what follows, we consider a global minimum found when candidate points reach a ball B_i of radius $\rho = 0.01\sqrt{N}$, where N is the Euclidean dimension of the function domain, that is,

$$B_i = \{y \in \mathbb{R}^N : \|y - y_i^G\| \leq \rho\} \tag{12}$$

where y_i^G is the global minimizer of the i-th function in a given test class and $i \in \{1, ..., 100\}$. Unlike in the original tests, just one function class (shown in Table 1) was used in the experiments. It should also be noted that all functions are non-differentiable, for (expected) greater difficulty. The present method was tested against the best one presented in [8], therein termed ALI. The authors proposed 4 new global optimization methods denoted as AG, AGI, AL and ALI. Figure 2 displays the number of function evaluations and respective global minima found by

the proposed technique, in a similar manner as those reported in [8] to compare AG and ALI for functions of class 5. As indicated in Table 2, the proposed algorithm produced better performance.

Table 1 Parameters pertaining to the function class used in the experiments

Function class	Domain dimension	No. of local minima	Global min- imum value	d	r_g	Function type
5	4	10	-1	0.66	0.33	ND

Table 2 ALI and the proposed method minimizing 100 ND-type class 5 functions

Method	Average number of function evaluations	Maximum number of function evaluations in individual minimization operations
ALI	14910	48210
Proposed	**10716.5**	**17350**

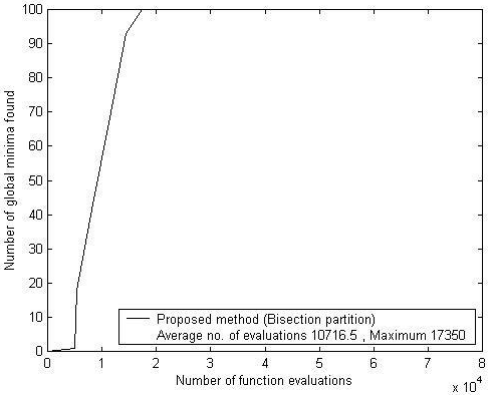


Fig. 2 Results for 100 class 5 functions - average 10716.5 and maximum 17350 evaluations

6 Conclusions

This work presented a multi-start approach for global minimization of multidimensional functions using dimensional reduction and a nonlinear stochastic method. When compared to previously published techniques, the method showed superior performance and adequacy for use in difficult cost functions. This is mainly due to the ability of space-filling curves in serializing the space, thereby allowing us to reduce the search domain to real line intervals. As an important and desirable

byproduct, the method is able to find points located in the attraction basins of global minimum points that will be used as seeds when jumping back to the original, multidimensional domain. That ability is invaluable in functions presenting large planar regions, where the great majority of methods get trapped, due mainly to the lack of differential indications. Although the reduced function typically presents a fractal-like graph even whenever the original one is smooth or "well-behaved", existing one-dimensional optimization techniques can obtain good estimates of near optimal points.

Finally, it should be said that although the method could be tested against many other classes of multimodal functions, in this work we have focused on (which we consider) the hardest test used in [8] (cited in section 4.5 and shown in Figure 12 of that reference), for the sake of exact comparison of our findings to the best performing method therein shown, namely ALI (in truth, all benchmarks presented in [8] were privately done and our method performed well, but, for lack of space, we are not presenting them here).

References

1. Boyarsky, A., Góra, P.: Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension. Birkhäuser, Boston (1997)
2. Cherruault, Y., Mora, G.: Optimisation Globale - Theorie des courbes alpha-denses. Economica, Paris (2005)
3. Choe, G.H.: Computational Ergodic Theory. Springer, Berlin (2005)
4. Dajani, K., Kraaikamp, C.: Ergodic Theory of Numbers. The Mathematical Association of America, Washington DC (2002)
5. Gaviano, M., Kvasov, D.E., Lera, D., Sergeyev, Y.D.: Software for generation of classes of test functions with known local and global minima for global optimization. ACM TOMS 29(4), 469–480 (2003)
6. Goertzel, B.: Global Optimization with Space-Filling Curves. Applied Mathematics Letters 12, 133–135 (1999)
7. Ingber, L.: Adaptive simulated annealing (ASA): Lessons learned. Control and Cybernetics 25(1), 33–54 (1996)
8. Lera, D., Sergeyev, Y.D.: Lipschitz and Hölder global optimization using space-filling curves. Applied Numerical Mathematics 60, 115–129 (2010)
9. Oliveira Jr., H.A.: Fuzzy control of stochastic global optimization algorithms and VFSR. Naval Research Magazine 16, 103–113 (2003)
10. Rosen, B.: Function optimization based on advanced simulated annealing, `http://www.inger.com`
11. Sagan, H.: Space-Filling Curves. Springer, New York (1994)