

**A Survey on Direct Search Methods
for Blackbox Optimization
and their Applications**

C. Audet

G-2012-53

October 2012

A Survey on Direct Search Methods for Blackbox Optimization and their Applications

Charles Audet

GERAD & Département de mathématiques et de génie industriel

École Polytechnique de Montréal

C.P. 6079, Succ. Centre-ville

Montréal (Québec) Canada, H3C 3A7

`charles.audet@gerad.ca`

October 2012

Les Cahiers du GERAD

G-2012-53

Copyright © 2012 GERAD

Abstract

Blackbox optimization typically arises when the functions defining the objective and constraints of an optimization problem are computed through a computer simulation. The blackbox is expensive to compute, can have limited precision and be contaminated with numerical noise. It may also fail to return a valid output, even when the input appears acceptable. Launching twice the simulation from the same input may produce different outputs. These unreliable properties are frequently encountered when dealing with real optimization problems. The term blackbox is used to indicate that the internal structure of the target problem, such as derivatives or their approximations, cannot be exploited as it may be unknown, hidden, unreliable or inexistent. There are situations where some structure such as bound or linear constraints may be exploited and in some cases, a surrogate of the problem is supplied or a model may be constructed and trusted. This chapter surveys algorithms for this class of problems, including a supporting convergence analysis based on the nonsmooth calculus. The chapter also lists numerous published applications of these methods to real optimization problems.

Acknowledgments: This work was supported by NSERC grant 239436 and AFOSR FA9550-12-1-0198.

1 Introduction

In many situations, one is interested in identifying the values of a set of variables that maximize or minimize some objective function. Furthermore, the variables cannot take arbitrary values, as they are confined to an admissible region and need to satisfy some prescribed requirements. Optimization studies problems of the form:

$$\min_{x \in \Omega} f(x), \quad (1)$$

where x represents the variables which must be taken in the admissible region Ω , a subset of \mathbb{R}^n , and where $f(\cdot)$ is the objective that we wish to minimize. The goal is to identify $x \in \Omega$ that has the least objective function value $f(x)$. The nature of f and of Ω dictates the type of optimization methods that should be used to tackle a given problem. In general, properties such as linearity, convexity, monotonicity and integrality may be used to identify appropriate optimization methods [68].

The paper focuses on optimization problems for which the nature of f or Ω is difficult or impossible to exploit. These are called *blackbox optimization problems*. Frequent examples are when the functions defining the problems are computed through a time consuming simulation. Some of the constraints defining Ω may be evaluated prior to launching the simulation, and others may only be evaluated *a posteriori*. In [116, 112] for example, launching a single simulation at a given tentative trial point $x \in \mathbb{R}^n$ to evaluate the objective function value $f(x)$ and to verify whether x belongs to Ω or not requires as much as 2000 computational hours. In other situations, the simulation can fail to return a value even when the tentative x belongs to Ω . In [37], 60% of the simulation calls return an error. There are also situations where the simulation is not deterministic. The simulation might involve random numbers, and may return different values of $f(x)$ for the same input x . In [29], the objective function value represents the time required to perform a series of tasks, and varies slightly even when the same conditions are prescribed.

There are situations where the evaluation of f is done through a blackbox simulation, but the problem is not as ill-conditioned as in the above-mentioned examples. Model-based derivative-free methods target this class of problems by constructing approximations of the objective function [44, 45, 46]. They do not assume explicit knowledge of the derivatives, but assume existence of the derivatives. The book [47] on derivative-free method distinguishes model-based and directional-based optimization method. The present work discusses directional-based optimization methods for general blackbox optimization on which no assumptions are made on the objective function or on the admissible region.

The chapter studies the optimization problem (1) and is structured as follows. Section 2 gives a high level overview of some directional-based methods designed to solve blackbox optimization problems. Section 3 describes typical exploitable specificities of the target problem. In particular, specific types of constraints and variables are discussed. Then, Section 4 describes how one can use surrogates and develop a model of the target problem to ease the solution process. These tools are used to guide the optimization so that it consumes less time evaluating the original expensive problem. That section also mentions work on the exploitation of parallelism in direct-search methods. Section 5 presents the theoretical foundations upon which these methods rely. The convergence analysis is hierarchical in the sense that the stronger the hypotheses on the objective and feasible region, the stronger the resulting theoretical guarantees. Finally, Section 6 surveys some selected published applications of these directional direct-search methods. Concluding remarks are drawn in the final section.

2 Directional direct search methods

Direct-search methods were introduced more than half a century ago. They are named this way because they interact directly with the function values, and do not attempt to use or estimate derivatives. Pioneer methods include the famous coordinate search (CS) method used by [62] on one of the first digital computers back in 1952, the Nelder-Mead [122] and the Hooke and Jeeves [87] algorithms. This section first gives a general overview of the CS method, followed by some of its descendants.

2.1 The coordinate search algorithm

Part of the structure of some modern direct search methods is present in the CS method for unconstrained optimization, *i.e.*, problem (1) with $\Omega = \mathbb{R}^n$. This iterative method can be simply described as follows.

At iteration $k \in \{0, 1, 2, \dots\}$, the current best known solution is denoted by x_k and called the current incumbent solution. At the initial iteration, the starting point $x_0 \in \mathbb{R}^n$ is supplied by the user of the method. An initial step size parameter $\Delta_0 > 0$ is also supplied by the user.

Then, at each iteration, a total of $2n$ trial points are generated in hopes of improving the current incumbent. Each trial point is obtained by varying a single coordinate of the vector x_k by a fixed step size of magnitude Δ_k . These are called the poll points and belong to the set

$$P_k = \{x_k \pm \Delta_k e_i : i = 1, 2, \dots, n\}$$

where e_i is the i -th coordinate vector in \mathbb{R}^n . The objective function f is then evaluated at each of these points, and there are two possible outcomes. One possibility is that a trial point $t \in P_k$ satisfying $f(t) < f(x_k)$ is identified. In that case the iteration is declared successful and x_{k+1} is set to t and the step size parameter Δ_{k+1} to Δ_k . The alternate possibility is that all $2n$ trial points were tested, but none improved the objective function value. In that case the iteration is said to be unsuccessful and x_{k+1} is set to x_k , but the step size parameter Δ_{k+1} is set to be half of Δ_k . The iteration then ends, the counter k is incremented by one and a new iteration is initiated.

This method is simple to implement. Furthermore, it may be applied to any unconstrained optimization problem, without any assumptions on the smoothness of the objective function. The CS algorithm is also known under the name *compass search* [96]. Section 5 reports that even if this method may seem naive, it is supported by a convergence analysis showing that it may produce a stationary point when the objective function is locally strictly differentiable.

2.2 The pattern search class of algorithms

Development and analyses of direct-search methods were not intensive for a few decades, but renewed interest occurred in the nineties. In [119], it was shown that the Nelder-Mead method could fail to converge to a local solution, even on a strictly convex unconstrained two-dimensional minimization problem.

An unified framework was proposed in [139], generalizing the following direct search methods: CS, Hooke and Jeeves, multidirectional search [57] as well as an evolutionary operation method [38]. The framework was called *pattern search*, or *generalized pattern search* (GPS), and one of the main contribution was to provide sufficient conditions on the target problem ensuring convergence to a stationary point for unconstrained optimization.

The GPS framework generalizes CS by increasing its flexibility. The main algorithmic improvements are the following.

1. The directions used to generate the poll set P_k are not restricted to the coordinate vectors.
2. When the iteration succeeds in improving the incumbent, the parameter Δ_{k+1} is allowed to increase or to remain the same as Δ_k .
3. At every iteration, the algorithm allows exploration at a finite number of trial points other than the poll points.

The first improvement allows a richer set of polling directions. Lewis and Torczon [104] propose the use of positive bases [54, 15] to generate the polling directions. Positive bases are not bases, but are minimal sets of directions whose nonnegative linear combinations span \mathbb{R}^n . The set of positive and negative coordinate directions in \mathbb{R}^n used by CS is an example of a positive basis. The flexibility of the construction of the polling directions is exploited in [13] for molecular geometry problems.

The second improvement allows the algorithm to dynamically adapt to the nature of the problem. Indeed, if the initial step size Δ_0 was chosen to be too small, then a series of successful iterations will increase Δ_k

and large steps will soon be taken. This modification was shown useful in practice, but the analysis of the behaviour of the parameter Δ_k gets more technical. Torczon [139] sets some rules on the way in which the parameter is increased or decreased, including a requirement that the factor by which it is modified be a rational number. She then gives sufficient conditions under which the limit inferior of the parameter Δ_k goes to zero as the iteration number k goes to infinity. This fundamental result is the first cornerstone of the convergence analysis. The rationality requirement was later shown to be necessary for the analysis [14].

