Noname manuscript No.
(will be inserted by the editor)

Global optimization of expensive black box problems with a known lower bound

Andrea Cassioli · Fabio Schoen

Abstract In this paper we propose an algorithm for the global optimization of computationally expensive black–box functions. For this class of problems, no information, like e.g. the gradient, can be obtained and function evaluation is highly expensive. In many applications, however, a lower bound on the objective function is known; in this situation we derive a modified version of the algorithm in [5]. Using this information produces a significant improvement in the quality of the resulting method, with only a small increase in the computational cost. Extensive computational results are provided which support this statement.

Keywords Global Optimization, black-box function, expensive objective functions, radial basis method, bumpiness

1 Introduction

In many practical applications the objective function (or the constraints) are very expensive to be evaluated (typically hours or even days of CPU time). In such situations it is desirable to reduce as much as possible the number of function evaluations. Traditional derivative–free methods based on direct search or gradient estimation via numerical differentiation are not viable as they require too many function evaluations. The most successful methods for similar problems are based on surrogate models of the objective function incrementally built from all the available function evaluations.

Many methods exist within this framework and excellent surveys are available in the literature – see e.g. [8,3]; most of the cited literature agree in suggesting that an interpolation or a regression model based on a sufficiently representative basis of functions is the best choice for function approximation. Typical choices for the bases are radial basis functions possibly extended with the use of low degree polynomials. Many existing methods base the choice of the next point where to evaluate the objective function simply on the surrogate model, typically choosing its minimizer as a predictor for the next evaluation point. Although

Dipartimento di Sistemi e Informatica
Università degli Studi di Firenze – Via di S.Marta 3, 50139 Firenze (Italy)
E-mail: cassioli@dsi.unifi.it,fabio.schoen@unifi.it

sensible, this policy is too myopic and tends to produce iterates that degenerate into a local search. The best approaches, in the authors' opinion, are those in which this step is performed by globally optimizing a suitable merit function on the feasible set. While the merit function might be the surrogate model itself, typically it chosen so that its global optimum corresponds to a point where it is "likely" to be able to observe a low function value. An excellent way to do so is through the introduction of a "bumpiness" function, as proposed by [5]: there, given a threshold, or aspiration, level, a point is chosen so that a surrogate model which includes an interpolation at such a point whise value is the chosen threshold is as little bumpy as possible.

This paper is devoted to an extension of the classical method of [5]; in many applications which require the optimization of costly black-boxes a lower bound on the optimal value is frequently available. The typical situation is in parameter calibration for simulation models: there usually a complex simulation model is built, e.g. for traffic micro–simulation or for continuous simulation of, say, geophysical systems, which depends on some parameters. For urban traffic, as an example, parameters might be linked to some internal constants used in the simulation like, e.g., parameters which govern the drivers' behavior: average gap acceptance in crossing, average speed, ... Parameter calibration is usually performed by taking some measurements at some location in the real network, like average queue lengths, and trying to change parameter setting until the differences between real data observation and data obtained in the simulation match as closely as possible. In these cases, the objective to be minimized might be the sum of absolute or squared differences between real and simulated data. Indeed, in many such applications, what we actually wish to minimize is the quadratic error between the outcome of a simulator $\theta(\mathbf{x})$ and experimental data $\bar{\theta}$, i.e.

$$f(\mathbf{x}) = \|\theta(\mathbf{x}) - \bar{\theta}\|^2.$$

The analytical expression for this function is unknown and its evaluation requires running a lengthy simulation, possibly several times to get reliable output. However it can be clearly assumed that the lower bound on this function is zero.

Analogously, remaining within the traffic simulation example, after a model has been calibrated, it is usually employed to make some decisions; if these decisions are, as an example, traffic light phase durations at some intersections, and the objective is to regulate them so that the average queue lengths at specific points in the network are minimized (more precisely, their sum or their maximum is minimized), then, again, we are faced with a costly black box for which it is clear that a lower bound equal to zero is available.

This situation arises in so many practical applications that it is quite surprising to notice that, to the authors' knowledge, it did not attract specific research. This paper is meant to be a first step towards the inclusion of this information in a global optimization method.

