ARTICLE IN PRESS

Applied Mathematics and Computation xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain

C. Voglis *, I.E. Lagaris

Department of Computer Science, University of Ioannina, P.O. Box 1186, Ioannina 45110, Greece

ARTICLE INFO

Keywords:
Global optimization
Multistart
Stohastic approach
Local search
Adaptive probability

ABSTRACT

A stochastic global optimization method based on Multistart is presented. In this, the local search is conditionally applied with a probability that takes in account the topology of the objective function at the detail offered by the current status of exploration. As a result, the number of unnecessary local searches is drastically limited, yielding an efficient method. Results of its application on a set of common test functions are reported, along with a performance comparison against other established methods of similar nature.

1. Introduction

Global optimization (GO) has received a lot of attention in recent years [1], due to the ever emerging scientific and industrial demand. For instance the description of the stable conformations of a molecule [2–4], the management of mutual funds [5–8], location and allocation issues [9,10], engineering design and the design of drugs [11,12], to mention a few topics, are in need of efficient global optimization techniques.

There exist several categories of GO methods. We distinguish two main classes: the deterministic [13,14] and the stochastic one. For a detailed account on classification of stochastic algorithms we refer to [15]. Deterministic methods provide a theoretical guarantee of locating the global optimum. Stochastic methods offer only a probabilistic (asymptotic) guarantee: their convergence proofs usually declare that the global optimum will be identified in infinite time with probability one. Moreover, stochastic methods adapt better to black-box formulations and extremely ill-behaved functions, whereas deterministic methods are usually based on at least some theoretical assumptions such as Lipschitz continuity and heavily depends on the problem at hand. A direct comparison between these two approaches may be found in [16], where the authors conclude that the stochastic approach is to be preferred. In addition deterministic methods suffer from the problem of dimensionality. For example, the complexity of interval global optimization [17] rises exponentially with the problem's dimension.

The problem we are interested in, may be expressed as:

$$\text{Find all } x_i^* \in S \subset R^n \text{ that satisfy :}$$
$$x_i^* = \arg\min_{x \in S_i} f(x), \quad S_i = S \cap \{x, \|x - x_i^*\| < \epsilon\}, \tag{1}$$

S is considered to be a bounded domain of finite measure and ϵ a positive infinitesimally small number. We are adopting the stochastic class of methods. One of the most widely used stochastic algorithms is the so called *Multistart* [18]. Its popularity stems from its simplicity and inherent parallelization [19–22]. Many stochastic methods have been developed around it

* Corresponding author.
E-mail address: voglis@cs.uoi.gr (C. Voglis).

0096-3003/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2009.03.012

starting from the classic papers of [18,23–25] were the popular Single Linkage Clustering, *Density Clustering* and *Multi-Level Single Linkage* procedures were introduced. Törn and Viitanen in [26] presented a *Topographical Clustering* algorithm which was extended by Ali and Storey in [27] to the well known *Topographical Multi-Level Single Linkage algorithm. More recently Hart in his PhD dissertation [28] proposes an adaptive method based on clustering and local searches, Locatelli [29] introduces the family of Random Linkage* algorithms and Schoen [30] and Locatelli [31] give an analysis *Two-phase methods*. More recently, Liang et. al. [32] introduce a function's landscape approximation, Bolton et. al. [33] provide a parallel framework based on clustering, while Tsoulos and Lagaris [34] proposed the so called *typical distance* clustering. Also related software may be found in [35].

In *Multistart* a point is sampled uniformly from the feasible region, and subsequently a local search is started from it. The weakness of this algorithm is that the same local minima may be found over and over again, wasting so computational resources. For this reason clustering methods have been developed that attempt to avoid repetitive discovery of the same minima [23–25,34,20].

The *Multistart* algorithm is presented below:

Multistart Algorithm
Initialize: Set k=1
Sample $x \in S$
$y_k = \mathscr{L}(x)$
Termination Control: If a stopping rule applies, STOP.
Sample: Sample $x \in S$
Main step: $y = \mathscr{L}(x)$
If $(y \notin \{y_i, i = 1, 2, \ldots, k\})$ Then
$k = k + 1$
$y_k = y$
Endif
Iterate: Go back to the Termination Control step.

The "**region of attraction**" of a local minimum associated with a local search procedure $\mathscr{L}$ is defined as:

$$A_i \equiv \left\{x \in S, \mathscr{L}(x) = x_i^*\right\}, \tag{2}$$

where $\mathscr{L}(x)$ is the minimizer returned when the local search procedure $\mathscr{L}$ is started at point x. If S contains a total of w local minima, from the definition above folows:

$$\cup_{i=1}^{w} A_i = S. \tag{3}$$

Let $m(A)$ indicate the Lebesgue measure of $A \subseteq R^n$. If we assume a deterministic search $\mathscr{L}$, then the regions of attraction do not overlap, i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$, and from Eq. (3) one obtains:

$$m(S) = \sum_{i=1}^{w} m(A_i). \tag{4}$$

If a point in S is sampled from a uniform distribution, the a priori probability p_i that it is contained in A_i is given by $p_i = \frac{m(A_i)}{m(S)}$. If K points are sampled from S, the a priori probability that at least one point is contained in A_i is given by:

$$1 - \left(1 - \frac{m(A_i)}{m(S)}\right)^K = 1 - (1 - p_i)^K. \tag{5}$$

From the above we infer that for large enough K, this probability tends to one, i.e. it becomes "asymptotically certain" that at least one sampled point will be found to belong to A_i. This holds $\forall A_i$, with $m(A_i) \neq 0$.

In this article we first define the "Ideal Multistart", a variation of Multistart in which every local minimum is found only once. This ideal version assumes that the region of attraction of a minimizer is determined as soon as the minimizer is located. Since this is a false hypothesis this version is of no practical value. It offers however a framework and a goal to work towards.

In section (2), we lay-out the new ideas involved and we present the corresponding algorithm, while in section (3), we give a description of the numerical experiments that were performed along with the results. Finally in section (5), our conclusions are summarized and we give a recommendation for future research.

2. Description of the method

"Ideal Multistart" starts by sampling a point from S and applying a local search leading to the first minimum y_1, with region of attraction A_1. Sampling points from S is continued until a point is found that does not belong to A_1. Once such a point is encountered, a local search is performed that leads to the second minimum y_2, having a region of attraction A_2. The next

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

sample point from which a local search will start, is a point that belongs neither to A_1 nor to A_2, i.e. it does not belong to their union ($A_1 \bigcup A_2$). This procedure goes on, until a stopping rule instructs termination. The detailed algorithm is laid out in the following paragraph.