The third generalization is introduced for practical reasons. A user of a pattern search method is often tempted to alter an optimization code to make it more efficient by integrating his knowledge of the target problem. Modifications of the poll set P_k could lead to loss of the structure necessary for the theoretical support. So, an additional exploration phase, later called the SEARCH step [37], is introduced at every iteration. The search step allows the evaluation of the objective function at finitely many trial points, located on a discretization of the space of variables. The discretization is called the mesh, and its coarseness is parameterized by the step size parameter Δ_k . The mesh is formally described in the next subsection.

2.3 The mesh adaptive direct search class of algorithms

There are two main practical and theoretical limitations to GPS. First, the directions used to construct the mesh and the poll set need to be chosen from a fixed finite set $D \subset \mathbb{R}^n$. The algorithm and its convergence analysis heavily rely on this requirement. A negative consequence of limiting the polling directions is exposed in [96] where CS is applied to a modification of the Dennis-Wood function [59] and the iterates converge to a non-stationary point. The objective function of this problem is simply the maximum of two strictly convex quadratics in \mathbb{R}^2 , and the problem is unconstrained. If one knew in advance the polling directions, it would be easy to devise a similar example for GPS.

The second limitation of GPS is that it cannot handle general constraints. In fact, it can only treat explicit linear constraints and bounds on the variables. In [23], the Mesh Adaptive Direct-Search (MADS) class of algorithms is introduced to address these limitations.

In MADS, the role of the step size parameter Δ_k is divided in two. That parameter is replaced by the mesh and the poll size parameters Δ_k^m and Δ_k^p . As its name indicates, the poll size parameter is used to construct the poll set P_k . The poll points are constructed around the current incumbent solution x_k and the distance separating the incumbent to each poll point is limited by Δ_k^p . In comparison, that distance is systematically equal to Δ_k with CS. The mesh size parameter dictates the coarseness or fineness of the mesh. In MADS, Δ_k^m is updated in a way that it converges to zero much faster than Δ_k^p . The consequence is that the tentative search and poll points can be chosen on a finer mesh than the mesh defined by the poll size parameter Δ_k^p .

As in GPS, MADS uses a fixed finite set of directions called D . Typically, D is composed of the $2n$ positive and negative coordinate directions. At iteration k , the mesh is a discretization of the space of variables on which all tentative search and poll points need to be selected. Formally, the mesh is defined as

$$M_k = \{x + \Delta_k^m Dz : x \in V_k, z \in \mathbb{N}^{n_D}\} \subset \mathbb{R}^n,$$

where V_k denotes the set of trial points where the simulation was launched by the start of iteration k . That set is also known as the *cache* as it contains the history of all evaluated trial points.

The poll set P_k is composed of mesh points whose distance (the infinity norm is frequently used) from the current incumbent solution x_k is bounded above by a constant $c > 0$ times the poll size parameter

$$P_k \subseteq \{x \in M_k : \|x - x_k\| \leq c\Delta_k^p\}.$$

Notice that if the constant c equals one, if $\Delta_k^p = \Delta_k^m$ and if $D = [I; -I]$, then the set on right-hand-side of this inclusion corresponds exactly to the poll set P_k of Section 2.1, with the additional point x_k .

The poll set P_k must contain at least $n + 1$ points since x_k must lie in the strict interior of its convex hull. An equivalent way of stating this last requirement is that the polling directions must form a positive spanning set.

Figure 1 illustrates the effect of different mesh and poll size parameters in \mathbb{R}^2 . In all plots, the arrows represent the directions of the positive spanning set D used to construct the mesh. The mesh is represented by the intersection of the lines. The darker lines delimit the region in which the poll points must be chosen. The incumbent solution x_k is located at the intersection of the arrows at the centre of each subfigure.

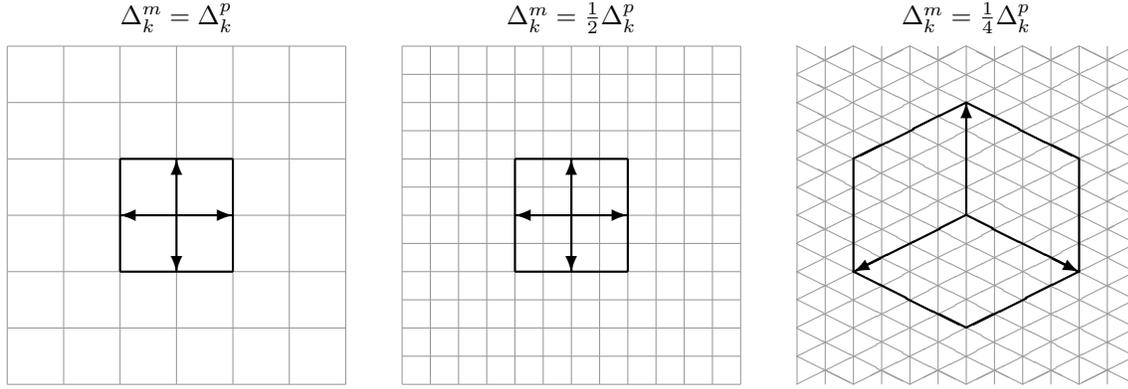


Figure 1: Examples of meshes in \mathbb{R}^2 .

The leftmost figure represents the situation in which both mesh and poll size parameters are equal, and in which the mesh directions of D are the positive and negative coordinate directions, i.e., it depicts the CS poll set. There is only one possibility to define the poll set with GPS.

The central figure represents an instance of MADS in which the mesh size parameter is half the poll size parameter. The directions are once again the positive and negative coordinate directions, but the poll points of P_k can be chosen with more flexibility as they do not need to be located at the endpoints of the arrows. Any combination that contain x_k in the strict interior of its convex hull can be chosen.

The rightmost figure is conceptually identical to the middle one, except that the mesh directions are not the coordinate directions. Again, there is great flexibility in choosing the poll points inside the hexagonal frame.

The key algorithmic element introduced in MADS is that the polling directions are not restricted to a fixed finite set. As Δ_k^m and Δ_k^p converge to zero, the set of normalized directions which may be used to construct the poll set P_k becomes dense in the unit sphere. In the leftmost figure, there are $3^2 - 1$ mesh points from which P_k can be chosen. In the central figure, with a ratio of $\frac{1}{2}$ for $\frac{\Delta_k^m}{\Delta_k^p}$ that number grows to $5^2 - 1$. For a ratio equal to $\lambda \in \mathbb{N}$, that number reaches $(2\lambda + 1)^2 - 1$ for a generalization of this example.

Figure 2 gives a high-level description of a MADS algorithm. Aside from the blackboxes defining the problem, the only element that the user must supply is a starting point x_0 , at which the simulation runs successfully. As detailed in the next section, this does not mean that the initial point must be feasible, but only requires that the simulation does not fail. This is modeled by requiring that $f(x_0) < \infty$.

In practice, a typical termination criteria for direct-search algorithms is based on an overall budget of calls to the simulation, or on wall clock time. Another possibility is to terminate when the mesh size parameter drops below a certain threshold.

The first instantiation of MADS, called LTMADS [23] uses randomly generated non-singular triangular matrices to generate the directions used to construct the poll set P_k . Three years later, the ORTHOMADS instantiation [10] used Householder matrices to generate orthogonal maximal positive bases.

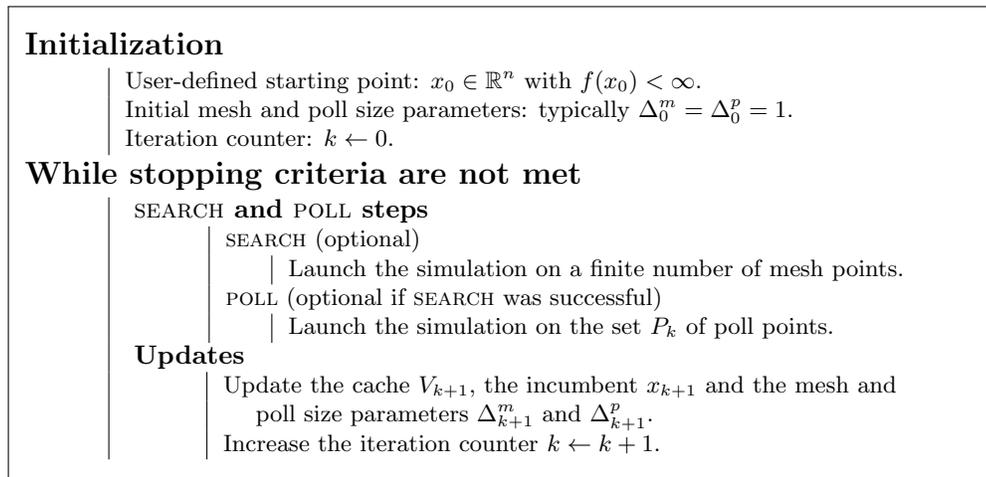


Figure 2: High level description of the MADS algorithm for solving the optimization problem (1).