The paper is organized as follows: at the end of this section a quick review of some similar approaches in the literature is presented; in Section 2 the basic global optimization methods based on radial basis interpolation models are quickly reviewed, mostly to introduce notation – most of the material is taken from [5] where a detailed analysis of these methods has been first proposed in a systematic

way. Our new technique is introduced in Section 3, while in Section 4 numerical results are provided. Conclusions and remarks are summarized in Section 5.

1.1 Brief literature review

The seminal work on the use of radial basis interpolant for the global optimization of black–box function is [5] in which the author introduce the use of the merit function and prove the convergence of the method under mild assumptions.

The use of a bumpiness measure is in our opinion an excellent strategy to guide the search towards new, promising points; however, also due to the non trivial implementation of bumpiness–based methods and to the requirement of running a global optimization method in order to optimize both the radial basis interpolant and the bumpiness at each iteration, some authors prefer to avoid the use of the bumpiness merit function, using alternative strategies.

In many applied fields a frequently used technique is kriging (see, e.g. [3]), a technique for interpolation which is very close to radial basis interpolation, although it is based on a stochastic model for the objective function; as with any interpolation or approximation scheme, models built according to the kriging methodology can be augmented by means of a merit function to guide the search for a good next evaluation point. Usually, however, in the kriging literature the standard approach of placing the next point at the global minimizer of the kriging model is adopted; in some cases models taking into account both the average and the standard deviation of the fitted model are exploited so that a balance between trusting the model (refinement) and looking where the variance is high (exploration) is obtained.

Recently some authors ([7,6,8,11]) proposed interesting extensions to the basic approach of Gutmann. These extension go into different directions like, e.g., incorporating a penalty term in a relaxed bumpiness minimization which allows for non exact interpolation (i.e., regression), or local refinements.

The possibility to handle constrained optimization problems has been exploited in [10]. A parallel implementation is presented in [12].

Apparently none of these approaches considered an extension like the one proposed in this paper and, on the other side, it can be easily checked that our modified method can be quickly adapted to all the cited modifications.

2 Global Optimization using Radial Basis surrogate models

Consider the problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

$$s.t.\ \mathbf{x} \in \mathscr{D} \subseteq \mathbb{R}^d$$

where $\mathscr{D}$ is a compact set. Assume that $f(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is an expensive function whose evaluation requires a significant amount of computational time and for which no analytic expression is available; assume also that f has been evaluated at k *pairwise distinct* points $\{\mathbf{x}^1, \ldots, \mathbf{x}^k \in \mathbb{R}^d\}$. At a generic iteration k define the sample set as the set of pairs $S^k = \{(\mathbf{x}^1, f(\mathbf{x}^1), (\mathbf{x}^2, f(\mathbf{x}^2), \ldots, (\mathbf{x}^k, f(\mathbf{x}^k))\}$. The

latter is the only information known at step k about the objective function to be minimized. The feasible set $\mathscr{D}$ is assumed to be analitically available and that its presence do not significantly increase the computational cost of optimization. In many applications, $\mathscr{D}$ turns out to be defined by simple bounds on each variable.

Global optimization methods based on surrogate functions require the definition

1. of a model of $f(\mathbf{x})$ based on the sample S^k
2. of a merit function based on such a model to be optimized in order to obtain the next evaluation point $\mathbf{x}^{k+1}$

If function evaluation, as we assume, is not too noisy, a surrogate model based on interpolation is usually employed; otherwise a regression model might be more reasonable. Here only the interpolation case will be developed, although the case of regression might be analyzed in quite a similar way. One of the best interpolation schemes for scattered multi-dimensional data is to define a surrogate model in terms of a linear combination of radial basis functions and a polynomial term:

$$s(\mathbf{x}|S^k) = \sum_{j=1}^{k} \lambda_j \phi(\|\mathbf{x} - \mathbf{x}^j\|) + \sum_{i=1}^{\hat{m}} c_i p_i(\mathbf{x}) \tag{1}$$

where $\phi(\cdot)$ is a suitable non negative function, $p_1(\cdot), \ldots, p_{\hat{m}}(\cdot)$ are a basis for the space of polynomials of a specific degree $\hat{m}$ and λ_j, c_i are coefficients.

