RECENT ADVANCES AND TRENDS IN GLOBAL OPTIMIZATION: DETERMINISTIC AND STOCHASTIC METHODS

Panos M. Pardalos
Department of Industrial and Systems Engineering
Department of Biomedical Engineering
and McKnight Brain Institute,
303 Weil Hall, University of Florida,
Gainesville, FL 32611, USA,
pardalos@ufl.edu

Fabio Schoen
Dipartimento di Sistemi e Informatica
Università degli Studi di Firenze
via di S.Marta 3, 50139 FIRENZE, ITALY
schoen@ing.unifi.it

Abstract

Global optimization has been expanding in all directions at an astonishing rate during the last few decades. Many new theoretical, algorithmic, and computational contributions of global optimization have been used to solve a wide spectrum of difficult problems in science and engineering. In particular, global optimization has recently emerged as a successful and versatile tool in many aspects of product and process design problems. The first part of the paper covers material regarding deterministic global optimization. We are going to focus on new algorithmic developments and some applications. In the second part of the paper we will outline some of the most basic stochastic techniques for global optimization and will present some elementary yet powerful approaches which might prove very useful for applications in the fields of process design, of innovative material research and of problems of biological interest; these problems share an enormous complexity which in most cases can be attacked only through heuristic methods.

keywords

Global optimization, nonconvex optimization, deterministic and stochastic methods.

1 Introduction

Global optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time one of the most striking trends in global optimization is the constantly increasing interdisciplinary nature of the field.

As early as 1978, in one of the first books on global optimization (Dixon and Szegö, 1978) a clear distinction was made between classical optimization and global optimization:
Many important practical problems can be posed as mathematical programming problems. This has been internationally appreciated since 1944 and has lead to major research activities in many countries, in all of which the aim has been to write efficient computer programs to solve subclasses of this problem. An important subclass that has proved very difficult to solve occurs in many practical engineering applications. Let us consider the design of a system that has to meet certain design criterion. The system will include features that may be varied by the designer within certain limits. The values given to these features will be the optimization variables of the problem. Frequently when the system performance is expressed as a mathematical function of the optimization variables, this function, which will sometimes be called the objective function, is not convex and possesses more than one local minimum. The problem of writing computer algorithms that distinguish between these local minima and locate the best local minimum is known as the global optimization problem, and is the subject of this volume.

Today the term **global optimization** is routinely used to denote a design process with increased speed and quality of solution. Many deterministic, stochastic, and heuristic approaches have been developed in the last decades to solve global optimization problems. In this brief overview we will only discuss some of the general important developments in deterministic and stochastic global optimization techniques. Recent developments can be found in the "Journal of Global Optimization," and the "Nonconvex Optimization and its Applications" book series published by Kluwer Academic Publishers (JOGO; NOIA).

In the following section we will give a short overview of some recent advances in the theory and in the development of numerical methods for nonconvex problems. In section 2 we will introduce some specific classes of deterministic global optimization problems; it will be shown that several practical problems can be formulated as deterministic global optimization problems with some special structure which, if suitably exploited, can lead to quite efficient algorithms with provable optimality certificates. Later, in section 3 we will describe a quite general setting in which many large scale problems with no known structure can be attacked by means of relatively simple heuristics which are based on a suitable coupling of random search and local optimization.

2 Deterministic Global Optimization[1]

2.1 DC Optimization

Many powerful techniques in global optimization are based on the fact that many objective functions can be expressed as the *difference of two convex functions* (so called *d.c. functions*). If $D(x)$ is an objective function in $\mathbb{R}^n$, then the representation $D(x) = p(x) - q(x)$, where p, q are convex functions is said to be a *d.c. decomposition* of D.

The space of d.c. functions is closed under many operations frequently encountered in optimization (i.e., sum, product, max, min, etc). Furthermore, based on the simple fact that every locally d.c. function is d.c., a large class of functions in optimization are d.c. For simplicity of notation, consider the d.c. program:

$$\begin{array}{ll} \min & f(x) - g(x) \\ \text{s.t.} & x \in D \end{array} \tag{1}$$

where D is a *polytope* in $\mathbb{R}^n$ with nonempty interior, and f and g are *convex functions* on $\mathbb{R}^n$.

[1] this section was contributed by P.M. Pardalos

By introducing an additional variable t, Problem (1) can be converted into the equivalent Global Concave Minimization problem:

$$\begin{array}{ll} \min & t - g(x) \\ \text{s.t.} & x \in D, f(x) - t \leq 0 \end{array} \tag{2}$$

with *concave objective function* $t - g(x)$ and *convex feasible set* $\{(x,t) \in \mathbb{R}^{n+1} : x \in D, f(x) - t \leq 0\}$. If (x^*, t^*) is an optimal solution of (2), then x^* is an optimal solution of (1) and $t^* = f(x^*)$.