2.1. Ideal Multistart

Ideal Multistart Algorithm
Initialize: Set k=1
Sample $x \in S$
$y_k = \mathscr{L}(x)$
Termination Control: If a stopping rule applies, STOP.
Sample: Sample $x \in S$
Main step: If $(x \notin \cup_{i=1}^{k} A_i)$ Then
$y = \mathscr{L}(x)$
$k = k + 1$
$y_k = y$
Endif
Iterate: Go back to the Termination Control step.

This algorithm invokes the local search procedure only w times, w being the number of existing minima of $f(.)$ in S. The main step is deterministic and requires the regions of attraction A_i of the already located minima to be known, which is not the case in practice. Hence we apply a stochastic modification to the main step, by allowing the local search to be performed with a probability, namely:

Main step (**Stochastic**):
Calculate the probability p, that $x \notin \cup_{i=1}^{k} A_i$
Draw a random number $\xi \in (0,1)$ from a uniform distribution
If $(\xi < p)$ Then
$y = \mathscr{L}(x)$
If $(y \notin \{y_i, i = 1, 2, \ldots, k\})$ Then
$k = k + 1$
$y_k = y$
Endif
Endif

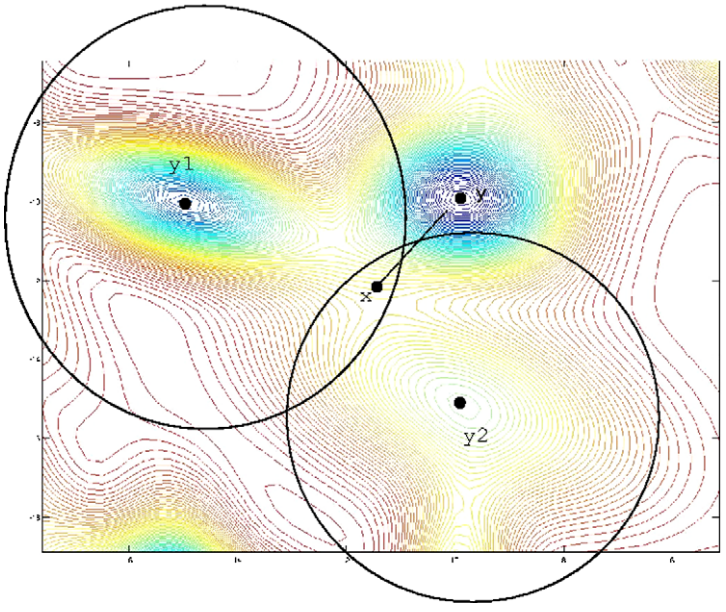


Fig. 1. A point x that would lead to a new minimum y, is inside the overlap region of the spheres around two recovered minima y_1 and y_2.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

This step requires the probability that a point does not belong to the region of attraction of any of the minima collected so far. This requirement is easier to fulfill, since even with a low accuracy estimate for the probability, the algorithm will succeed. Notice that an overestimated probability ($p \to 1$) will transform the algorithm into the usual *Multistart*. On the other hand underestimation ($p \to 0$) is not of considerable cost, since no local search is performed. Performance however will be optimized if reasonably accurate estimates for the probability can be calculated. Several ways may be designed to accomplish this goal. We suggest one in the following paragraph.

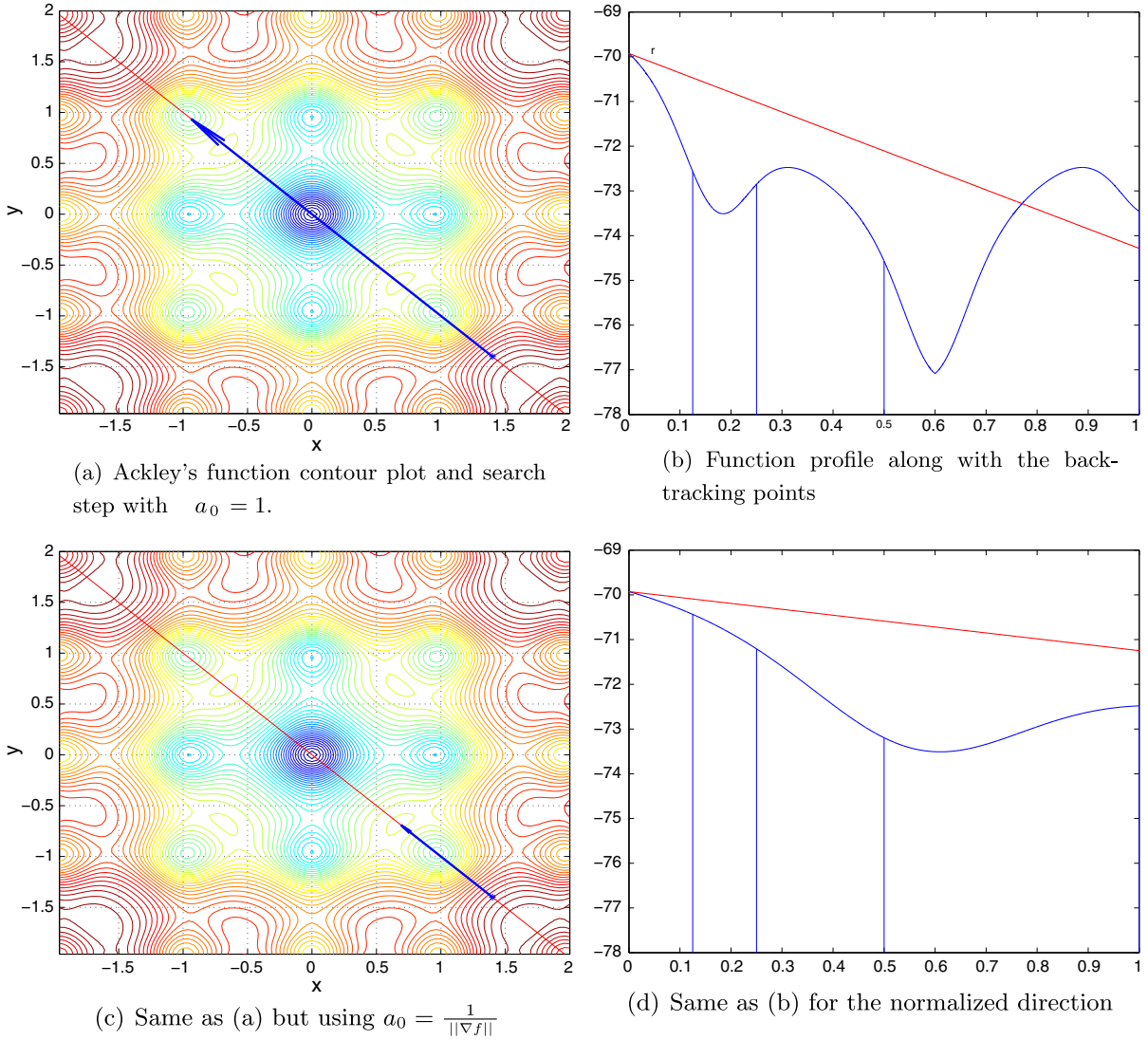
2.2. Estimating the local search probability

The required probability may depend on several factors, such as the distance from existing minimizers, the direction of the gradient, the number of times each minimizer has been discovered, etc. We consider how each factor influences the probability and combine them together to get the required estimate.