In matrix form it is possible to write the above linear system as

$$\begin{aligned} \Phi\lambda + P\mathbf{c} &= \mathbf{f} \\ P^T\lambda \quad &= 0, \end{aligned} \tag{2}$$

where $\Phi_{ij} = \phi(\|\mathbf{x}^i - \mathbf{x}^j\|), P_{i\ell} = p_\ell(\mathbf{x}^i)$ and $\mathbf{f}_i = f(\mathbf{x}^i)$. Under suitable and quite weak assumptions on the degree of the polynomial (a minimum degree is required depending on the radial basis in use) and on the location of the points in the initial sample set, the linear system (2) can be shown to admit a unique solution. Indeed system (2) can be derived from the KKT conditions applied to the following optimization problem

$$\begin{aligned} \min_{\lambda,\mathbf{c}} \quad & \tfrac{1}{2}\lambda^T\Phi\lambda \\ \Phi\lambda + P\mathbf{c} &= \mathbf{f} \end{aligned} \tag{3}$$

Unique solution of problem (3) can be obtained enforcing $P^T\lambda = 0$, which, under suitable assumptions, ensure the conditional positive definiteness of Φ (see [5]).

It has been shown that function $\lambda^T\Phi\lambda$ can be used as a measure of bumpiness of the interpolating function; indeed, when restricted to the one-dimensional case using a cubic radial basis function $\phi(r) = r^3$ it can be shown that the radial basis interpolation coincides with the natural cubic spline interpolation, which is known to possess a minimal total curvature; moreover, in that case, the value of $\lambda^T\Phi\lambda$ is proportional to the total curvature of the spline.

2.1 A merit function to select new evaluation points

Once an interpolant $s(\mathbf{x}|S^k)$ has been formed, the task is how to choose the next point $\mathbf{x}^{k+1}$ in which $f(\mathbf{x})$ has to be evaluated. The common practice is to use a merit function $\mu(\mathbf{x})$ as a measure of the quality of new candidate points. A possible choice is to use the interpolant itself, i.e. $\mu(\mathbf{x}) \equiv s(\mathbf{x}|S^k)$; this can be indeed a good choice during the first steps of the method, but it is easily seen that using the interpolation itself as a merit function quickly leads to over-emphasize the initial sample and the consequence typically is that more and more observations are placed within a small neighborhood of the best sampled point.

An interesting possibility is that introduced in [5]; in that approach it is assumed that a threshold value τ is chosen as an aspiration level or an estimate of a likely value for the global minimum of the problem. Assuming that τ is a good estimate of the value at the global minimum, the idea of the method is to look for a point $\mathbf{x}$ at which it is in some sense more likely that $f(\mathbf{x}) = \tau$. The notion of likelihood introduced in [5] is not connected with probability, but is based on the bumpiness measure introduced in (3). If that measure is used, then a new point would be placed in such a way that the bumpiness of the interpolant corresponding to the extended sample $\{S^k \bigcup (\mathbf{x}, \tau)\}$ is minimized. After some computations, it turns out that minimizing the bumpiness reduces to minimizing the following function:

$$\mu(\mathbf{x}) = (-1)^{m_\phi} \frac{[\phi(0) - (\lambda, \mathbf{c})^T \begin{bmatrix} \Phi & P \\ P^T & 0 \end{bmatrix}^{-1} (\lambda, \mathbf{c})]}{(s(\mathbf{x}|S^k) - \tau)^2} \tag{4}$$

where $m_\phi \geq 1$ is an integer that depends on the radial basis in use(see [5]). The choice of the aspiration level is crucial to ensure the convergence to a global optimum and is particularly important for the efficiency of the method.

2.2 A Global Optimization algorithm for costly black–box functions

The overall algorithm for the global optimization of black–box expensive function is depicted in Algorithm 1. It is a kind of meta-algorithm, as many parts have to be defined in order to obtain an implementable method.

The convergence of Algorithm 1 can be proved (see [5]) provided that the choice of τ^k satisfy some quite mild assumption: basically it is sufficient to require that, infinitely often, τ_k is sufficiently lower than the minimum of the current interpolant model. In practice, τ^k is chosen by means of a cyclic rule which not only satisfy the convergence requirements, but also induces an alternation of refinement/exploration phases. In fact, a small value for τ^k (possibly even $-\infty$) produces a new observation which is as far as possible from sample points; on the contrary, choosing τ^k close to the minimum of $s(\cdot|S^k)$ is likely to generate a new iterate which is quite close to the currently best observation.