Therefore, any d.c. program of type (1) can be solved by an algorithm for minimizing a concave function over a convex set. Concave function have the nice combinatorial property that if they have a solution, then a solution occurs at an extreme point of the feasible domain. This property is very useful when the feasible domain is defined by a set of linear constraints. Several algorithms have been proposed over the years for the solution of concave minimization problems. These algorithms are based on branch and bound techniques, cutting plane methods, integer programming approaches, optimality conditions, and their efficiency depends on the problem structure (see (Horst and Pardalos, 1995; Pardalos and Romeijn, 2002; Horst et al., 2000; Pardalos and Rosen, 1987)).

2.2 Monotonic Optimization

Monotonicity with respect to some variables (partial monotonicity) or to all variables (total monotonicity) is a natural property exhibited by many problems encountered in applications. The most general problem of *d.i. monotonic optimization* is:

$$\begin{array}{ll} \min & f(x) - g(x) \\ \text{s.t.} & f_i(x) - g_i(x) \leq 0, i = 1, \ldots, m \end{array} \tag{3}$$

where are all functions are increasing on R^n_+. As with the class of DC functions, the class of monotonic functions is closed under many operations. Numerous global optimization problems can be reformulated as monotonic optimization problems. Such problems include multiplicative programming, nonconvex quadratic programming, polynomial programming, and Lipschitz optimization problems.

Assume without loss of generality that $g(x) = 0$. Note that, we have

$\{\forall i \quad f_i(x) - g_i(x) \leq 0\} \Leftrightarrow \max\limits_{1 \leq i \leq m} \{f_i(x) - g_i(x)\} \leq 0 \Leftrightarrow$

$\Leftrightarrow F(x) - G(x) \leq 0$, where

$$F(x) = \max_i \{f_i(x) + \sum_{i \neq j} g_j(x)\},$$

$$G(x) = \sum_i g_i(x).$$

The functions $F(x)$ and $G(x)$ are both increasing. The original problem reduces to the problem of optimizing a monotonic function over a normal domain (Tuy, 2000):

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & F(x) + t \leq F(b), \\ & G(x) + t \geq F(b), \\ & 0 \leq t \leq F(b) - F(0), \\ & x \in [0, b] \subset R^n_+. \end{array} \tag{4}$$

A set $G \subseteq R^n_+$ **normal** if for any two points x, x' such that $x' \leq x$, if $x \in G$, then $x' \in G$.

The solution of the above occurs at the boundary of the feasible domain. Monotonicity analysis was initiated in the early 70's (Papalambros and Wilde, 1979) for optimal design problems. A recent paper that described some algorithmic approaches to solve several types of monotonic optimization problems can be found in (Tuy, 2000)

2.3 Is Continuous Optimization different than Discrete Optimization?

It is clear that $z \in \{0,1\} \Leftrightarrow z - z^2 = z(1-z) = 0$ or $z \in \{0,1\} \Leftrightarrow z + w = 1, z \geq 0, w \geq 0, zw = 0$. Hence, integer constraints are equivalent to continuous nonconvex constraints (complementarity!). Discrete optimization, is a special case of global continuous optimization. It seems that the main difference, regarding the problem computational difficulty, is between convex and nonconvex optimization problems. It can

be shown that the linear complementarity problem is equivalent to the linear mixed integer feasibility problem (Horst et al., 2000). Complementarity is a fundamental tool in continuous optimization since it expresses optimality conditions. Generalizations of complementarity problems lead to variational inequalities and problems with equilibrium constraints (Giannessi et al., 2001).

Given matrices $A_{n\times n}$, $B_{n\times l}$ and a vector $b \in \mathbb{R}^n$ with rational entries, the **mixed integer feasibility problem** is to find (x, z), such that $x \in \mathbb{R}^n$, $x \geq 0$, $z \in \{0,1\}^l$ that satisfy $Ax + Bz = b$.

The mixed integer feasibility problem can be reduced to the solution of a **linear complementarity complementarity problem** $LCP(M, q)$: Find v, y such that

$$v \geq 0, y \geq 0, v^T y = 0, v = My + q,$$

where

$$y = \begin{pmatrix} z \\ x \\ \theta \end{pmatrix}, v = \begin{pmatrix} w \\ s \\ t \end{pmatrix}$$

$$M = \begin{pmatrix} -I & 0 & 0 \\ B & A & 0 \\ -B & -A & 0 \end{pmatrix}, q = \begin{pmatrix} e \\ b \\ -b \end{pmatrix},$$

where $\theta \in \mathbb{R}^n$ and $e \in \mathbb{R}^l$ is the vector of all 1's.

Given the $LCP(M, q)$ consider the following mixed zero-one integer problem:

$$\begin{aligned} 0 \leq My + \alpha q \leq e - z, \\ \alpha \geq 0, 0 \leq y \leq z, \\ z \in \{0,1\}^n. \end{aligned} \tag{5}$$

Let (α^*, y^*, z^*) with $\alpha^* > 0$ be a feasible point of the above mixed integer problem (which is always feasible). Then $x^* = y^*/\alpha^*$ solves the LCP. If $\alpha^* = 0$ is the only feasible value of α, then the LCP has no solution. For details and further results on this equivalence see (Pardalos, 1994; Pardalos, 1996).