Let us define the *maximum attractive radius (MAR)* as:

$$R_i = \max_j \left\{ \left\| x_j^{(i)} - y_i \right\| \right\}, \tag{6}$$

where $x_j^{(i)}$ are the sampled points which led the subsequent local search to the i^{th} minimizer y_i.



(a) Ackley's function contour plot and search step with $a_0 = 1$.

(b) Function profile along with the backtracking points

(c) Same as (a) but using $a_0 = \frac{1}{||\nabla f||}$

(d) Same as (b) for the normalized direction

Fig. 2. Illustration of the modified line search.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

Given a sampled point x, let y be anyone of the recovered minimizers, with *MAR* denoted by R. If $\|y-x\| < R$, then x is likely to be inside the region of attraction of y. If however $\nabla f(x)^T(y-x) \geqslant 0$, i.e. the direction from x to y is ascent, then x is likely to be outside y's region of attraction. Letting $z \equiv \|y-x\|/R$, then an estimate of the probability that $x \notin A(y)$ may be given by:

$$p(x \notin A(y)) = \begin{cases} 1, & \text{if } z > 1 \text{ or } \nabla f(x)^T(y-x) \geqslant 0, \\ \phi(z,l) * \left[1 + \frac{(y-x)^T \nabla f(x)}{\|y-x\| \nabla f(x)|}\right], & \text{otherwise}, \end{cases} \tag{7}$$

l is the number of times y has been recovered so far, while $\phi(z,l)$ is a model with the following properties:

$$\begin{aligned} &\lim_{z\to 0} \phi(z,l) \to 0, \\ &\lim_{z\to 1} \phi(z,l) \to 1, \\ &\lim_{l\to\infty} \phi(z,l) \to 0, \\ &0 < \phi(z,l) < 1. \end{aligned} \tag{8}$$

Notice that the factor inside the square brackets in Eq. (7), varies from zero to one, as the gradient from anti-parallel becomes perpendicular to $y-x$.

The probability that $x \notin \cup_{i=1}^{k} A_i$ is given by the product $\prod_{i=1}^{k} p(x \notin A_i)$ and may now be approximated by the probability that $x \notin A_n, A_n$ being the region of attraction of the nearest to x discovered minimizer y_n. The rationale for this approximation is that if $x \notin B(y_i, R_i)\ \forall i \neq n$, where $B(y,R)$ is a sphere of radius R centered at y, then the above approximation is exact since all other probabilities as following from Eq. (7) equal 1. If on the other hand x is inside the intersection of two or more overlapping spheres, the product of small terms may result to too small a probability for a point that could lead to a new minimum (see in Fig. 1, an example). The spheres are expected to overlap, due to the manner their radii are chosen by Eq. (6). Hence the approximation is prudent, and essentially in most cases does not overestimate the local search probability. One may employ alternative approximations, by considering for example the first two (or more) nearest minimizers. This is an issue that needs further consideration and is outside the scope of the present article.

2.3. Local search properties

The probability model is based on distances from the discovered minima. It is implicitly assumed that the closer to a minimum a point is, the greater the probability that falls inside its region of attraction. This implies that the regions of attraction are contiguous and surround the minima. This is not true for all local search procedures and hence this assumption influences the local search choice. For example widely used methods such as Newton or quasi Newton, employing either a line

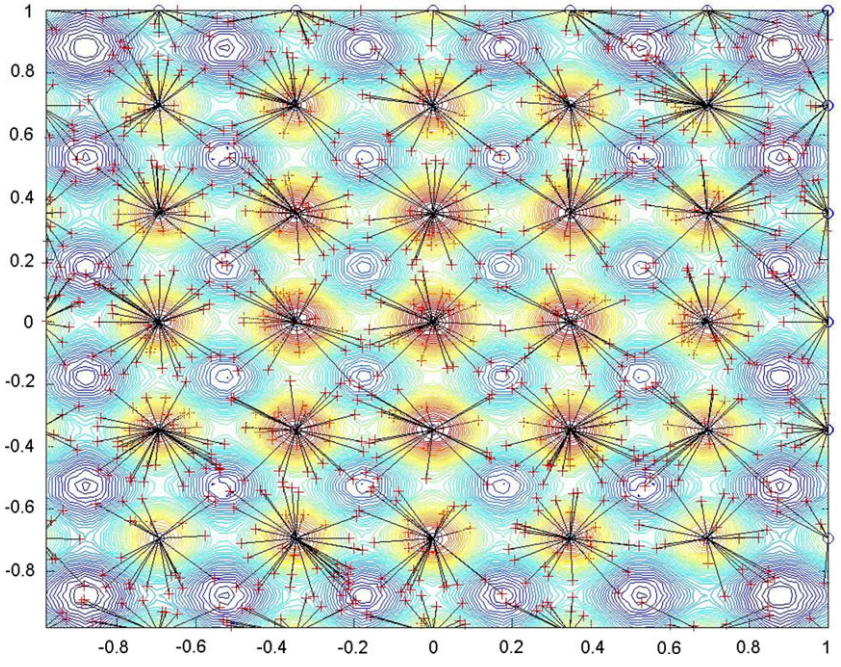


Fig. 3. A suitable local search, with contiguous regions of attraction.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

search or a trust region strategy, create disjoint regions of attraction. Hence these methods have to be modified so that their regions of attraction are contiguous, resembling those of a descent method with an infinitesimal step. In Fig. 3 we connect start-points (marked by +) to the minimum they arrive via a local search. This is a desirable local search since its regions of attraction are contiguous. Start points are attracted towards the close by minima.

In this work we apply the BFGS method with a modified line search. This modification creates contiguous regions of attraction ensuring a strictly descent path [23].

We present the associated algorithm below:

Modified Local Search Algorithm
Input:
$k = 0, B_k = I, \epsilon > 0$
Step 1 (Calculate descent direction):
$p_k = -B_k^{-1}\nabla f(x_k)$
If $\|\nabla f(x)\| > \epsilon$ **Then**
$p_k = -\frac{1}{\|\nabla f(x_k)\|} B_k^{-1}\nabla f(x_k)$
End if
Step 2 (Line search):
$min_a(f(x_k + \alpha p_k))$, **yielding** a_k
Step 3 (Next iterate):
$x_{k+1} = x_k + \alpha_k p_k$
Step 4 (Update approximation):
$\gamma_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
$\delta_k = x_{k+1} - x_k$
$B_{k+1} = bfgs\_update(B_k, \gamma_k, \delta_k)$
Step 5 (Termination Control):
If termination conditions are met stop, Else set
$k \leftarrow k + 1$ **and repeat from Step 1.**

To illustrate the behavior of this normalization at Step 1 of the line search we provide Fig. 2a–d. The single minimum appearing in Fig. 2d is the first minimum in Fig. 2b. Note that in Fig. 2c the line search ends up to the nearest minimum while that of Fig. 2a in a different minimum further apart.