Moreover, as stated before, some conditions on the choice of the interpolation form and the initial sample are required in order to obtain a unique interpolant at every iteration; these conditions basically require that the polynomial added to the radial basis linear combination has a suitable degree depending on the basis

Algorithm 1: The GO algorithm

Input: a suitable initial sample S

1 $k \leftarrow 0$;

2 $S^0 \leftarrow S$;

3 **while** *stopping criterion* **do**

4 build $s(\cdot|S^k)$;

5 select τ^k;

6 let $\mathbf{x}^k \in \arg\min_{\mathbf{x}\in\mathscr{D}} \mu(\mathbf{x}, s(\mathbf{x}|S^k), \tau^k)$;

7 $S^{k+1} = S^k \cup (\mathbf{x}^k, f(\mathbf{x}^k))$;

8 $k \leftarrow k + 1$;

9 **end**

which is used, and that the initial sample forms a unisolvent set. For the commonly used choices of cubic or thin-plate spline bases, these requirements correspond to having a linear polynomial added and imposing that the initial sample contains $d + 1$ affinely independent points. If a Gaussian basis is chosen no polynomial is required and no condition is necessary for the initial sample.

It is important to notice that in step 6 of the above algorithm, a global minimization is required to find the optimum of the merit function; moreover, for methods based on the bumpiness measure, another global optimization is required at each iteration in which a finite threshold τ^k is chosen. In fact, in order to be able to guarantee that the next iterate will be distinct from the sample, it is required to chose τ^k strictly lower than the global minimum of the current interpolant. Thus one or two runs of a global optimization method are required at each iteration of the algorithm. Although these two optimization runs are expensive, it can be assumed that they are negligible when compared with the cost for function evaluation.

3 A new algorithm for the optimization of black-box functions with known lower bound

In this section we introduce some modifications to Algorithm 1 in order to take into account the availability of a lower-bound on the objective function.

We assume that a constant $\underline{f}$ is given such that

$$\min_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x}) \geq \underline{f}.$$

This assumption is fulfilled quite often in practice by the problems for which Algorithm 1 is designed. Our aim is two–fold: we would like to be able to build more accurate surrogate models and also to avoid selecting an unrealistic aspiration level τ.

For what concern the first point, given a surrogate model $s(\mathbf{x}|S)$ we would like to impose that the global minimum of the surrogate model is greater or equal of $\underline{f}$, i.e

$$\min_{\mathbf{x}\in\mathscr{D}} s(\mathbf{x}|S) \geq \underline{\mathrm{f}}. \tag{5}$$

Recalling that a global minimization of the interpolant is required at every step in order to choose a threshold τ which guarantees that the new iterate is different from previous ones, checking condition (5) becomes trivial, if we trust the global optimum found. At iteration k, let $\bar{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathscr{D}} s(\mathbf{x}|S^k)$; then if $s(\bar{\mathbf{x}}|S^k) \geq \underline{\mathrm{f}}$ nothing has to be done and we trust the interpolant satisfies the lower bound. Otherwise, we can temporarily include $\bar{\mathbf{x}}$ in the sample set, i.e. we let

$$\mathbf{x}^{k+1} = \bar{\mathbf{x}}$$

and we impose the a new interpolant is built such that at point $\bar{\mathbf{x}}$ the value of the interpolant is at least $\underline{\mathrm{f}}$. In order to do so the following optimization problem has to be solved:

$$\begin{aligned} \min_{\lambda,\mathbf{c},\xi}\ & \tfrac{1}{2}\lambda^T \Phi\lambda \\ & \Phi\lambda + P\mathbf{c} - \begin{bmatrix} 0 \\ \xi \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \underline{\mathrm{f}} \end{bmatrix} \\ & P^T\lambda = 0 \\ & \xi \geq 0 \end{aligned} \tag{6}$$

where ξ are slack variables.

As it can be easily seen, the constraints in the above problem allow for at least one solution, but, differently from pure interpolation problems, the solution is generally non unique. So the objective function is chosen in such a way that among all the interpolants of f at the sampled points which satisfy the lower bound condition at $\bar{\mathbf{x}}$, the one is chosen which is the least bumpy.

It can be shown, using the conditional semi–definiteness of Φ, that problem (6) is a quadratic convex problem that can be solved efficiently. Once a new surrogate model is built, then we can again check whether its global minimum is above $\underline{\mathrm{f}}$. If not, we can repeat the process, adding one more point and one more constraint to bound the interpolant. This process is repeated up to a certain number of times, after which the last interpolant built is taken as the new surrogate model. The overall scheme to build the surrogate model is depicted in algorithm 2.