During the last few decades, several approaches have been developed that link the discrete universe to the continuous universe through geometric, analytic, and algebraic techniques. Such techniques include global optimization formulations, semidefinite programming, and spectral theory (Pardalos, 1996; Pardalos and Rosen, 1987).

2.4 Hierarchical (Multilevel) Optimization

The mathematical study of hierarchical structures can be found in diverse scientific disciplines including environment, ecology, biology, chemical engineering, classification theory, databases, network design, transportation, game theory and economics. The study of hierarchy occurring in biological structures reveals interesting properties as well as limitations due to different properties of molecules. Multi-level (or hierarchical) programming problems have been studied extensively in their general setting during the last decade. In general, hierarchical optimization problems are nonconvex and therefore is not easy to find globally optimal solutions (Migdalas et al., 1997).

Hierarchical optimization can be used to study properties of these hierarchical designs. In hierarchical optimization, the constraint domain is implicitly determined by a series of optimization problems which must be solved in a predetermined sequence. Hierarchical (or multi-level) optimization is a generalization of mathematical programming. The simplest two-level (or bilevel) programming problem describes a hierarchical system which is composed of two levels of decision makers and is stated as follows:

$$\begin{aligned} \textbf{(BP)} \quad & \min_{y \in Y} && \varphi(x(y), y) \\ \text{subject to} \quad & \psi(x(y), y) \leq 0 \\ & \text{where} \quad x(y) && = \arg\min_{x \in X} f(x, y) \\ \text{subject to} \quad & g(x, y) \leq 0, \end{aligned}$$

where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are closed sets, $\psi : X \times Y \to \mathbb{R}^p$ and $g : X \times Y \to \mathbb{R}^q$ are multifunctions, φ and f are real-valued functions. The set $\mathcal{S} = \{(x, y) : x \in X, y \in Y, \psi(x, y) \leq 0, g(x, y) \leq 0\}$ is the *constraint set* of **BP**.

Many algorithmic developments are based on the properties of special cases of **BP** (and the more general problem) and reformulations to equivalent or approximating models, presumably more tractable. Most of the exact methods are based on branch and bound or cutting plane techniques and can handle only moderately size problems (Migdalas et al., 1997). Only algorithmic approaches that consider the special structure of the problem can be successful in practice. As we show next, a general optimization problem can be reformulated as a multilevel optimization problem. In particular this is very important if a model describes a hierarchical system or design.

2.5 Multivariate Partition Approach

The basic idea of this approach is to partition all the variables appearing in the optimization problem into several groups, each of which consists of some variables, and regard each group as a set of active variables for solving the original optimization problem. With this approach we can formulate optimization problems as multi-level optimization problems.

Consider the following problem:

$$\min_{x \in D \subseteq \mathbb{R}^n} f(x), \tag{1}$$

where D is a robust set and $f(x)$ is continuous.

Let $\{\Delta_i, i = 1, \ldots, p\}$ be a partition of $S = \{x_1, \ldots, x_n\}$, $p > 1$.

The above problem is equivalent to the following multilevel optimization problem:

$$\min_{y_{\sigma_1} \in D_{\sigma_1}} \{ \min_{y_{\sigma_2} \in D_{\sigma_2}} \ldots \{ \min_{y_{\sigma_p} \in D_{\sigma_p}} f(\Delta_1, \ldots, \Delta_p) \} \ldots \}, \tag{2}$$

where $\sigma = (\sigma_1, \ldots, \sigma_n)$ is any permutation of $\{1, 2, \ldots, p\}$. The components of the vector y_{σ_i} coincide with the elements of Δ_i and D_{σ_i} is defined as a feasible domain of y_{σ_i}.

The multivariate partition approach has been used to develop efficient algorithms for the spherical code problem (Huang et al., 2001) and the minimization of Lennard- Jones potential energy function (Huang et al., 2002).

3 Stochastic algorithms[2]

When the size of the problem to be solved increases it is out of question to look for an exact algorithm with certifiable quality, unless in very special cases. Most problems encountered in practice do not lend themselves to an analysis which might lead to the development of an exact global optimization method, capable of stopping in a finite time with a deterministic certificate attesting the quality of the global optimum estimate found. So it is quite natural that, when dealing with large scale, computationally expensive, unstructured global optimization problems, usually we have to resort to an heuristic procedure. It is important to notice, however, that heuristic is not a synonymous of easy, simple to implement, nor inefficient. Many heuristic procedures for global optimization are extremely sophisticated and display impressive performance even over very challenging problems. Also it is to be remarked that heuristic is neither a synonymous of stochastic: many heuristic search methods for global optimization do actually contain a stochastic component, but this is not a rule. What, in the authors' opinion, is extremely important to keep in mind when choosing an heuristic algorithm for global optimization, is that, being the problem so difficult in general, there is no hope of finding a general purpose method capable of producing satisfactory results in most cases: algorithms have to be tuned as much as possible on known problem characteristics if we want to have reasonable chances of discovering a good local optimum, if not the global one.