In Fig. 4 we connect start-points (marked by +) to the minimum the arrive via a different local search. This illustrates an undesirable local search since its regions of attraction are disjoint. Start points are attracted towards distant minima.

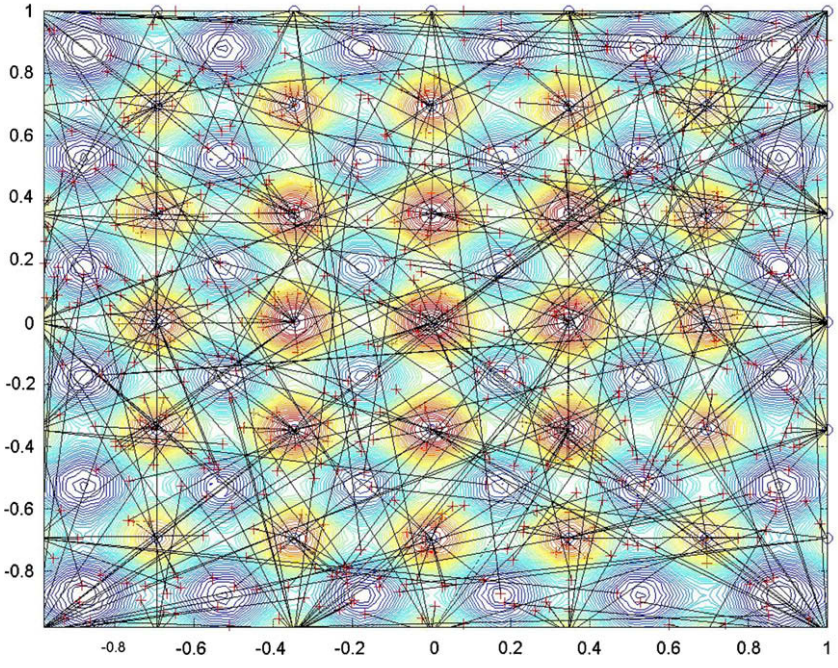


Fig. 4. An improper local search, with disjoint regions of attraction.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

2.4. Asymptotic guaranty

The probability that minimizer y is found with one trial is given by:

$$p_y^{(i)} = \int_{x \in A(y)} p_{LS}^{(i)}(x) \frac{dx}{|S|}, \tag{9}$$

where $1/|S|$ is the pdf of the uniform distribution and $p_{LS}^{(i)}(x)$ is the local search probability at x. The superscript i denotes the state of the process, i.e. the number of minima discovered so far, the number of times each minimizer is found, the *MAR's*, etc. The probability that after k trials y is not found is then given by:

$$\pi_y^{(k)} = \prod_{i=1}^{k} (1 - p_y^{(i)}) \leqslant \left(1 - \min_i \{p_y^{(i)}\}\right)^k. \tag{10}$$

From the definition of $p_y^{(i)}$ in Eq. (9), we have:

$$p_y^{(i)} = \int_{x \in A_1(y)} p_{LS}^{(i)}(x) \frac{dx}{|S|} + \int_{x \in A_2(y)} p_{LS}^{(i)}(x) \frac{dx}{|S|}, \tag{11}$$

where

$$A_1(y) = \{x \in A(y); (y_c - x)^T \nabla f(x) \leqslant 0\},$$
$$A_2(y) = \{x \in A(y); (y_c - x)^T \nabla f(x) > 0\}$$

and $y_c = y_c(x)$, is the closest to x discovered minimizer.

If y is not found yet (and hence $y_c \neq y$), then $A_2(y) \neq \emptyset$ and hence $|A_2(y)| \neq 0$. Note that

$$\forall x \in A_2(y), \quad p_{LS}^{(i)}(x) = 1$$

and hence from Eq. (11)

$$p_y^{(i)} \geqslant \frac{|A_2(y)|}{|S|} > 0, \quad \forall i = 1, 2, \ldots, k.$$

At the limit as $k \to \infty$ we deduce from above and Eq. (10) that $\pi_y^{(k)} \to 0$, i.e. asymptotically all minimizers will be found.

2.5. A model for $\phi(z, l)$

Many models may be constructed with the desired properties described in (8). We propose one that is simple to visualize and easy to implement.

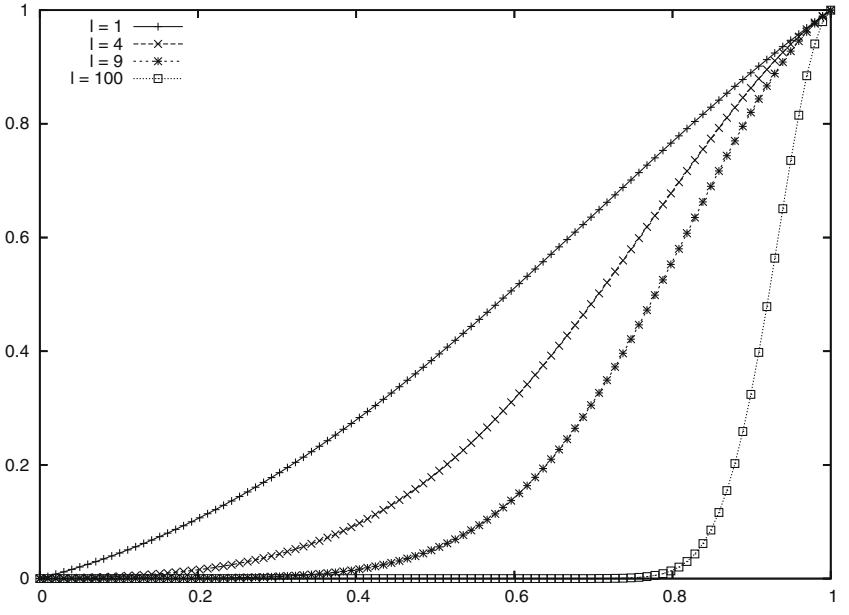


Fig. 5. Model plots for several l values.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

$$\phi(z,l) = ze^{-l^2(z-1)^2}, \quad \forall z \in (0,1). \tag{12}$$

A graphical representation is depicted in (Fig. 5).

2.6. The ADAPT Algorithm

The proposed algorithm, in summary, is presented below:

ADAPT Algorithm
Input:
The input function $f : R^n \to R$
The search domain $S \subseteq R^n$
A local search procedure $\mathscr{L}(x)$ having the properties described in Section 2.3.