2.4 Sufficient decrease methods

As mentioned in the previous section, MADS generalizes GPS by decoupling the role of the mesh and poll size parameters. An alternate strategy to allow a rich set of directions consists in removing the mesh requirement, but in accepting a trial point as being the new incumbent solution only if the decrease in the objective function value is sufficiently important. For continuously differentiable functions, such strategies were proposed and analyzed in [111] and [72] as well as in the frame-based method described in [48, 127] and in the generating set search (GSS) methods [96, 98].

For Lipschitz continuous functions, such sufficient decrease methods ensure a minimal displacement in the space of variables when accepting a new incumbent. This displacement plays a role similar to that of the mesh requirement and prevents the iterates to converge prematurely to an undesirable solution.

3 Exploitable specificities of the target problem

Even if the target problem is provided as a blackbox, there are situations where some structure is available and may be exploited. This section lists some examples.

3.1 Integer and categorical variables

Integer variables can easily be handled by mesh-based methods such as GPS and MADS. Indeed, the mesh M_k imposes a discrete structure on the space of variables. It suffices to make sure that the mesh points are integers, and the natural stopping criteria consists in terminating as soon as the mesh size parameter drops below the value one. This is illustrated by the two first plots of Figure 1. Using the coordinate directions to define the mesh and an integral mesh size parameter ensures that all trial points satisfy the integrality requirement.

Discrete optimization variables are said to be *categorical* if the objective function or the constraints can not be evaluated unless the variables take one of a prescribed enumerable set of values. Categorical variables differ from integer variables as they do not possess any natural ordering. These variables need to be accompanied by a user-defined notion of neighbourhood, specific to the target problem. The GPS and MADS algorithms are extended to allow such variables in [20] and [7], respectively. In these papers, the variables are partitioned into three groups: continuous, integer and categorical variables. There is no special treatment for the continuous variables. The integer variables are handled by making sure that the mesh only contains integer points. This is done easily by specifying a minimal mesh size for these variables. The user-defined neighbourhood is used to defined the poll points of the categorical variables.

3.2 Explicit bound and linear constraints

GPS algorithms are adapted in [105, 106] to handle bound and linear inequalities. The extension consists of making sure at each iteration that the poll directions positively span the tangent cone at the nearly active constraints. In bound constraint optimization, it suffices to take the positive and negative coordinate directions. For linear inequalities and equalities, the explicit knowledge of the constraints is used to generate the tangent cone [103]. Degeneracy issues can be handled by the strategy proposed in [11].

3.3 General constraints

Of the methods enumerated above, only MADS can handle general nonsmooth constraints. The union of normalized MADS poll directions grows asymptotically dense in the unit sphere, and this allows the treatment of constraints by the extreme barrier in which infeasible trial points are simply rejected from consideration. This aggressive treatment of constraints is necessary when the simulation cannot be launched when a constraint is violated. For example, a simulation that computes logarithms or square roots cannot be trusted when negative values are entered. These are called unrelaxable constraints. There are also practical situations where the simulation fails inexplicably. These are often referred to as hidden constraints [40].

The extreme barrier has the merit of being simple to implement, but there are more subtle ways to handle the relaxable constraints whose amount by which they are violated is available. For continuously differentiable functions, augmented Lagrangian approaches are presented in [107, 97] for GPS and GSS, respectively.

In [67, 66], a filter method is proposed for nonlinear programming. The main component of a filter method is a constraint violation function h that aggregates the violations of each individual constraint. The function h is nonnegative, and equal to zero only when the corresponding trial point is feasible. Filter methods exploit tradeoffs between the reduction of the objective f and the constraint violation h . Filter methods are adapted to GPS [22, 9] and to frame-based methods [56].

More recently, another mechanism called the progressive barrier was proposed [24] to treat nonsmooth quantifiable constraints. The progressive barrier imposes a maximal threshold on the constraint violation h which is progressively reduced. Trial points whose constraint violation value exceed the threshold are rejected from consideration. Among all infeasible trial points that are not rejected by the progressive barrier, a local exploration is conducted around the one with the best objective function value.

An hybrid method is presented in [26] for the situation where the initial point x_0 does not satisfy all of the quantifiable constraints. Under this strategy, the constraints are initially handled by the progressive barrier, and as soon as an individual constraint is satisfied by an incumbent solution, then the treatment of that constraint is done by the extreme barrier.

3.4 Minimax optimization problems

The lack of differentiability may come from a variety of sources. One one them occurs when the objection function of Problem (1) is obtained by taking the maximum of finitely many functions $f^i : X \rightarrow \mathbb{R} \cup \{\infty\}$ for $i = 1, 2, \dots, q$:

$$f(x) = \max\{f^1(x), f^2(x), \dots, f^q(x)\}.$$

The maximum operator introduces nondifferentiability, even if the finitely many functions are differentiable. Minimax optimization problems can be used to model worst case scenarios: one may wish to minimize the highest possible loss, *i.e.*, to minimize the largest of the $f^i(x)$ values..

This class of problem is in [109] in the case where Ω is defined by linear constraints, and the functions f^i are twice continuously differentiable. The finite minimax structure is exploited by a smoothing technique based on an exponential penalty function.

More recently, [83] studies the unconstrained case where the functions are continuously differentiable. They exploit the structure of the problem by identifying the active manifold and then treat the objective as a smooth function restricted to the manifold.

3.5 Multi-objective optimization and trade-off studies

There are situations where one is interested in analyzing the tradeoffs between multiple objectives $f^{(p)}$, $p = 1, 2, \dots, q$. There is no single objective function that encapsulates the totality of the design process. In such a situation, the goal of the optimization is not to produce a single solution, but to produce the collection of Pareto undominated solutions. The required computational effort increases rapidly with the number of objectives. A feasible solution x is said to be dominated by another $x' \in \Omega$ when $f^{(p)}(x') \leq f^{(p)}(x)$ for every objective functions with a strict inequality for at least one objective function. The Pareto set is defined to be the set of undominated solutions.

The first direct search algorithm for biobjective optimization was introduced in [30] and then generalized to more than two objectives in [31] by incorporating the normal-boundary intersection [53] technique. The method approximates the Pareto front by launching a series of optimization on single-objective reformulations of the problem. A different mechanism is proposed in [51].

In [27], two strategies are developed to analyze the sensitivity of an optimal solution to general constraints, including bounds on variables, with the help of a direct-search solver. A simple method performed immediately after a single optimization by inspecting the cache, and a detailed one performing biobjective optimization on the minimization of the objective versus the constraint of interest. The resulting analysis helps in identifying the relative importance of the constraints.

4 Tools for dealing with costly blackboxes

In addition to the specificities outlined in the previous section, there are situations where additional tools are available. This section discusses surrogates, models and parallelism.

4.1 Static surrogate functions

In blackbox optimization, the functions defining the target problem are expensive to evaluate. A frequently used strategy consists in designing a second blackbox optimization problem called the surrogate. A surrogate needs to share some similarities with the expensive optimization problem, but must be cheaper to evaluate.

Static surrogates may be constructed by reducing the number of internal iterations, or through simplified physics model for example. The surrogate management framework [58, 37] manages the interplay between the fidelity of the surrogate problem to ensure that the optimization process converges to a solution of the original target problem. The variable precision of a surrogate is exploited in [125] to reduce the overall computational effort.

In the context of GPS and MADS algorithms, the surrogate may be used at many places in the algorithm. A first obvious usage consists in solving the surrogate optimization problem and to use the best solution(s) as starting points for the optimization of the true problem. Other usages consist in solving subproblems on the surrogate, and only to evaluate the true function at the solution of the subproblem. For example, in [20], an extended poll is conducted when categorical variables are modified. The extended poll can be viewed as a descent in a subspace of variables, using the surrogate problem. In [16], the variable neighborhood search (VNS) metaheuristic [120, 121] is used to attempt to escape locally optimal solutions. The descent is performed on the surrogate problem since VNS may be expensive in terms of function evaluations.

In addition, surrogates may be used at every iteration to order the tentative search and poll points so that the most promising ones are treated first. The list of tentative points is sorted with regards to their surrogate values and then, the expensive simulation is launched on the most promising ones first. An ordering in the presence of constraints is proposed in [43]. The process terminates as soon as a new incumbent is generated, thereby reducing the number of expensive function calls.