In this section, rather than presenting a survey of existing algorithms, a task which can be accomplished by looking at the vast literature on global optimization (see, for example, the (NOIA) volume series), we would like to introduce the reader to some elementary concepts which might provide a guidance towards the choice of a stochastic global optimization method for a specific problem class.

First of all, let us limit the analysis to a special case in global optimization: let us assume that the

[2] this section was contributed by F. Schoen

problem under consideration can be formalized as

$$\min_{x \in S \subset \mathbb{R}^n} f(x) \tag{6}$$

where S is a "simple" compact set (like, e.g., a box) and f is a "sufficiently smooth" objective function. By the assertion that S is simple we mean that this problem is considered to be an unconstrained one and that we generally expect that the global optimum is found in the interior of S. The presence of S is just meant to enable us with relative easyness to draw a sample (e.g., an uniform one) – in other words, we do not consider here the complications arising when dealing with *constrained* global optimization problems: the presence of the feasible set is just meant to be an aid to the algorithm designer who would like to limit the search to a compact region. Many practical problems can indeed be formalized this way. As an example, we might cite the well known problem of finding the minimum energy conformation of a cluster of particles interacting via two-body forces:

$$\min \sum_{i<j} v\left(\|X_i - X_j\|\right) \tag{7}$$

where X_i, $i = 1, \ldots, N$ are the coordinates, in $\mathbb{R}^3$, of the centers of N particles and $v(\cdot)$ is the pairwise potential; two well known examples of pair potentials are the Lennard-Jones potential (see, for example (Northby, 1987))

$$v(r) = \frac{1}{r^{12}} - \frac{2}{r^6} \tag{8}$$

and the Morse potential (see (Morse, 1929))

$$v(r) = (\exp(\rho(1-r)) - 1)^2 - 1 \tag{9}$$

where $\rho > 0$ is a parameter.

These two models, which are widely studied both as accurate models of some microclusters, and as approximate models of more complex materials like, e.g., fullerene (Doye et al., 2001) or even proteins, give rise to large scale global optimization problems which are essentially unconstrained, being the coordinates of particles free to be placed in $\mathbb{R}^3$; however, as it is immediately seen, the energy to be minimized depends only on the relative distance between pairs of particles, so that it is invariant with respect to rigid transformations in general and translations in particular. Thus it might safely be assumed that the geometric center of the cluster is in the origin of $\mathbb{R}^3$ and that no two atoms are "too far each other": for example, as the minimum of both (8) and of (9) is attained at $r = 1$, it is quite easy to see that in the global minimum there cannot exist two particles which are at a distance greater than $N - 1$ (actually, as it seems quite reasonable, putative global optima for these cluster models possess a very compact structure and it is widely believed, even if no formal proof has yet been published, that the diameter of globally optimal clusters grows as $\sqrt[3]{N}$).

Thus, going back to our problem statement, here is a case in which no constraint appear in problem formulation, but a box can be quite easily defined containing all of the "interesting" configurations, including the global optimum. The availability of such a box is instrumental towards the definition of good stochastic search methods.

For what concerns the objective function, here we do not assume very much on its smoothness. Usually we require that f is continuous on S, so that, thanks to Weierstrass' theorem, a global optimum surely exists. The more we know on f's structure, the better: should we have the possibility of observing not only the value of f at a sample point x, but also the value of its gradient, $\nabla f(x)$, then we could exploit this knowledge in local exploration.

Before presenting some general ideas on stochastic global optimization approaches, we would like to pause for a while and consider the role of stochastic elements in heuristic algorithms. Let us immagine a scenario in which the objective function is a black box, i.e. it is a complex procedure, possibly a process simulation, in which we provide a tentative value of the control variables $x \in \mathbb{R}^n$ and, after some lengthy computation, the value of $f(x)$ is returned. We have no access to the analytic form of f, nor we can hope to have the possibility of computing the gradient or other information on the function. We even do not know if the function is continuous or not. What can we do in similar cases? Any heuristic algorithm devoted to such a problem should be composed of an

equilibrated mix of two procedures:

1. an *exploration* procedure, whose aim is to observe, as evenly as possible, different regions in S trying to cover this set with a sufficiently dense or a sufficiently well chosen sample;

2. a *refinement* procedure, whose aim is to locally explore the neighborhood of a promising point in search of a better estimate of the global optimum.