Initialize: **Set k=1**
Sample $x \in S$
$y_k = \mathscr{L}(x)$
$r_k = \|x - y_k\|, n_k = 1$
Termination Control: **If a stopping rule applies, STOP.**
Sample: **Sample** $x \in S$
Main step: $\mathbf{i} = \underset{j=1,\ldots,k}{argmin} \|x - y_j\|$
$d = \|x - y_i\|$
If $(d < r_i)$ **Then**
If $(\nabla f(x)^T (y_i - x) < 0)$ **Then**
$z = \frac{\|y_i - x\|}{r_i}$
$p = \phi(z, n_i)\left[1 + \frac{(y_i - x)^T \nabla f(x)}{\|(y_i - x)^T \nabla f(x)\|}\right]$
Else
$p = 1.0$
Endif
Else
$p = 1.0$
Endif
Let ξ **be a uniform random in [0, 1]**
If $(\xi < p)$ **Then**
$y = \mathscr{L}(x)$
If (y **is new minimum) Then**
$k = k + 1, r_k = \|x - y_k\|, n_k = 1$
Else {*We discovered the l-th local minimum*}
$r_l = \max(r_l, \|x - y_l\|), n_l = n_l + 1$
Endif
Else {*Assuming that x belongs in the region of attraction of the i-th minimum*}
$r_i = \max(r_i, \|x - y_i\|), n_i = n_i + 1$
Endif
Iterate: **Go back to the Termination Control step.**

3. Experiments and comparison

The method has been tested on a number of test problems that are listed in Appendix A. These test functions have been used in the past by many authors and hence they constitute a convenient platform for comparison. We count for every problem the number of local searches, the number of function and gradient evaluations and we report averages on thirty experiments performed with different random number sequences. We also count the number of minima found. All experiments used the "*Double-Box*" stopping rule [36], with the suggested compromise factor (0.5). The local search used by ADAPT is a modification of BFGS so that the resulting regions of attraction have the properties described in Section 2.3. A comparison is made with the standard "*Multistart*" with the "*Topological Multilevel Single Linkage*" (TML) method [35] and with *MinFinder* [34].

All of the above methods use as a local minimizer subroutine `TOLMIN` due to Powell [54]. We coded Multistart, while the codes for *TML* and *MinFinder* were obtained from the corresponding authors and were run with the default parameters. Observing the results listed in Table 1 we note that the performance of the new method (*ADAPT*) is overall superior. *MinFinder* has similar performance on functions M0, Borne, Shubert(N = 5, 10) while it has an edge with functions having a periodicity in their contour plots like Holder, Levy No3, Rastrigin($N = 2$), and Shubert(N = 2).

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

Table 1
Method comparison using the Double-Box rule [36].

Function	TML				Multistart				MinFinder				Adapt			
	Min.	FC	GC	LS	Min.	FC	GC	LS	Min.	FC	GC	LS	Min.	FC	GC	LS
Ackley	121	10,259	14,457	1207	121	23,281	36,543	2054	121	7510	11,926	208	121	7340	4600	539
Bird	158.5	84,798	103,889	2507	141.5	212,196	150,529	3737	172.8	122,639	145,460	1832	171.7	56,008	55,296	1468
Bohachefsky	25	139,190	125,684	2369	25	241,501	187,175	2547	18.7	25,907	32,243	538	24.3	18,332	23,112	215
Giunta	196	104,812	16,606	719	196	45,688	67,311	1212	196	18,753	20,972	791	196	10,211	17,821	771
Grienwank	527.2	1,883,423	1,461,617	39,090	526.4	1,912,452	1,892,111	38,727	529	1,133,908	1,284,982	30,577	528.5	231,123	27,811	15,733
Guillin Hills	25	81,153	69,847	2451	24.9	87,411	76,563	2617	24.8	22,901	23,570	820	24.7	17,751	31,811	691
Holder	85	28,749	23,346	622	85	69,038	34,468	988	85	8289	8977	261	85	16,788	16,461	352
Langermann	257	129,521	124,360	3566	260	185,669	111,478	4169	270	503,470	500,675	19,123	270	80,578	80,386	2479
Levy No3	527	170,541	171,643	6999	527	494,578	277,868	8909	527	59,830	91,479	2320	527	146,574	179,502	5481
Levy No5	508	173,026	183,092	5011	508	365,258	175,718	6783	508	81,037	160,683	2733	508	84,152	83,462	2644
Liang	224.6	90,506	51,538	2464	233.1	180,419	79,899	3161	236	637,784	676,941	22,607	235.8	73,215	50,569	2340
Piccionni	43	58,042	42,536	1475	42.9	74,125	72,918	2090	43	33,333	36,238	1222	43	48,123	45,647	987
Rastrigin	49	11,340	14,812	741	49	22,233	17,063	1705	49	1730	2833	85	49	17,810	7481	136
Voglis	60.8	16,938	1888	944	60.5	35,408	2505	2304	61	21,684	23,126	694	61	16,932	1267	437
Schaffer	93.7	48,401	23,295	865	94.8	56,722	73,922	1811	94.5	22,370	24,682	876	94.7	18,922	16,779	702
Shubert	399.6	890,899	2594	2297	398.2	1,062,260	10,732	10,475	400	16,551	36,065	665	400	193,211	10,780	1439
M0	65.1	53,817	35,311	1654	64.5	85,266	87,221	2741	64	14,667	16,005	799	65.7	16,659	17,033	1023
M3	25.8	30,601	20,295	1507	25.8	47,188	33,872	2184	23.7	8914	11,285	790	25.6	8752	17,168	711
Borne	595.6	1,090,974	322,968	11,324	593.4	3,821,280	3,343,620	15,922	598.2	1,916,295	2,138,766	68,865	598.2	1,880,314	2,626,185	70,161
Rast($N = 5$)[a]	243	131,646	36,084	662	243	399,909	111,467	2011	243	59,298	64,349	1022	243	30,350	37,634	1214
Griew($N = 5$)[b]	160.1	1,859,878	1,619,266	27,717	159.6	2,154,055	2,023,821	32,101	170	2,074,573	2,193,769	32,710	169.8	1,833,628	2,052,681	34,911
Griew($N = 10$)[c]	11.1	86,815	78,843	1277	11.4	145,552	125,187	2141	10.4	85,454	85,209	1270	12.7	76,221	121,176	1782
Shub($N = 5$)[d]	32	12,221	15,622	508	32	17,881	19,822	811	32	6136	6520	158	32	7022	8112	205
Shub($N = 10$)[e]	1021.2	1,779,357	220,435	3977	1002.1	1,866,619	1,973,621	33,230	1024	563,927	565,624	10,302	1024	406,503	526,229	12,853

[a] 243 minima in $[-0.5, 0.5]^5$.
[b] 171 minima in $[-5, 5]^5$.
[c] 13 minima in $[-3, 3]^5$.
[d] 32 minima in $[-1, 1]^5$.
[e] 1024 minima in $[-1, 1]^10$.