Notice that for a surrogate to be efficient, it does not need to be an accurate model of the true problem. The introduction of the chapter mentioned a problem [29] in which the objective function value represents the time required to perform a set of tasks. The surrogate used in that work consists of the time to perform

a small number of these tasks. The units for the surrogate values are seconds, and the units for the true simulations are hours. The surrogate value is not at all a good approximation, but it is very useful as it shares some similarities with the true problem.

4.2 Dynamic models

Static surrogates are usually supplied by the developer of the optimization problem. An alternate way to define a surrogate is to construct approximations of the objective and constraints. These models are then used in the SEARCH and POLL steps of Algorithm 2 as detailed in the previous subsection. As the algorithm is deployed on a problem, more and more simulations are launched, and the newly collected information can be used to recalibrate and improve the fidelity of the models. This dynamic way of constructing models is done in the update step of algorithm 2.

A natural option is to consider the quadratic models described in [52, 43] which are constructed and updated by considering past evaluations from the cache that are close to the current incumbent solution. For target problems that are not noisy, quadratic models may lead to significant improvements and increase the speed of convergence to a local solution.

Other strategies to dynamically construct models include DACE Kriging [34, 110, 134, 50], treed Gaussian processes [78] and radial basis functions [33, 128, 123, 142].

4.3 Parallelism

Most modern machines now have multiple processors. A first parallel synchronous version of GPS is presented in [57]. The Asynchronous Parallel Pattern Search algorithm APPS [88, 80, 95] removes this synchronization barrier. Adaptation to GSS are presented in [82, 81]. The asynchronous versions are especially useful when the blackbox has heterogeneous computing times depending on the trial point where it is evaluated. A convergence analysis is presented in [99] for the smooth case.

Based on a remark of [60] stating that the parallel variable distribution of Ferris and Mangasarian [63] should be paired with GPS, [25] proposes another strategy to exploit parallelism.

5 Theoretical foundations

None of the methods surveyed in this paper can guarantee convergence to a global minimizer of Problem (1). As stated in its title [136], global optimization requires global information. In blackbox optimization there is no information available, even less global information.

The convergence analysis looks at the sequence of trial points, and studies some of its accumulation points as the iteration number goes to infinity. Of course, this is a theoretical analysis since in practice one cannot let $k \rightarrow \infty$. But the analysis is useful as it shows limiting behaviors. Based on local properties of the objective and constraints, the analysis ensures that some necessary optimality conditions are met. This section summarizes the analysis for smooth and nonsmooth problems.

5.1 Smooth unconstrained optimization

Like Newton's method for unconstrained optimization, the CS and GPS algorithm may get stuck at saddle points if the polling directions are not properly chosen. For example, if Newton's method or CS is applied to the unconstrained minimization of the quadratic function $f(x) = x_1x_2$ from the origin, then the sequence of iterates stagnate at the origin which is a stationary point.

However, [5] shows that GPS and CS cannot converge to a strict local maximizer, unlike Newton's method. This may appear to be surprising that a method using first and second derivatives of a C^2 function ensures weaker convergence results than a method that uses only function values, without using nor estimating derivatives. This study is generalized in [6] for MADS.

It is shown in [61] that if $\nabla f(\cdot)$ is Lipschitz continuous, then at an unsuccessful iteration k of GPS, $\nabla f(x_k)$ is bounded above by a constant times the current mesh size parameter, whose limit inferior converges to zero [139]. An interesting consequence of this result is that it provides a theoretical justification of the stopping criteria based on a small mesh size parameter.

5.2 Nonsmooth analysis for unconstrained optimization

As mentioned in the introduction, the target problems for which the direct-search methods are designed are typically blackbox problems. The function values returned by the blackbox are the result of an expensive simulation. There is no reason to believe that they should be differentiable nor even continuous functions. Therefore, the convergence results outlined in the previous subsection are certainly true, but are somewhat incompatible with the target problems.

The paper [21] studies the convergence of the GPS method under less restrictive assumptions for unconstrained optimization. They propose a hierarchy of convergence results based on local smoothness of the objective function. In order to achieve this, they consider the set of unsuccessful iterations, i.e., the iterations where the incumbent solution x_k is shown to have an objective function value less than or equal to that of the neighbouring poll points in P_k . Such an incumbent is called a mesh local optimizer. They then consider subsequence of unsuccessful iterations for which the corresponding incumbent solutions converge to a limit point denoted \hat{x} and for which the corresponding subsequence of mesh size parameters converges to zero. Such a subsequence of iterates is called a refined subsequence, and \hat{x} a refined point.

The fundamental convergence result does not require any assumption on the objective function. It is called the zero-th order result, and states that \hat{x} is the limit of mesh local optimizers on meshes that get arbitrarily fine.

The Clarke calculus for nonsmooth function [42] generalizes notions such as the directional derivative and the gradient to non-differentiable functions. The fundamental convergence theorem of [21] states that the Clarke generalized directional derivative at a refined point \hat{x} in a direction d used infinitely often in the refining subsequence

$$f^\circ(\hat{x}; d) := \limsup_{y \rightarrow \hat{x}, t \downarrow 0} \frac{f(y + td) - f(y)}{t}$$

is nonnegative, if f is locally Lipschitz near \hat{x} . The convergence analysis then progressively adds assumptions on f such as local regularity and strict differentiability [102].

The convergence analysis of MADS [23, 17] strengthens this result to $f^\circ(\hat{x}; d) \geq 0$ for every direction $d \in \mathbb{R}^n$. This last result can be stated in an equivalent way as follows: 0 belongs to the generalized gradient of f at \hat{x} . A similar convergence analysis is developed for the DIRECT algorithm [64] and to sampling methods for perturbed Lipschitz functions [65].

5.3 Nonsmooth stationarity for constrained optimization

In smooth optimization, a necessary optimality condition states that if \hat{x} is a local minimizer of the function f over the domain $\Omega \subseteq \mathbb{R}^n$, then the directional derivative of f at \hat{x} in every tangent direction d to Ω is nonnegative:

$$f'(\hat{x}; d) \geq 0, \quad \text{for every } d \in T_\Omega(\hat{x}).$$

Another way to see this optimality condition is to state that there are no feasible descent directions at \hat{x} .

Using the Clarke calculus, and generalizations of the tangent cone, [23] shows that the refined point \hat{x} generated by the MADS algorithm under the extreme barrier satisfies

$$f^\circ(\hat{x}; d) \geq 0, \quad \text{for every } d \in T_\Omega^H(\hat{x})$$

under the assumption that f is locally Lipschitz near \hat{x} , and where $T_\Omega^H(\hat{x})$ is called the hypertangent cone [89, 130], and is a nonsmooth generalization of the tangent cone. The Rockafellar upper subderivative [130] is defined for non-Lipschitz functions, and is analyzed in [141] to enrich the convergence hierarchy.

When handling some relaxable quantifiable constraints with the progressive [24] rather than the extreme barrier, the method may generate feasible or infeasible refined points. The same necessary optimality conditions are guaranteed for the feasible refined points. However, the analysis shows that an infeasible refined point \hat{x} satisfies

$$h^\circ(\hat{x}; d) \geq 0, \quad \text{for every } d \in T_X^H(\hat{x})$$

where h is the constraint violation function mentioned in Section 3.3, and X is the domain corresponding to the unrelaxable or non-quantifiable constraints, and are handled by the extreme barrier. Roughly speaking, this suggests that the algorithm reached at a local minima of the constraint violation. This occurs in particular when there are no feasible solutions to the optimization problem.

6 Applications to real blackbox problems

The methods discussed above were created to be applied on real blackbox optimization problems. The present section lists some of these applications. They are presented in non-disjoint groups and the list is not exhaustive. Other applications can be found in the introductory chapter of [47].

Some of these applications were solved by the NOMAD [8, 100, 101] open source C++ implementation of MADS for single or biobjective blackbox optimization problems of the form (1). NOMAD integrates the features listed in the present paper, as well as others such as latin hypercube sampling [138], periodic variables [28] and batch or library modes.

6.1 Shape optimization

The PhD thesis [112] applies the surrogate management framework mentioned in Section 4.1 to identify the shape of a hydrofoil trailing-edge that minimizes the aerodynamic noise propagated to the far field. Computations require large-eddy simulations. Unconstrained results for laminar flow are presented [114], and deformation of upper and lower surfaces of the trailing-edge in laminar flow with lift and drag constraints are analyzed in [115] resulting in as much as 70% reduction in noise. Reynolds-averaged Navier-Stokes calculations are incorporated for constraint evaluation to make the optimization more efficient in [116] leading to a 89% noise reduction.