Very often, although by no means always, the exploration procedure contains some random sampling mechanism. As an extreme example, this procedure might simply be composed of a uniform random sampling routine capable of generating uniform vectors in S. In other cases this procedure might be composed of a random sampling routine capable of generating a point in the intersection between S and a sphere centered at the current point with prescribed radius. In other cases the procedure consists in finding a random direction out from the current point and then exploring the function along such a direction. The inclusion of randomness in the sampling mechanism has some advantages: first of all it is an easy way to avoid neglecting important parts of S: it is not too difficult, in the examples reported before, to impose conditions that guarantee that, with probability one, every subset of S with positive Lebesgue measure will eventually be visited. Of course a deterministic sampling plan might be used instead of random sampling, and there exist in fact important examples of sampling through quasi-random deterministic sequences with better filling properties than random ones (Niederreiter, 1992). However, as the dimension increases, the feasible space becomes in some sense too large to permit any kind of systematic exploration and random sampling, thanks to its unbiasedness and to the easiness with which random points can be drawn, is generally preferred.

For what concerns refinement, this is usually best performed by means of local optimization, if this is at all possible. When analytical gradients are available, it is possible to use one of the many reliable methods for unconstrained minimization, like, e.g., conjugate gradients or limited memory BFGS algorithms (Liu and Nocedal, 1989). Of course if derivatives are not available, other local methods can be employed like, e.g., grid or direct search strategies (Kolda et al., 2003). Often, however, in this case an effect similar to that of local optimization can be accomplished by sampling in small neighborhoods of the current point.

In order to exploit the characteristics of the problem at hand, besides using gradients if available, we can distinguish two interesting situations:

1. the objective function is multimodal, but there are relatively few local optima with quite large basins of attraction;

2. the objective function is characterized by a huge number of local optima, but these can be considered as small perturbations around a function which is similar to the one described in the previous point.

Of course these two possibilities do not exhaust the possible situations, but they are representative of large classes of practical problems: functions of type 1 often arise when dealing with black box functions in which the model, although analytically not available, may be thought of as smooth and continuous; quite often these problems can be solved by simple variants of Multistart, i.e. by sampling a random uniform point and starting from there a local search, repeating the procedure until some stopping condition is met. As already pointed out, a local search can be implemented through sampling in a small neighborhood of the current point. If the objective function is expensive to evaluate, then it is advisable to build an approximate model of $f(x)$ based on the observations; an interesting approach within this framework was proposed in (Gutmann, 2001), where an interpolation of the objective function is built, based upon radial basis functions, and the next sample point is chosen as the global minimizer of the interpolant.

For what concerns problems of the second type let us first observe that these occur very frequently and are believed to be very common in fields like molecular clusters, protein–protein docking as well as protein folding. In these contexts the energetic landscape is usually composed of a few set of "funnels" with quite stable states at their bottom; each funnel,

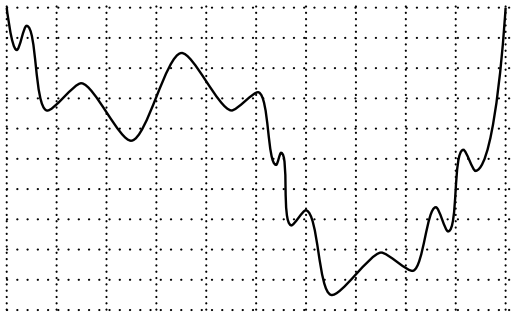

Figure 1: A multimodal function with many local optima, but only two funnels.

Figure 2: The effect of local searches on multimodal functions.

however, is not a simple valley which an hypothetical monotonic gradient search could smoothly traverse towards the minimum. They are, on the contrary, extremely oscillating landscapes which cause any local descent method to be trapped very far from the bottom of the funnel. Even more sophisticated algorithms like simulated annealing are extremely inefficient in these cases as the time scale needed to eventually reach the bottom of the funnel is usually astronomically large. Here, however, quite simple methods can be efficiently employed, based upon an elementary observation. If we substitute the objective function $f(x)$ with another one, $f^1(x) = f(\mathcal{LS}(x))$, in which to each point $x \in S$ a value is associated which corresponds to the local minimum found by a local optimization method $\mathcal{LS}(\cdot)$ applied to f and initialized at x, then the new objective function would be piecewise constant, with far less oscillations than the original one and with the global optimum corresponding to the region of attraction of the global optimum of $f(x)$.

As a simple example in $\mathbb{R}^1$ we can look at the picture in figure 1. There we represent a function which, although characterized by many local optima, might also be considered as a perturbation of a much simpler one, possessing a so-called funnel structure – in other words we might think of this function as a perturbation of a much simpler one, characterized by few local minima with relatively large basins of attraction. If we assume we could replace the objective function $f(x)$ with $f^1(x)$, then we see, looking at figure 2, that this function has the same global optimum value as the original one and it is far less "oscillating"; more important, the funnel structure becomes more apparent in f^1 when compared to the original objective function. Function $f^1(x)$ might be optimized using techniques similar to those suggested in case 1. Of course the analytical form of $f^1(x)$ is not available, but often we are able to sample f^1 at specific points. Algorithms may be designed which try to exploit the simpler form of $f^1(x)$ with respect to the original $f(x)$. An interesting such algorithm is based upon the use of the so called Monotonic Basin Hopping algorithm as a local descent method for $f^1(x)$, coupled with a simple Multistart algorithm. Here we start a local search on f from a random point (exploration); then a random point is generated in a neighborhood of the local optimum found and a local optimization of f is started from there. If this optimization leads to a better optimum, this is accepted and the procedure restarts from this point; otherwise the original local optimum is retained and another point is sampled, until a stopping rule calls for stopping. This way the algorithm performs a monotonic descent path on function f^1 from one local optimum to a neighboring one until a deep valley is reached. This approach, and many variants of it, is currently considered as the best one, both in performance and in the quality of the results, for the optimization of molecular cluster conformations. The local descent Monotonic Basin Hopping (MBH) algorithm can be

described as follows.