Table 2
Adapt results using different stopping rules.

Function	Boender and Kan				Observables				Expected Minimizers			
	Min.	FC	GC	LS	Min.	FC	GC	LS	Min.	FC	GC	LS
Ackley	121	11,245	6539	759	121	4167	3630	367	121	4265	2891	326
Bird	169	84,260	84,659	2021	170	35,551	36,867	985	171	33,165	31,874	866
Bohachefsky	22	26,213	34,576	318	24	12,815	16,532	150	24	11,615	14,865	137
Giunta	193	14,077	27,661	1077	194	7156	11,693	521	195	7829	10,953	462
Grienwank	527	273,304	36,644	21,407	528	121,838	15,231	10,498	529	136,513	22,405	9179
Guillin Hills	25	23,790	45,749	969	24	12,235	19,661	469	23	10,713	18,995	415
Holder	85	22,165	23,301	504	85	11,022	11,525	240	81	10,014	10,368	216
Langermann	263	120,996	109,319	3397	268	54,378	52,878	1660	269	45,700	47,968	1456
Levy No3	521	215,190	204,894	7478	524	97,859	121,728	3661	524	90,826	108,808	3206
Levy No5	508	115,365	118,926	3622	508	56,203	53,068	1770	508	46,168	49,922	1552
Liang	234	99,512	60,608	3206	235	53,217	35,288	1568	236	45,557	29,356	1375
Piccioni	43	65,782	61,847	1367	42	31,060	32,534	666	42	29,770	27,678	587
Rastrigin	49	28,537	12,238	214	49	12,873	5345	97	49	11,901	4920	91
Voglis	59	26,614	1136	621	60	12,833	850	298	61	10,620	510	265
Schaffer	95	23,812	24,878	982	95	13,096	9704	476	95	12,128	9537	420
Shubert	400	270,985	17,194	1982	400	132,697	9041	966	400	115,116	5153	851
M0	66	26,782	23,763	1417	66	10,261	11,503	689	66	10,160	10,899	608
M5	19	9573	24,757	992	23	7286	10,728	480	24	4660	9226	425
Borne	598	2,685,857	3,312,841	95,370	597	1,152,416	1,678,605	46,790	598	1,033,720	1,576,821	40,893
Rast($N = 5$)	238	44,294	49,774	1676	240	19,100	27,478	816	240	14,214	24,037	720
Griew($N = 5$)	166	2,654,250	2,776,645	47,468	169	1,182,435	1,329,089	23,286	170	1,026,242	1,170,489	20,355
Griew($N = 10$)	11	101,772	164,941	2449	13	52,465	82,684	1194	13	42,192	73,400	1050
Shub($N = 5$)	29	9288	12,639	304	31	5379	5627	143	29	4573	5526	131
Shub($N = 10$)	1021	557,748	754,471	17,493	1023	279,352	372,875	8577	1024	254,876	310,294	7501

Lagaris and Tsoulos [36] report a comparison among five stopping rules. From their results it can be seen that Multistart favors the *Expected minimizers* [36] rule with the *Observables* [36] and *Double-box* criteria following closely. We conducted experiments using the criteria of Boender and Kan [37], the *Observables* and the *Expected minimizers*. Similar behavior is observed as it can be deduced by inspecting the results displayed in Table 2. Although the *Double-Box* is not the best performer, it is fairly easy to implement and has negligible computational overhead.

4. A parallel scheme

A sample Master–Slave parallel implementation is displayed below. The Master CPU creates candidate start points. The Slave CPUs perform local searches. Note, that since our method uses one point per iteration, each search is independent, enabling so maximum utilization of the Slave CPUs. On the other hand, most clustering methods use a collection of points, as for example in [23,24,34], that in turn create dependencies in the application of the local searches, a fact that makes the parallelization less profitable.

Definitions.

- M-list: A list that holds the minimizers (managed by the Master CPU);
- S-list: A list of possible starting points (managed by the Master CPU);
- L-list: There is one such list for every Slave CPU. Each contains the minimizers discovered by the corresponding CPU.

Master CPU:
1. Check if a stopping rule applies. If so terminate.
2. Take in account the updated minimizers list (M-list).
3. Create candidate start points and add them to the starting list (S-list) and assign to each one a zero flag.

Slave CPUs:
1. If no zero flag start-points exist in the S-list, wait.
2. Pick from the S-list a start–point with zero flag, change its flag to one, and apply a local search.
3. Add the minimizer to a temporary local minimizer list (L-list).

Updater CPU:
1. Pick a minimizer from the L-list and check if it is a new minimizer and remove it from the list.
2. If so, add it to the M-list.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

5. Conclusions and further work

The adaptive character of the method enables a reasonably accurate estimate of the probability that a point belongs to a region of attraction. This in turn, on one hand saves a large fraction of local search applications, and on the other hand prevents the systematic overlook of regions of attraction, reducing therefore the risk of loosing minima. The method is robust and efficient as has been deduced from the results of the computational experiments. Most of the stochastic global optimization approaches use a population of points to proceed and thus the population size is an additional parameter that affects the performance of the method. The present work in contrast, uses a single point per iteration without any adjustable parameters. This feature adds another (obvious) advantage in the case where the parallel implementation is of interest.

A parallel algorithm that would benefit from a cluster of tightly coupled processors or from a parallel shared memory system would be significant development. Such systems are nowadays widely available and offer the possibility of solving harder problems. Work in this direction is underway.

Other models for the probability, such as adaptively grown Gaussian mixtures may be considered and some early preliminary results are promising.

Appendix A. Test functions

A.1. Ackley's test function [40]

The number of existing minima in $[-5,5]^2$ is 121

$$f(x) = -\alpha e^{-b\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n} cos(cx_i)} - \alpha e^1.$$

A.2. Bird's test function [41]

This function has 173 minima in $[-50,50]^2$

$$f(x_1,x_2) = sin(x_1)\; e^{(1-cos(x_2))^2} + cos(x_2)\; e^{(1-sin(x_1))^2} + (x(1)-x(2))^2.$$

A.3. Bohachevsky's test function [42]

This function has 25 minima in $[-10,10]^2$

$$f(x_1,x_2) = x_1^2 + 2x_2^2 - 0.3\; cos(3\pi x_1) - 0.4\; cos(4\pi x_2) + 0.7.$$

A.4. Giunta's test function [44]

This test function has 196 minima inside $[-20,20]^2$

$$f(x_1,x_2) = 0.6 + \sin y_1 + \sin^2 y_1 + \frac{1}{50}\sin 4y_1 + \sin y_2 + \sin^2 y_2 + \frac{1}{50}\sin 4y_2,$$

where $y_1 = \frac{16}{15}x_1 - 1$ and $y_2 = \frac{16}{15}x_2 - 1$. where $y_1 = \frac{16}{15}x_1 - 1$ and $y_2 = \frac{16}{15}x_2 - 1$.