A framework for coupling optimal shape design to time-accurate three-dimensional blood flow simulations in idealized cardiovascular geometries is presented in [113]. Results on idealized Y-shaped baffle for the Fontan surgery for children with congenital heart defects are shown in [143]. Uncertainties in the simulation input parameters as well as shape design variables are accounted for in [132] using the adaptive stochastic collocation technique of [131].

In [126], the evolution and propagation of cracks in two dimensional elastic domains are studied. The simulation requires a finite element discretization and a set of partial differential equations with nonlinear boundaries. The ultimate goal is to determine the optimal shape resulting in a crack path with as much energy as possible without completely destroying the specimen.

6.2 Positioning problems

There are situations in which the blackbox is expensive to evaluate, but some properties of the variables are known and may be exploited. One such family of problems are those where some or all of the variables represent spacial coordinates. Over the last years, direct-search methods were applied to some nonsmooth positioning problems.

Following the 2004 Indian Ocean tsunami, an effort led by the National Oceanic and Atmospheric Administration's Pacific Marine Environmental Labs to identify the optimal position to deploy tsunami detection buoys in the Pacific Ocean [135]. The question was formulated as an optimization problem, where the variables were the coordinates of the buoys, and the objective was to maximize the warning time to coastal cities in the event of a tsunami. Constraints on water depth and bottom roughness were incorporated

and NOMAD would then position a selected number of buoys within a sub-region of the ocean so as to optimize the detection time for a set of unit sources.

In [70] a water supply problem and a hydraulic capture problem are proposed as a challenge to blackbox optimization community. The objective of the water supply problem is to minimize the cost to supply a specific quantity of water subject to a set of constraints on the net extraction rate, pumping rates, and hydraulic head. The decision variables are the two-dimensional locations and pumping rates of the wells, as well as the number of wells. The objective of the hydraulic capture CP is to minimize the cost needed to prevent an initial contaminant plume from spreading by using wells to control the direction and extent of advective fluid flow. The methods used in the study are APPS [88], Boeing DE [34, 50], DIRECT [90, 71], IFFCO [75, 93], NOMAD and NSGA-II [55]. Additional tests with IFFCO are conducted in [69].

More recently, researchers [41, 117] use snow-monitoring devices to estimate the quantity of water stored in snow over a vast domain. When the accumulated snow melts in spring, important quantities of water are liberated, and a precise estimation is necessary for efficient management of hydroelectrical dams. In [12], the question of identifying the optimal position of these snow-monitoring devices is studied, to minimize the overall kriging approximation error of the quantity of water. Different strategies of exploiting the fact that the variables represent locations are presented.

In [118], the question of positioning antennas in an irregular-shaped domain and assigning radio frequencies in a telecommunications network is studied as a blackbox optimization problem. The proposed methodology combines MADS for the positioning problem with a tabu search for the radio frequencies.

6.3 Parameter estimation

There are situations in which a model is characterized by a set of parameters. The question of assigning good values to the parameters may be formulated as an optimization problem in which the difference between observed and actual data points needs to be minimized. Below are examples of such applications.

In [108] An automated flow is described for total-ionizing dose (TID)-aware SPICE model generation that includes TID response and its dependence on process variability and layout. A differential evolution algorithm is adapted for global exploration, and a modified GPS strategy is introduced for local exploration. The optimizer efficiently reduces the value of different kinds of objective functions in the extraction at reasonable cost and avoids premature convergence in most practical cases.

Facial recognition systems are studied in [39]. Given a database of image samples of known individuals the task is to design a system that for any input image, identifies the input with one of the known individuals. The classification problem involves designing a function to map feature vectors to the appropriate class label. The MADS algorithm was shown to outperform heuristics both in accuracy and processing time.

In [137], a conditional averaging approach to estimate the parameters of a land surface water and energy balance model is presented. The parameters are then used to classify net radiation and precipitation. The paper proposes an objective function that approximates the temperature-and moisture-dependent errors in terms of atmospheric forcing, surface states, and model parameters. Minimization of the approximated error yields parameters for model applications.

A method for determining the fire front positions for optically thin flames and the rate of spread of forest or vegetation fires is presented in [91]. The first step of the method measures the heat fluxes coming from the flame by a specific thermal sensor in four horizontal directions. In the second step, these heat fluxes are approximated by a radiative transfer equation. Then, the positions of the fire front and the flame characteristics are determined by applying an inverse method. The rate of spread is deduced by applying a least-square regression on the position values.

In [84], a method for evaluating the kinetic constants in a rate expression for catalytic combustion applications using experimental light-off curves is presented. The method uses a finite element reactor model to simulate reactor performance. The heat and mass transfer models used account for developing flow in the entrance region. A GPS algorithm is used to determine the best fit parameters.

6.4 Tuning of algorithms

Many algorithms depend on a set of parameters. As long as the parameters satisfy some prescribed requirements, the algorithm may be trusted to perform adequately. The question of finding good parameter values has been studied under different names, including parameter tuning, software automatic tuning and parameter optimization. Direct-search methods are used in [85] to optimize computations involving matrices. In [29], the question of adjusting the four trust-region algorithmic parameters so to minimize the overall computational time to solve a large collection of CUTEr [77] test problems was studied and solved using NOMAD. The sensitivity to the parameters was studied in [76].

A more general blackbox formulation of this question is proposed in [18] through the OPAL framework and applied to the DFO algorithm [44]. The user of the OPAL framework must supply a target algorithm together with a set of metrics defining the notions of acceptable parameter values and of performance of the algorithm as well as a collection of representative sets of valid input data for the target algorithm.

Parallelism may be used at various levels within the OPAL framework. It may be used at the blackbox solver level to concurrently test parameter values, or it can be used to assess the quality of a set of parameters on different test problems. These two strategies and a combination of both are studied in [19].

The OPAL framework is applied in [124] on a matrix multiplication algorithm where optimization with respect to blocking, loop unrolling and compiler flags takes place. This application requires the use of categorical variables.

A methodology based on variable selection and a sensitivity analysis of inputs is applied on several instructive data sets, and a analysis of automatic computer code tuning is presented in [79]. A set of extensible and portable search problems in automatic performance tuning is proposed in [32].

6.5 Engineering design applications

The papers [36, 35] and thesis [133] study a design problem of interest to Boeing. It consists of minimizing a vibration measure of a helicopter rotor blade. A simplified surrogate simulation code is used, requiring only a few minutes in contrast with hours for the true simulation. In addition, approximately 60% of the simulation calls fail to return a value. This work is the first to illustrate the use of the surrogate management framework [37].

Categorical variables are studied in [94] to minimize the power required to maintain heat shields at given temperatures in a thermal insulation system. Variables defining this optimization problem include thicknesses, temperatures but also the types of materials used as insulators as well as the number of shields. This means that the number of variables defining the problem is itself an optimization variable. Nonlinear constraints are considered in [4]. A formulation of the problem without the use of categorical variables is presented in [3].

The environmental impact of a commercial aircraft departure is defined by noise nuisance in the protected zones near airports, local air quality, and global warming. A multiobjective, constrained, nonlinear optimization problem is formulated to obtain optimal departure procedures in [140].

Spent potliners are a toxic byproduct generated by the aluminum industry. A process treatment of spent potliners is presented in [49], and the seven input parameters are optimized over four blackbox constraints. The simulation requires the Aspen simulation software [1].

The MADS algorithm is applied in [92] in combination with full-field electromagnetic simulations to tailor the broadband spectral response of gold and silver split-ring resonator metamaterials.

In a series of paper [73, 74, 129], the FactSage thermodynamic software [2] coupled with the NOMAD software to optimize alloy and process designs. Mono and biobjective constrained problems are studied. The FactSage database contains thermodynamic properties as functions of temperature, pressure and composition for over 5000 pure substances and hundreds of multicomponent solid and liquid solutions.

7 Discussion

Since the nineties, there has been a renewed interest in direct search methods for nonsmooth blackbox optimization without derivatives. Recent methods can now handle general constraints, multiple objectives, integer and categorical variables, and can exploit models and surrogates to guide the optimization. An important effort has been deployed to embed the methods into a general framework and to develop a convergence analysis for this framework. Many of these methods are now supported by a hierarchical analysis that ensures necessary stationary conditions based on local properties of the objective and constraints.

Section 6 shows that there are numerous applications of these optimization methods on real problems. With only a few exceptions, most of these blackboxes are not openly distributed to the optimization community. Some of them use proprietary codes, others can only be released internally. Consequently, there are not many real blackbox optimization problems that can be shared and used for benchmarking different methods.