Procedure MBH(x,r, MaxNoImprove)
    // x: *starting point*
    // *MaxNoImprove: stopping criterion*
    // r: *radius used in local*
    // *perturbation of the current point*
    NoImprove = 0;
    $x^\star = \mathcal{LS}(x)$;
    record = $f(x^\star)$;
    while (NoImprove < MaxNoImprove)
        y = random uniform point in $S(x^\star, r)$;
        $y^\star = \mathcal{LS}(y)$;
        current = $f(y^\star)$;
        if (current < record)
            // *new record found in* $S(x^\star, r)$;
            record = current;
            $x^\star = y^\star$;
            NoImprove = 0;
        else
            NoImprove = NoImprove + 1;
        end if
    end while
return (record);
end procedure

In the above procedure, r is a real parameter representing the radius of an hypersphere $S(x^\star, r)$, centered at the current point, where sampling is performed; the choice of an appropriate value for r is quite critical, as too small a radius makes the procedure too local, while a choice of a large r value makes the algorithm too inefficient, as it reduces to sampling in a very large region. An interesting research area is that of exploring the possibility of letting r vary in an adaptive way, but we are not aware of any rigorous analysis of this kind of approach. The parameter `MaxNoImprove` is used to stop the search when there is sufficient evidence that no better local optima can be found in the neighborhood (of radius r) of the current point; in other words, when `MaxNoImprove` attempts have been perfomred with no improvement, it is believed that the bottom of the funnel has been reached and the current search is stopped. Inside the procedure, calls are made to $\mathcal{LS}(y)$, which is a local search algorithm which, starting from a feasible point y, returns a local optimum $y^\star$ such that $f(y^\star) \leq f(y)$. This local procedure can be a standard local optimization method, if the objective function is sufficiently smooth and we have the possibility of evaluating derivatives, or it might be a heuristic procedure, possibly based on random sampling or on direct search. MBH is usually called as a procedure in a simple global optimization method. A simple example of the use of MBH for global optimization is given by the following procedure, where MBH is included in Multistart:

Procedure MultiMBH(r, MaxNoImprove,MaxIter)
    // *MaxIter: stopping criterion*
    record = $+\infty$
    iter = 1;
    while (iter $\leq$ MaxIter)
        iter = iter + 1;
        x = random uniform point in S;
        current = MBH(x, r, MaxNoImprove)
        if current < record
            record = current;
        endif
    end while
end procedure

It is quite impressive to see that a very simple algorithm like MBH can be extremely efficient when dealing with problems with a huge number of local optima. A well known example in this framework is the global optimization of Lennard-Jones clusters. Here it is reported in the literature that the first "unbiased" algorithm (i.e., an algorithm which did not included any prior knowledge on the problem or on the location of the optima) capable of finding all of the putative global optima for clusters with less than 110 particles was a variant of MultiMBH (Wales and Doye, 1997) inspired by simulated annealing, in which the acceptance of the current point in MBH is allowed not only when a new record is found, but also, with a prescribed probability, when the `current` value found is not too worse than the `record`. This way the algorithm performs also some ascent step in the landscape of local optima. However, successive experiments showed that using a monotonic procedure (Leary, 2000) does not prevent the discovery of

all putative optima: these results confirm the fact that in a funnel-like structure it is a good idea to try to descend as quickly as possible towards the bottom of a funnel; there is generally no much benefit in slowing down the descent, by allowing ascent steps, in the hope of having the chance of exploring a different funnel: this task, often, is best performed by restarting the procedure. This point of view led us to introduce a new family of algorithms for Lennard-Jones (Locatelli and Schoen, 2002; Locatelli and Schoen, 2003) and for Morse clusters (Doye et al., 2003) which displayed an improvement of one or two orders of magnitude in the number of local searches needed to obtain the putative optimal configurations, in particular for those clusters which are considered as the most difficult to optimize (like 75–78, 98, 102 particles in Lennard-Jones clusters, or most of the configurations with $\rho = 14$ in Morse clusters). The idea which led to such an improvement was that an effort had to be made in order to find good starting points for MBH: the standard MBH method coupled to Multistart can sometimes be too inefficient, in particular when the funnel containing the global optimum is very narrow; our procedure in some sense was meant to augment the probability of sampling a starting point in the "right" funnel; this was accomplished not through a modification of the sampling procedure (something which would have required some prior knowledge on the global optima), but through a modification of the local search method which, in some sense, is made more global. We cannot go into details here, but we refer the interested reader to the bibliography. We can also cite that experiments using the same kind of ideas were performed in order to find the minimum energy coupling of two macro molecules interacting through a complex force field (in particular, we used the (AMBER) force field). Also in this case, where the objective consists in finding the correct docking configuration of two large molecules, typically proteins, we obtained quite impressive improvements with respect to more traditional optimization algorithms (Addis and Schoen, 2003a; Addis and Schoen, 2003b).