A.5. Griewank's test function [45]

This function has 529 minima inside $[-100,100]^2$

$$f(x) = \frac{1}{200}\sum_{i=0}^{n} x_i^2 - \prod_{i=1}^{n} \cos\frac{x_i}{\sqrt{i}} + 1.$$

A.6. Guillin Hills's test function [34]

This test function possesses 25 minima inside $[0,1]^2$

$$f(x) = 3 + \sum_{i=1}^{n} \frac{c_i(x_i+9)}{x_i+10}\sin\left(\frac{\pi}{1-x_i+\frac{1}{2k}}\right),$$

where $c_i = 2,\ i = 1,\ldots,n$ and $k = 5$.

A.7. Holder test function [41]

This function has 85 minima inside $[-20,20]^2$

$$f(x_1,x_2) = -\cos x_1 \cos x_2 e^{1-\frac{\sqrt{x_1^2+x_2^2}}{\pi}}.$$

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

A.8. Langermanns's test function [46]

This test function has 270 minima inside $[0,7]^2$

$$f(x_i) = \sum_{k=0}^{5} c_k e^{\sigma_k} \cos \lambda_k.$$

In current implementation $a = (3,5,2,1,7)^T, c = (1,2,5,2,3)^T$ where $\sigma_k = \sum_{i=1}^{n} -\frac{(x_i-a_k)^2}{\pi}$ and $\lambda_k = \sum_{i=1}^{n} \pi(x_i - a_k)^2$.

A.9. Levy's 3rd test function [48]

This test function has 527 minima inside $[-10,10]^2$

$$f(x_1,x_2) = \sum_{k=1}^{5} k\cos((k-1)x_1 + k) \sum_{k=1}^{5} k\cos((k+1)x_2 + k).$$

A.10. Levy's 5th test function [48]

This test function has 508 minima inside $[-10,10]^2$

$$f(x_1,x_2) = f_{Levy3}(x_1,x_2) + (x_1 + 1.42513)^2 + (x_2 + 0.80032)^2.$$

A.11. Liang's test function [49]

This test function has 236 local minima inside $[1,4]^2$

$$f(x_1,x_2) = -(x_1\sin(20x_2) + x_2\sin(20x_1))^2 \cosh(\sin(10x_1)x_1) - (x_1\cos(20x_2) - x_2\sin(10x_1))^2 \cosh(\cos(10x_2)x_2).$$

A.12. Piccioni's test function [50]

This test function has 28 minima inside $[-5,5]^2$

$$f(x) = -10\sin(\pi x_1)^2 - \sum_{i=1}^{n-1}(x_i - 1)^2(1 + 10\sin(\pi x_{i+1})) - (x_n - 1)^2.$$

A.13. Rastrigin's test function [51]

This test function has 49 minima inside $[-1,1]^2$

$$f(x) = 10n + \sum_{i=1}^{n}\left(x_i^2 - 10\cos(2\pi x_i)\right).$$

A.14. Voglis's test function

This test function has 61 minima inside $[-25,25]^2$

$$f(x) = \alpha_0\left(\frac{1}{2}x^T Q_0 x + x^T d_0\right) + \sum_{i=1}^{80} \alpha_k e^{-\frac{1}{2}x^T Q_k x + x^T d_k}.$$

Function dimension $n = 2$, Q_j specific positive definite 2×2 matrices, d_j 2-dimensional vectors and α_j appropriate scaling constants.

A.15. Schaffer's test function [41]

This test function has 95 minima inside $[-3,3]^2$

$$f(x_1,x_2) = 0.5 + \frac{\sin(x_1^2 + x_2^2)^2 - 0.5}{\left(1 + 0.001(x_1^2 + x_2^2)\right)^2} + 0.1\sin(10x_1) + 0.1\sin(10x_2).$$

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

A.16. Shubert's test function [52]

This test function has 400 minima inside $[-10, 10]^2$

$$f(x) = -\sum_{i=1}^{n}\sum_{j=1}^{5} j\sin((j+1)x_i + j).$$

A.17. M0 test function [52]

This test function has 66 minima inside $[-5, 1]^2$

$$f(x) = \sin\left(2.2\pi x_1 + \frac{\pi}{2}\right)\frac{2-x_2}{2}\frac{3-x_1}{2} + \sin\left(\frac{\pi}{2}x_2^2 + \frac{\pi}{2}\right)\frac{2-x_2}{2}\frac{3-x_1}{2}.$$

A.18. M3 test function [52]

This test function has 26 minima inside $[-2, 2]^2$

$$f(x) = -(x_2^2 - 4.5x_2^2)x_1x_2 - 4.7\cos(3x_1 - x_2^2(2+x_1))\sin(2.5\pi * x_1) + (0.3 * x1)^2.$$

A.19. Siam Problem 4 Function [53]

This test function has 600 minima inside $[-1, 1]^2$

$$f(x) = \exp(\sin(x_1)) + \sin(60\exp(x_2)) + \sin(70\sin(x_1)) + \sin(\sin(80x_2)) - \sin(10(x_1+x_2)) + \frac{x_1^2 + x_2^2}{4}.$$