Comparing direct search methods for blackbox optimization on smooth problems from the CUTer [77] or from the Hock and Schittkowsky [86] collection (or on some perturbed variants) is not ideal, as these problems do not possess the same kind of difficulties than blackbox problems without derivatives from real applications. Hopefully, more and more real test problems will be shared to help developers to design more efficient optimization methods.

References

- [1] Aspentech. <http://www.aspentech.com/>.
- [2] FactSage. <http://http://www.factsage.com/>.
- [3] K. Abhishek, S. Leyffer, and J. Linderoth. Modeling without categorical variables: a mixed-integer nonlinear program for the optimization of thermal insulation systems. *Optimization and Engineering*, 11:185–212, 2010. 10.1007/s11081-010-9109-z.
- [4] M.A. Abramson. Mixed variable optimization of a load-bearing thermal insulation system using a filter pattern search algorithm. *Optimization and Engineering*, 5(2):157–177, 2004.
- [5] M.A. Abramson. Second-order behavior of pattern search. *SIAM Journal on Optimization*, 16(2):315–330, 2005.
- [6] M.A. Abramson and C. Audet. Convergence of mesh adaptive direct search to second-order stationary points. *SIAM Journal on Optimization*, 17(2):606–619, 2006.
- [7] M.A. Abramson, C. Audet, J.W. Chrissis, and J.G. Walston. Mesh adaptive direct search algorithms for mixed variable optimization. *Optimization Letters*, 3(1):35–47, 2009.
- [8] M.A. Abramson, C. Audet, G. Couture, J.E. Dennis, Jr., S. Le Digabel, and C. Tribes. The NOMAD project. Software available at <http://www.gerad.ca/nomad>.
- [9] M.A. Abramson, C. Audet, and J.E. Dennis, Jr. Filter pattern search algorithms for mixed variable constrained optimization problems. *Pacific Journal of Optimization*, 3(3):477–500, 2007.
- [10] M.A. Abramson, C. Audet, J.E. Dennis, Jr., and S. Le Digabel. OrthoMADS: A deterministic MADS instance with orthogonal directions. *SIAM Journal on Optimization*, 20(2):948–966, 2009.
- [11] M.A. Abramson, O.A. Brezhneva, J.E. Dennis Jr., and R.L. Pingel. Pattern search in the presence of degenerate linear constraints. *Optimization Methods and Software*, 23(3):297–319, 2008.
- [12] S. Alarie, C. Audet, V. Garnier, S. Le Digabel, and L.A. Leclaire. Snow water equivalent estimation using blackbox optimization. Technical Report G-2011-09, Les cahiers du GERAD, 2011. To appear in *Pacific Journal of Optimization*.
- [13] P. Alberto, F. Nogueira, H. Rocha, and L.N. Vicente. Pattern search methods for user-provided points: application to molecular geometry problems. *SIAM Journal on Optimization*, 14(4):1216–1236, 2004.
- [14] C. Audet. Convergence results for pattern search algorithms are tight. *Optimization and Engineering*, 5(2):101–122, 2004.
- [15] C. Audet. A short proof on the cardinality of maximal positive bases. *Optimization Letters*, 5(1):191–194, 2011.
- [16] C. Audet, V. Béchar, and S. Le Digabel. Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *Journal of Global Optimization*, 41(2):299–318, 2008.
- [17] C. Audet, A.L. Custódio, and J.E. Dennis, Jr. Erratum: Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 18(4):1501–1503, 2008.

- [18] C. Audet, C.-K. Dang, and D. Orban. Algorithmic parameter optimization of the DFO method with the OPAL framework. In *Software Automatic Tuning: From Concepts to State-of-the-Art Results*, chapter 15, pages 255–274. Springer, k. naono and k. teranishi and j. cavazos and r. suda edition, 2010.
- [19] C. Audet, C.-K. Dang, and D. Orban. Efficient use of parallelism in algorithmic parameter optimization applications. Technical Report G-2011-03, Les cahiers du GERAD, 2011. To appear in *Optimization Letters*.
- [20] C. Audet and J.E. Dennis, Jr. Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 11(3):573–594, 2001.
- [21] C. Audet and J.E. Dennis, Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889–903, 2003.
- [22] C. Audet and J.E. Dennis, Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.
- [23] C. Audet and J.E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.
- [24] C. Audet and J.E. Dennis, Jr. A progressive barrier for derivative-free nonlinear programming. *SIAM Journal on Optimization*, 20(4):445–472, 2009.
- [25] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Parallel space decomposition of the mesh adaptive direct search algorithm. *SIAM Journal on Optimization*, 19(3):1150–1170, 2008.
- [26] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Globalization strategies for mesh adaptive direct search. *Computational Optimization and Applications*, 46(2):193–215, 2010.
- [27] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Trade-off studies in blackbox optimization. Technical Report G-2010-49, Les cahiers du GERAD, 2010. To appear in *Optimization Methods and Software*.
- [28] C. Audet and S. Le Digabel. The mesh adaptive direct search algorithm for periodic variables. *Pacific Journal of Optimization*, 8(1):103–119, 2012.
- [29] C. Audet and D. Orban. Finding optimal algorithmic parameters using derivative-free optimization. *SIAM Journal on Optimization*, 17(3):642–664, 2006.
- [30] C. Audet, G. Savard, and W. Zghal. Multiobjective optimization through a series of single-objective formulations. *SIAM Journal on Optimization*, 19(1):188–210, 2008.
- [31] C. Audet, G. Savard, and W. Zghal. A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research*, 204(3):545–556, 2010.
- [32] P. Balaprakash, S.M. Wild, and B. Norris. Spapt: Search problems in automatic performance tuning. *Procedia Computer Science*, 9:1959–1968, 2012. Proceedings of the International Conference on Computational Science, ICCS 2012.
- [33] M. Björkman and K. Holmström. Global optimization of costly nonconvex functions using radial basis functions. *Optimization and Engineering*, 1:373–397, 2000.
- [34] A.J. Booker. Well-conditioned Kriging models for optimization of computer simulations. Technical Report M&CT-TECH-00-002, Boeing Computer Services, Research and Technology, M/S 7L-68, Seattle, Washington 98124, 2000.
- [35] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.W. Moore, and D.B. Serafini. Managing surrogate objectives to optimize a helicopter rotor design – further experiments. AIAA Paper 1998-4717, Presented at the 8th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, 1998.
- [36] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, and V. Torczon. Optimization using surrogate objectives on a helicopter test example. In J. Borggaard, J. Burns, E. Cliff, and S. Schreck, editors, *Optimal Design and Control*, Progress in Systems and Control Theory, pages 49–58, Cambridge, Massachusetts, 1998. Birkhäuser.
- [37] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, V. Torczon, and M.W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization*, 17(1):1–13, 1999.
- [38] G.E.P. Box. Evolutionary operation: A method for increasing industrial productivity. *Appl. Statist.*, 6:81–101, 1957.
- [39] C.-D. Caeanu, X. Mao, G. Pradel, S. Moga, and Y. Xue. Combined pattern search optimization of feature extraction and classification parameters in facial recognition. *Pattern Recognition Letters*, 32(9):1250–1255, JUL 1 2011.
- [40] T.D. Choi and C.T. Kelley. Superlinear convergence and implicit filtering. *SIAM Journal on Optimization*, 10(4):1149–1162, 2000.
- [41] Y. Choquette, P. Lavigne, P. Ducharme, A. Houdayer, and J.-P. Martin. Apparatus and Method for Monitoring Snow Water Equivalent and Soil Moisture Content Using Natural Gamma Radiation, September 2010. US Patent No. 7800051 B2.