Returning to the characterization of global optimization problems, as described in (Locatelli, 2003) we observe that, if local searches of any kind are available, then we might try to solve a global optimization problem by simply sampling f; otherwise, we might apply a global optimization algorithm to the function $f^1(x) = f(\mathcal{LS}(x))$ obtained by applying local searches to f; but we might continue, and apply a global optimization algorithm to the function $f^2(x) = f^1(\mathcal{LS}(x))$ obtained by the application of local searches to f^1, and so on. Here the presentation is over simplified, as it should be remarked that the function f^1 is piecewise constant and traditional local optimization would make no progress from any starting point; what is here denoted by $\mathcal{LS}$ is a more refined procedure which is capable of generating a not ascending path along the steps of f^1; this task might be accomplished, for example, by algorithm MBH. Unimodal optimization problems are those for which a single descent algorithm applied to $f(x)$ would lead to the global optimum; "easy" global optimization problems are those for which a simple global optimization algorithm, like Multistart, equipped with a suitable descent method (like an unconstrained local search, or a direct search if derivatives are unavailable) can be successfully applied directly to $f(x)$. More difficult problems, like e.g. molecular conformation problems, can be attacked through simple global optimization methods applied to $f^1(x)$ (an example of this is Monotonic Basin Hopping); in these cases Multistart alone is not sufficient, if applied to $f(x)$, but it becomes a viable algorithm if applied to $f^1(x)$. Even more challenging problems might be candidates for algorithms working at the f^2 level, but at the moment of writing we know very few attempts of this kind.

Another promising approach, which however is still in its infancy, is that of trying to smooth the piecewise constant function resulting from the application of local optimizations to $f(x)$, i.e. to $f^1(x) = f(\mathcal{LS}(x))$. Performing a local smoothing on this function might help in revealing which are the descent paths – it should be recalled that we are speaking here of descent paths in the function transformed by local searches, which is piecewise constant. A procedure based on local smoothing can be defined in a way which is quite similar to a MBH algorithm: at each step we collect local information on $f^1(x)$ by means of a small sample of local searches performed from

points uniformly sampled in a sphere of radius r centered at the current iterate. Like in MBH, as soon a a global record is observed, the procedure is restarted, and the current point becomes the new record found. However, differently from MBH, if a certain number of local searches have been performed with no improvement in the global record, instead of stopping the search, we try to extract information which might lead to an improvement. Without going into much detail, we remark here that this information can be extracted by means of an approximate smoothing of the observed local optima in the neighborhood of the current one. In particular the current point is moved to the minimum of the function

$$\hat{L}_g^B(x) = \frac{\sum_{i=1}^K \mathcal{LS}(y_i) g(\|y_i - x\|)}{\sum_{i=1}^K g(\|y_i - x\|)}$$

where $B(x)$ is a ball centered at the current point x, $y_i, i = 1, \ldots, K$ are K points sampled in such a sphere and g is a smoothing kernel (e.g., a gaussian density function). Numerical results obtained with classical, high-dimensional, test functions with a huge number of local optima show that the approach is quite promising. This method can be considered within the framework presented in this section as a possibility for building a descent algorithm on the flat steps of $f^1(x)$. MBH is a simpler possibility, but it might be sometimes limited by the fact that if no record is found in the current sphere the algorithm stops; this new approach is aimed at trying to extract some information on the shape of $f^1(x)$ by means of smoothing.

Concluding this brief analysis of stochastic methods, we would like to stress here that, apart from the specific method used in sampling, local approximation, restart, it seems that a promising direction for designing stochastic algorithms for difficult global optimization with high number of local optima comes from the following suggestions:

1. try to exploit as much as possible *local* optimization: do not waste function evaluations and CPU time in trying to *approximate* a local optimum

2. use sampling, possibly guided by the objective function itself, to explore the feasible set, in the hope of placing an observation in the funnel leading to the global optimum

3. explore funnels as quickly as possible, with a descent algorithm like, e.g., MBH, trying to find descent paths in the transformed function f^1

Conclusions

In these short notes it was clearly impossible to offer an exhaustive account of recent advances in a dynamic field like global optimization is nowadays. Hundreds of papers and tens of new books on the subject appear each year. So, rather than presenting a long list of global optimization problem classes and/or algorithms, we preferred to outline some general trends in contemporary algorithmic research, with a style and a focus which is obviously strongly connected with the authors' own personal experience on the field. We hope at least to have given the readers the feeling that global optimization is no more an impossible task and that large scale, difficult multimodal problems can be, quite often, solved at optimality through a carefully selected mixture of modelling, implicit enumeration, local searches, sampling.