References

[1] M. Pardalos Panos, H. Romeijn Edwin, Hoang Tuy, Recent developments and trends in global optimization, Journal of Computational and Applied Mathematics 124 (2000) 209–228.
[2] C.D. Maranas, C.A. Floudas, Global minimum potential energy conformations of small molecules, Journal of Global Optimization 4 (1994) 135–170.
[3] J.D. Wales, H.A. Scheraga, Global optimization of clusters, crystals, and biomolecules, Science 285 (1999) 1368–1372.
[4] M. Saunders, Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum, Journal of the American Chemical Society 109 (1987) 3150–3152.
[5] W.T. Ziemba, R.G. Vickson, Stochastic Optimization Models in Finance, World Scientific, 2006.
[6] R.T. Rockafella, S. Uryasev, Optimization of conditional value-at-risk, Journal of Risk 2 (2000) 21–41.
[7] D. Bertsimas, C. Darnell, R. Soucy, Portfolio construction through mixed integer programming, Sloan Working Papers, Sloan School of Management, Massachusetts Institute of Technology, 1998.
[8] D.L. Iglehart, S. Voessner, Optimization of a trading system using global search techniques and local optimization, Journal of Computational Intelligence in Finance (1998).
[9] C.A. Floudas, P.M. Pardalos, A Collection of Test Problems for Constrained Global Optimization Algorithms, Springer-Verlag New York, Inc., 1990.
[10] D.M. Himmelblau, Applied Nonlinear Programming, McGraw-Hill, 1972.
[11] C.A. Floudas, P.M. Pardalos, Optimization in Computational Chemistry and Molecular Biology: Local and Global Approaches, Kluwer Academic Publishers, 2000.
[12] H.D. Waterbeemd, D.A. Smith, K. Beaumont, D.K. Walker, Property-based design: optimization of drug absorption and pharmacokinetics, Journal of Medical Chemistry 44 (2001) 1313–1333.
[13] C.A. Floudas, Deterministic Global Optimization: Theory, Methods and Applications, Springer, 1999.
[14] R. Horst, H. Tuy, Global Optimization: Deterministic Approaches, Springer, 1993.
[15] C.G.E. Boender, H. Romeijn Edwin, Stochastic methods, in: R. Horst, P.M. Pardalos (Eds.), Handbook of Global Optimization, Kluwer, Dordrecht, 1995, pp. 829–871.
[16] Leo. Liberti, S. Kucherenko, Comparison of deterministic and stochastic approaches to global optimization, International Transactions In Operational Research 12 (2005) 263–285.
[17] E.R. Hansen, G.W. Walster, Global Optimization Using Interval Analysis, CRC Press, 2004.
[18] C.G.E. Boender, A.H.G. Rinnooy Kan, G.T. Timmer, L. Stougie, A stochastic method for global optimization, Mathematical Programming 22 (1982) 125–140.
[19] Z. Ugray, L. Lasdon, John Plummer, J. Kelly, R. Marti, Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization, McCombs Research Paper Series No. IROM-07-06, 2006, SSRN: <http://ssrn.com/abstract=886559>.
[20] W. Tu, R.W. Mayne, An approach to multi-start clustering for global optimization with non-linear constraints, International journal for numerical methods in engineering 53 (2002) 2253–2269.
[21] J.S. Maltz, E. Polak, T.F. Budinger, Multistart optimisation algorithm for joint spatial and kinetic parameter estimation in dynamic ECT, in: Nuclear Science Symposium, 1998, Conference Record, 1998 IEEE, vol. 3, pp. 1567–1573.
[22] L.C.W. Dixon, M. Jha, Parallel algorithms for global optimization, Journal of Optimization Theory and Applications 79 (1993) 385–395.
[23] A.H.G. Rinnooy Kan, G.T. Timmer, Stochastic global optimization methods. Part I: clustering methods, Mathematical Programming 39 (1987) 27–56.
[24] A.H.G. Rinnooy Kan, G.T. Timmer, Stochastic global optimization methods. Part II: multi level methods, Mathematical Programming 39 (1987) 57–78.
[25] A.A. Törn, A search clustering approach to global optimization, in: L.C.W. Dixon, G.P. Szegö (Eds.), Towards Global Optimization, vol. 2, North-Holland, Amsterdam, 1978.
[26] A. Törn, S. Viitanen, Topographical global optimization using pre-sampled points, Journal of Global Optimization 5 (1994) 267–276.
[27] M.M. Ali, C. Storey, Topographical multilevel single linkage, Journal of Global Optimization 5 (1994) 349–358.
[28] Hart, W.E. (1994), Adaptive Global Optimization with Local Search, Ph.D Dissertation, University of California, San Diego.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012

[29] M. Locatteli, Random linkage: a family of acceptance/rejection algorithms for global optimisation, Mathematical Programming 85 2 (1998) 379–396.
[30] F. Schoen, Two-phase methods for global optimization, in: Handbook of Global Optimization, Kluwer Academic Publishers, 2002, pp. 151–178.
[31] M. Locatteli, On the multilevel structure of global optimization problems, Computational Optimization and Applications 30 1 (2005) 5–22.
[32] K.-H. Liang, X. Yao, C. Newton, Combining landscape approximation and local search in global optimization, in: Proc. of the 1999 Congress on Evolutionary Computation, 1999, pp. 1514–1520.
[33] H.P.J. Bolton, J.F. Schutte, A.A. Groenwold, Multiple Parallel Local Searches in Global Optimization, LNCS 1908, 2000, pp. 88–95.
[34] I.G. Tsoulos, I.E. Lagaris, MinFinder: locating all the local minima of a functions, Computer Physics Communications 174 (2006) 166–179.
[35] F.V. Theos, I.E. Lagaris, D.G. Papageorgiou, PANMIN: sequential and parallel global optimization procedures with a variety of options for the local search strategy, Comput. Phys. Commun. 159 (2004) 63–69.
[36] I.E. Lagaris, I. Tsoulos, Stopping rules for box-constrained stochastic global optimization, Applied Mathematics and Computation 197 (2008) 622–632.
[37] C.G.E. Boender, A.H.G. Kan Rinnooy, Bayesian stopping rules for multistart global optimization methods, Mathematical Programming 37 (1987) 59–80.
[40] D.H. Ackley, A Connectionist Machine for Genetic Hillclimbing, Kluwer Academic Publishers, 1987.
[41] S.K. Mishra, Some new test functions for global optimization and performance of repulsive particle swarm method, Munich Personal RePEc Archive, Paper No. 2718, 2006.
[42] M.E. Bohachevsky, M.E. Johnson, M.L. Stein, Generalized simulated annealing for function optimization, Technometrics 28 (1986) 209–217.
[44] A.A. Giunta, Aircraft Multidisciplinary Design Optimization using Design of Experiments Theory and Response Surface Modeling Methods, MAD Center Report 97-05-01, Virginia Polytechnic Institute & State Univ. Blacksburg, 1997.
[45] A.O. Griewank, Generalized descent for global optimization, Journal of Optimization Theory and its Application 34 (1981) 11–39.
[46] GEATbx, GEATbx: The Genetic and Evolutionary Algorithm Toolbox for Matlab, 2006. <http://www.geatbx.com/index.html>.
[48] A.V. Levy, A. Montalvo, The tunneling algorithm for the global minimization of functions, SIAM Journal on Scientific and Statistical Computing 6 (1) (1985) 15–29.
[49] J.J. Liang, P.N. Suganthan, Dynamic multi-swarm particle swarm optimizer, International Swarm Intelligence Symposium IEEE (2005) 124–129.
[50] S. Lucidi, M. Piccioni, Random tunneling by means of acceptance–rejection sampling for global optimization, Journal of Optimization Theory and Applications 62 (2) (1989) 255–277.
[51] L.A. Rastrigin, Statistical Search Methods, Nauka, Moscow, 1968.
[52] B.O. Shubert, A sequential method seeking the global maximum of a function, SIAM J. Numer. Anal. 9 (1972) 379–388.
[53] <http://www.siam.org/siamnews/01-02/challenge.pdf>.
[54] M.J.D. Powell, TOLMIN: A Fortran package for linearly constrained C optimization calculations, DAMTP 1989/NA2, Dept. of Applied Mathematics and Theoretical Physics, Univ. of Cambridge, 1989.

Please cite this article in press as: C. Voglis, I.E. Lagaris, Towards "Ideal Multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Appl. Math. Comput. (2009), doi:10.1016/j.amc.2009.03.012