- [42] F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, as Vol. 5 in the series Classics in Applied Mathematics.
- [43] A.R. Conn and S. Le Digabel. Use of quadratic models with mesh adaptive direct search for constrained black box optimization. Technical Report G-2011-11, Les cahiers du GERAD, 2011. To appear in *Optimization Methods and Software*.
- [44] A.R. Conn, K. Scheinberg, and Ph.L. Toint. DFO (derivative free optimization). Software available at <http://oss.software.ibm.com/developerworks/opensource/coin/>.
- [45] A.R. Conn, K. Scheinberg, and Ph.L. Toint. On the convergence of derivative-free methods for unconstrained optimization. In M.D. Buhmann and A. Iserles, editors, *Approximation Theory and Optimization: Tributes to M.J.D. Powell*, pages 83–108. Cambridge University Press, Cambridge, United Kingdom, 1997.
- [46] A.R. Conn, K. Scheinberg, and L.N. Vicente. Global convergence of general derivative-free trust-region algorithms to first and second order critical points. *SIAM Journal on Optimization*, 20(1):387–415, 2009.
- [47] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*. MOS/SIAM Series on Optimization. SIAM, Philadelphia, 2009.
- [48] I.D. Coope and C.J. Price. Frame-based methods for unconstrained optimization. *Journal of Optimization Theory and Applications*, 107(2):261–274, 2000.
- [49] Y. Courbariaux, J. Chaouki, and C. Guy. Update on spent potliners treatments: kinetics of cyanides destruction at high temperature. *Ind. Eng. Chem. Res.*, 43(18):5828–5837, August 2004.
- [50] E.J. Cramer and J.M. Gablonsky. Effective parallel optimization of complex computer simulations. In *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, August 2004.
- [51] A.L. Custódio, J.F.A. Madeira, A.I.F. Vaz, and L.N. Vicente. Direct multisearch for multiobjective optimization. *SIAM Journal on Optimization*, 21(3):1109–1140, 2011.
- [52] A.L. Custódio, H. Rocha, and L.N. Vicente. Incorporating minimum Frobenius norm models in direct search. *Computational Optimization and Applications*, 46(2):265–278, 2010.
- [53] I. Das and J.E. Dennis, Jr. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998.
- [54] C. Davis. Theory of positive linear dependence. *American Journal of Mathematics*, 76:733–746, 1954.
- [55] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, apr 2002.
- [56] J.E. Dennis, Jr., C.J. Price, and I.D. Coope. Direct search methods for nonlinearly constrained optimization using filters and frames. *Optimization and Engineering*, 5(2):123–144, 2004.
- [57] J.E. Dennis, Jr. and V. Torczon. Direct search methods on parallel machines. *SIAM Journal on Optimization*, 1(4):448–474, 1991.
- [58] J.E. Dennis, Jr. and V. Torczon. Managing approximation models in optimization. In N.M. Alexandrov and M.Y. Hussaini, editors, *Multidisciplinary Design Optimization: State of the Art*, pages 330–347. SIAM, Philadelphia, 1997.
- [59] J.E. Dennis, Jr. and D.J. Woods. Optimization on microcomputers: The Nelder–Mead simplex algorithm. In Arthur Wouk, editor, *New Computing Environments: Microcomputers in Large-Scale Computing*, pages 116–122. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1987.
- [60] J.E. Dennis, Jr. and Z. Wu. *Parallel continuous optimization*, pages 649–670. Sourcebook of parallel computing. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
- [61] E.D. Dolan, R.M. Lewis, and V. Torczon. On the local convergence of pattern search. *SIAM Journal on Optimization*, 14(2):567–583, 2003.
- [62] E. Fermi and N. Metropolis. Numerical solution of a minimum problem. Los Alamos Unclassified Report LA-1492, Los Alamos National Laboratory, Los Alamos, USA, 1952.
- [63] M.C. Ferris and O.L. Mangasarian. Parallel variable distribution. *SIAM Journal on Optimization*, 4(4):815–832, 1994.
- [64] D.E. Finkel and C.T. Kelley. Convergence analysis of the DIRECT algorithm. Technical Report CRSC-TR04-28, Center for Research in Scientific Computation, 2004.
- [65] D.E. Finkel and C.T. Kelley. Convergence analysis of sampling methods for perturbed lipschitz functions. *Pacific Journal of Optimization*, 5(2):339–350, 2009.
- [66] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming, Series A*, 91:239–269, 2002.
- [67] R. Fletcher, S. Leyffer, and Ph.L. Toint. On the global convergence of an SLP-filter algorithm. Technical Report NA/183, Dundee University, Department of Mathematics, 1998.

- [68] R. Fourer and D. Orban. Dr. Ampl – a meta-solver for optimization problem analysis. *Computational Management Science*, 7(4):437–463, 2009.
- [69] K.R. Fowler, C.T. Kelley, C.T. Miller, C.E. Kees, R.W. Darwin, J.P. Reese, M.W. Farthing, and M.S.C. Reed. Solution of a well-field design problem with implicit filtering. *Optimization and Engineering*, 5(2):207–234, 2004.
- [70] K.R. Fowler, J.P. Reese, C.E. Kees, J.E. Dennis Jr., C.T. Kelley, C.T. Miller, C. Audet, A.J. Booker, G. Couture, R.W. Darwin, M.W. Farthing, D.E. Finkel, J.M. Gablonsky, G. Gray, and T.G. Kolda. Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. *Advances in Water Resources*, 31(5):743–757, 2008.
- [71] J.M. Gablonsky and C.T. Kelley. A locally-biased form of the DIRECT algorithm. *Journal of Global Optimization*, 21:27–37, 2001.
- [72] U.M. García-Palomares and J.F. Rodríguez. New sequential and parallel derivative-free algorithms for unconstrained optimization. *SIAM Journal on Optimization*, 13(1):79–96, 2002.
- [73] A.E. Gheribi, C. Audet, S. Le Digabel, E. Bétille, C.W. Bale, and A.D. Pelton. Calculating optimal conditions for alloy and process design using thermodynamic and properties databases, the FactSage software and the Mesh Adaptive Direct Search algorithm. *CALPHAD: Computer Coupling of Phase Diagrams and Thermochemistry*, 36:135–143, 2012.
- [74] A.E. Gheribi, C. Robelin, S. Le Digabel, C. Audet, and A.D. Pelton. Calculating all local minima on liquidus surfaces using the factsage software and databases and the mesh adaptive direct search algorithm. *The Journal of Chemical Thermodynamics*, 43(9):1323–1330, 2011.
- [75] P. Gilmore and C.T. Kelley. An implicit filtering algorithm for optimization of functions with many local minima. *SIAM Journal on Optimization*, 5(2):269–285, 1995.
- [76] N.I.M. Gould, D. Orban, A. Sartenaer, and Ph.L. Toint. Sensitivity of trust-region algorithms on their parameters. *4OR*, 3(3):227–241, 2005.
- [77] N.I.M. Gould, D. Orban, and Ph.L. Toint. CUTer (and SifDec): a constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2003.
- [78] R.B. Gramacy and S. Le Digabel. The mesh adaptive direct search algorithm with treed Gaussian process surrogates. Technical Report G-2011-37, Les cahiers du GERAD, 2011.
- [79] R.B. Gramacy, M.A. Taddy, and S.M. Wild. Variable selection and sensitivity analysis via dynamic trees with an application to computer code performance tuning. Technical Report 1108.4739, arXiv, August 2011.
- [80] G.A. Gray and T.G. Kolda. Algorithm 856: APPSPACK 4.0: Asynchronous parallel pattern search for derivative-free optimization. *ACM Transactions on Mathematical Software*, 32(3):485–507, 2006.
- [81] J.D. Griffin and T.G. Kolda. Nonlinearly-constrained optimization using heuristic penalty methods and asynchronous parallel generating set search. *Applied Mathematics Research eXpress*, 25(5):36–62, 2010.
- [82] J.D. Griffin, T.G. Kolda, and R.M. Lewis. Asynchronous parallel generating set search for linearly-constrained optimization. *SIAM Journal on Scientific Computing*, 30(4):1892–1924, 2008.
- [83] W. Hare and M. Macklem. Derivative-free optimization methods for finite minimax problems. *To appear in Optimization Methods and Software*, 2012.
- [84] R.E. Hayes, F.H. Bertrand, C. Audet, and S.T. Kolaczowski. Catalytic combustion kinetics: Using a direct search algorithm to evaluate kinetic parameters from light-off curves. *The Canadian Journal of Chemical Engineering*, 81(6):1192–1199, 2003.
- [85] N.J. Higham. Optimization by direct search in matrix computations. *SIAM Journal on matrix analysis and applications*, 14:317–333, 1993.
- [86] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, Berlin, Germany, 1981.
- [87] R. Hooke and T.A. Jeeves. Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 8(2):212–229, 1961.
- [88] P.D. Hough, T.G. Kolda, and V. Torczon. Asynchronous parallel pattern search for nonlinear optimization. *SIAM Journal on Scientific Computing*, 23(1):134–156, 2001.
- [89] J. Jahn. *Vector Optimization: Theory, Applications, and Extensions*. Springer, Berlin, 2004.
- [90] D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Application*, 79(1):157–181, 1993.
- [91] Chetehouna K., O. Sero-Guillaume, I. Sochet, and A. Degiovanni. On the experimental determination of flame front positions and of propagation parameters for a fire. *International Journal Of Thermal Sciences*, 47(9):1148–1157, SEP 2008.
- [92] L.A. Sweatlock K. Diest and D.E. Marthaler. Metamaterials design using gradient-free numerical optimization. *Journal of Applied Physics*, 108(8):1–5, 2010.