Acknowledgements

The research of Fabio Schoen has been partially supported by Progetto FIRB "Ottimizzazione Non Lineare su Larga Scala"

References

JOGO. Journal of global optimization. `http://www.kluweronline.com/issn/0925-5001`.

NOIA. Nonconvex optimization and its applications. `http://www.wkap.nl/prod/s/NOIA`.

Addis, B. and Schoen, F. (2003a). Docking of atomic clusters through nonlinear optimization. To appear in *J. of Global Optimization*.

Addis, B. and Schoen, F. (2003b). A randomized global optimization method for protein-protein docking. Technical Report DSI 4-2003, Dip. Sistemi e Informatica - Univ. Firenze.

AMBER. Amber homepage. `http://www.amber.ucsf.edu/amber/amber.html`.

Dixon, L. and Szegö, G. (1978). *Towards Global Optimization 2*. North-Holland.

Doye, J. P., Leary, R. H., Locatelli, M., and Schoen, F. (2003). The global optimization of Morse clusters by potential energy transformations. *INFORMS Journal On Computing (to appear)*.

Doye, J. P. K., Wales, D. J., Branz, W., and Calvo, F. (2001). Modeling the structure of clusters of C_{60} molecules. *Physical Review B*, 64:235409–1–235409–11. .

Giannessi, F., Maugeri, A., and Pardalos, P., editors (2001). *Equilibrium Problems and Variational Models*. Kluwer Academic Publishers, Boston.

Gutmann, H.-M. (2001). A radial basis function method for global optimization. *Journal of Global Optimization*, 19:201–227.

Horst, R. and Pardalos, P., editors (1995). *Handbook of Global Optimization*. Kluwer Academic Publishers.

Horst, R., Pardalos, P. M., and Thoai, N. (2000). *Introduction to Global Optimization (2nd Edition)*. Kluwer Academic Publishers, Boston.

Huang, H. X., Pardalos, P., and Shen, Z. (2001). A point balance algorithm for the spherical code problem. *Journal of Global Optimization*, 19:329–344.

Huang, H. X., Pardalos, P., and Shen, Z. (2002). Equivalent formulations and necessary optimality conditions for the Lennard-Jones problem. *Journal of Global Optimization*, 22:97–118.

Kolda, T. G., Lewis, R. M., and Torczon, V. (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482.

Leary, R. H. (2000). Global optimization on funneling landscapes. *Journal of Global Optimization*, 18(4):367–383. .

Liu, D. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, B 45:503–528.

Locatelli, M. (2003). On the multilevel structure of global optimization problems. Technical report - Univ. di Torino.

Locatelli, M. and Schoen, F. (2002). Fast global optimization of difficult Lennard-Jones clusters. *Computational Optimization and Applications*, 21(1):55–70.

Locatelli, M. and Schoen, F. (2003). Efficient algorithms for large scale global optimization: Lennard-Jones clusters. *Computational Optimization and Applications*, 26:173–190.

Migdalas, A., Pardalos, P., and Varbrand, P., editors (1997). *Multilevel Optimization: Algorithms and Applications*. Kluwer Academic Publishers, Boston.

Morse, P. M. (1929). Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys. Review*, 34(1):57–64.

Niederreiter, H. (1992). *Random number generation and Quasi-Monte Carlo methods*. SIAM, Philadelphia.

Northby, J. A. (1987). Structure and binding of Lennard–Jones clusters: $13 \leq n \leq 147$. *Journal of Chemical Physics*, 87:6166–6178.

Papalambros, P. and Wilde, D. (1979). Global noniterative design optimization using monotonicity analysis. *Trans. ASME, J. of Mechanical Design*, 101:643–649.

Pardalos, P. (1996). Continuous approaches to discrete optimization problems. In Di Pillo, G. and Giannessi, F., editors, *Nonlinear Optimization and Applications*, pages 313–328. Plenum Publishing.

Pardalos, P. M. (1994). The linear complementarity problem. In Gomez, S. and Hennart, J., editors, *Advances in Optimization and Numerical Analysis*, pages 39–49. Kluwer Academic Publishers, Boston.

Pardalos, P. M. and Romeijn, H. E., editors (2002). *Handbook of Global Optimization Vol. 2* Kluwer Academic Publishers, Boston.

Pardalos, P. M. and Rosen, J. B. (1987). *Constrained Global Optimization: Algorithms and Applications*. Springer-Verlag, Berlin.

Tuy, H. (2000). Monotonic optimization. *SIAM Journal of Optimization*, 11(2):464–494.

Wales, D. J. and Doye, J. P. K. (1997). Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *Journal of Physical Chemistry A*, 101:5111–5116.