- [93] C.T. Kelley. *Iterative Methods for Optimization*. Number 18 in Frontiers in Applied Mathematics. SIAM, Philadelphia, 1999.
- [94] M. Kokkolaras, C. Audet, and J.E. Dennis, Jr. Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system. *Optimization and Engineering*, 2(1):5–29, 2001.
- [95] T.G. Kolda. Revisiting asynchronous parallel pattern search for nonlinear optimization. *SIAM Journal on Optimization*, 16(2):563–586, 2005.
- [96] T.G. Kolda, R.M. Lewis, and V. Torczon. Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.
- [97] T.G. Kolda, R.M. Lewis, and V. Torczon. A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints. Technical Report SAND2006-5315, Sandia National Laboratories, USA, 2006.
- [98] T.G. Kolda, R.M. Lewis, and V. Torczon. Stationarity results for generating set search for linearly constrained optimization. *SIAM Journal on Optimization*, 17(4):943–968, 2006.
- [99] T.G. Kolda and V. Torczon. On the convergence of asynchronous parallel pattern search. *SIAM Journal on Optimization*, 14(4):939–964, 2004.
- [100] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Transactions on Mathematical Software*, 37(4):44:1–44:15, 2011.
- [101] S. Le Digabel and C. Tribes. NOMAD user guide. Technical Report G-2009-37, Les cahiers du GERAD, 2009.
- [102] E.B. Leach. A note on inverse function theorem. In *Proceedings of the AMS*, volume 12, pages 694–697, 1961.
- [103] R.M. Lewis, A. Shepherd, and V. Torczon. Implementing generating set search methods for linearly constrained optimization. *SIAM Journal on Scientific Computing*, 29(6):2507–2530, 2007.
- [104] R.M. Lewis and V. Torczon. Rank ordering and positive bases in pattern search algorithms. Technical Report 96–71, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, Virginia 23681–2199, 1996.
- [105] R.M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):1082–1099, 1999.
- [106] R.M. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM Journal on Optimization*, 10(3):917–941, 2000.
- [107] R.M. Lewis and V. Torczon. A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Optimization*, 12(4):1075–1089, 2002.
- [108] M. Li, Y.F. Li, Y.J. Wu, S. Cai, N.Y. Zhu, N. Rezzak, R.D. Schimpf, D.M. Fleetwood, J.Q. Wang, X.X. Cheng, Y. Wang, D.L. Wang, and Y. Hao. Including Radiation Effects and Dependencies on Process-Related Variability in Advanced Foundry SPICE Models Using a New Physical Model and Parameter Extraction Approach. *IEEE Transactions On Nuclear Science*, 58(6, Part 1):2876–2882, DEC 2011. IEEE Radiation Effects Data Workshop (REDW)/48th IEEE International Nuclear and Space Radiation Effects Conference (NSREC), Las Vegas, NV, JUL 25-29, 2011.
- [109] G. Liuzzi, S. Lucidi, and M. Sciandrone. A derivative-free algorithm for linearly constrained finite minimax problems. *SIAM Journal on Optimization*, 16(4):1054–1075, April 2006.
- [110] S. Lophaven, H. Nielsen, and J. Søndergaard. Dace: A matlab kriging toolbox version 2.0. Technical Report IMM-REP-2002-12, Informatics and Mathematical Modelling, Technical University of Denmark, 2002.
- [111] S. Lucidi and M. Sciandrone. On the global convergence of derivative-free methods for unconstrained optimization. *SIAM Journal on Optimization*, 13:97–116, 2002.
- [112] A.L. Marsden. *Aerodynamic Noise Control by Optimal Shape Design*. PhD thesis, Stanford University, 2004.
- [113] A.L. Marsden, J.A. Feinstein, and C.A. Taylor. A computational framework for derivative-free optimization of cardiovascular geometries. *Computer Methods in Applied Mechanics and Engineering*, 197(21–24):1890–1905, 2008.
- [114] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Optimal aeroacoustic shape design using the surrogate management framework. *Optimization and Engineering*, 5(2):235–262, 2004.
- [115] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Suppression of airfoil vortex-shedding noise via derivative-free optimization. *Physics of Fluids*, 16(10):L83–L86, 2004.
- [116] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Trailing-edge noise reduction using derivative-free optimization and large-eddy simulation. *Journal of Fluid Mechanics*, 572:13–36, 2007.
- [117] J.-P. Martin, A. Houdayer, C. Lebel, Y. Choquette, P. Lavigne, and P. Ducharme. An unattended gamma monitor for the determination of snow water equivalent (SWE) using the natural ground gamma radiation. In *Nuclear Science Symposium Conference Record*, pages 983–988. IEEE, 2008.
- [118] A. Marty. *Optimisation du placement et de l’assignation de fréquence d’antennes dans un réseau de télécommunications*. Master’s thesis, École Polytechnique de Montréal, 2011.

- [119] K.I.M. McKinnon. Convergence of the Nelder-Mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9:148–158, 1998.
- [120] N. Mladenović and P. Hansen. Variable neighborhood search. *Computers and Operations Research*, 24(11):1097–1100, 1997.
- [121] P. Hansen N. Mladenović. Variable neighborhood search: principles and applications. *European Journal of Operational Research*, 130(3):449–467, 2001.
- [122] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [123] R. Oeuvray and M. Bierlaire. BOOSTERS: A derivative-free algorithm based on radial basis functions. *International Journal of Modelling and Simulation*, 2008, to appear.
- [124] D. Orban. Templating and automatic code generation for performance with python. Technical Report G-2011-30, Les cahiers du GERAD, 2011.
- [125] E. Polak and M. Wetter. Precision control for generalized pattern search algorithms with adaptive precision function evaluations. *SIAM Journal on Optimization*, 16(3):650–669, 2006.
- [126] M. Prechtel, G. Leugering, P. Steinmann, and M. Stingl. Towards optimization of crack resistance of composite materials by adjustment of fiber shapes. *Engineering Fracture Mechanics*, 78(6):944–960, APR 2011.
- [127] C.J. Price and I.D. Coope. Frames and grids in unconstrained and linearly constrained optimization: A nonsmooth approach. *SIAM Journal on Optimization*, 14:415–438, 2003.
- [128] R.G. Regis and C.A. Shoemaker. Constrained global optimization of expensive black box functions using radial basis functions. *Journal of Global Optimization*, 31:153–171, 2005.
- [129] E. Renaud, C. Robelin, A.E. Gheribien, and P. Chartrand. Thermodynamic evaluation and optimization of the Li, Na, K, Mg, Ca, Sr // F, Cl reciprocal system. *Journal Of Chemical Thermodynamics*, 43(8):1286–1298, AUG 2011.
- [130] R.T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canad. J. Math.*, 32(2):257–280, 1980.
- [131] S. Sankaran, C. Audet, and A.L. Marsden. A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation. *Journal of Computational Physics*, 229(12):4664–4682, 2010.
- [132] S. Sankaran and A.L. Marsden. The impact of uncertainty on shape optimization of idealized bypass graft models in unsteady flow. *Physics of Fluids*, 22(12):121902, December 2010.
- [133] D.B. Serafini. *A Framework for Managing Models in Nonlinear Optimization of Computationally Expensive Functions*. PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1998.
- [134] J. Søndergaard. *Optimization Using Surrogate Models — by the Space Mapping Technique*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, 2003.
- [135] M. C. Spillane, E. Gica, and V. V. Titov. Tsunameter Network Design for the U.S. DART Array. *AGU Fall Meeting Abstracts*, page A1368, December 2009.
- [136] C.P. Stephens and W. Baritomba. Global optimization requires global information. *Journal of Optimization Theory and Applications*, 96:575–588, 1998.
- [137] J. Sun, G.D. Salvucci, D. Entekhabi, and L. Farhadi. Parameter estimation of coupled water and energy balance models based on stationary constraints of surface states. *Water Resources Research*, 47, FEB 10 2011.
- [138] B. Tang. Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397, 1993.
- [139] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, 1997.
- [140] R. Torres, C. Bès, J. Chaptal, and J.-B. Hiriart-Urruty. Optimal, environmentally-friendly departure procedures for civil aircraft. *Journal of Aircraft*, 48(1):11–22, 2011.
- [141] L.N. Vicente and A.L. Custódio. Analysis of direct searches for discontinuous functions. *Mathematical Programming*, 133(1-2):299–325, 2012.
- [142] S.M. Wild and C.A. Shoemaker. Global convergence of radial basis function trust region derivative-free algorithms. *SIAM J. Optimization*, 21(3):761–781, 2011.
- [143] W. Yang, J.A. Feinstein, and A.L. Marsden. Constrained optimization of an idealized y-shaped baffle for the fontan surgery at rest and exercise. *Computer Methods in Applied Mechanics and Engineering*, 199(33-36):2135–2149, July 2010.