In procedure 2, the statement "build $s(\mathbf{x}|S^k \cup S_f)$" corresponds to solving the quadratic convex problem (6) with exact interpolation imposed on all points in S^k and lower bounds imposed through slack variables to all points in S_f. It might be wirth observing that, even if the global optimum of the interpolant s is only imprecisely found, nevertheless the inequality introduced in (6) remains valid.

Once the surrogate model $s(\mathbf{x}|S^k \cup S_f)$ has been built, any suitable rule can be used to select the aspiration level τ^k based on $s(\mathbf{x}|S^k \cup S_f)$. In fact, as explained in Section 2, being τ^k a value we would like $f(\mathbf{x})$ to attain, unrealistic values should be avoided; on the other hand, to ensure convergence, it must hold $\tau^k < \min_{\mathbf{x}} s(\mathbf{x}|S^k \cup S_f)$. Thus, knowing that both the objective function and the surrogate model are bounded from below by $\underline{\mathrm{f}}$, we can avoid choosing $\tau^k < \underline{\mathrm{f}}$.

Once a new point $\mathbf{x}^{k+1}$ is selected and $f(\mathbf{x}^{k+1})$ evaluated, all auxiliary points are removed from the sample set and the algorithm is restarted with the new sample set.

Of course, if the bounding procedure on the interpolant fails, a threshold τ^k strictly lower than the minimum of the interpolant is chosen.

Algorithm 2: The procedure to augment the surrogate model

Input: $S, \bar{n}, \underline{f}$
$i = 0, S_f = \emptyset$;
while $i < \bar{n}$ **do**
  build $s(\mathbf{x}|S \cup S_f)$;
  $\bar{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathscr{D}} s(\mathbf{x}|S \cup S_f)$;
  if $f(\bar{\mathbf{x}}) > \underline{f}$ **then**
    return $s(\mathbf{x}|S \cup S_f)$
  $S_f = S_f \cup (\bar{\mathbf{x}}, f(\bar{\mathbf{x}}))$;
  $i = i + 1$;
end
return $s(x|S \cup S_f)$

4 Computational Results

The performance of Algorithm 1 extended by Algorithm 2 has been investigated running extensive tests. For all the tests we have performed, the global minimum $f^\star$ was a priori known, so that we could choose a lower bound of the form

$$\underline{f} = f^\star - \delta,$$

In the experiments, we choose $\delta \in \{0, 1, \infty\}$, i.e the tightest lower bound, a strict lower bound and, in the third case, no lower bound is used at all (in fact in this case the method is the standard one). The initial set of points have been generated using the improved Latin hypercube algorithm as in [9]. The total number of function evaluations allowed has been set to 200.

We used the cubic radial basis with a linear polynomial, although quite similar results were obtained using, e.g., the thin plate spline. For what concerns the rule used to update the aspiration level, following [5], at the $k-$th iteration we selected the aspiration level as

$$\tau^k = \begin{cases} -\infty & \text{if } k \mod 2 = 0 \\ \min_{\mathbf{x}\in\mathscr{D}} s(\mathbf{x}|S^k) - 0.1|\min_{\mathbf{x}\in\mathscr{D}} s(\mathbf{x}|S^k)| & \text{if } k \mod 2 = 1 \end{cases}$$

and then we apply our proposal which include taking into account the lower bound.

In practice, as pointed out in [6], it is often convenient to use as merit function the interpolant itself if

$$\min_{j=1,\dots,k} f_j - \min_{\mathbf{x}\in\mathscr{D}} s(x|S^k) > \Delta,$$

where $\Delta > 0$ is a parameter. The rationale is that if this condition holds the surrogate model can not be trusted enough to use the bumpiness.

For what concerns this updating rule, it was chosen just for the sake of simplicity: preliminary tests suggests that no significant differences in terms of performance comparison arise changing the radial basis or using a more complex strategy to update the threshold parameter.

Both the surrogate model and the bumpiness function were globally optimized running a simple Multistart scheme in which local searches were performed by the L–BFGS–B algorithm (see [1]); a total of 200 local search were performed. Problem (6) was solved using SNOPT 7.0 SQP solver (see [4]).

The first test suite, listed in Table 1, is based on classic test function which are commonly used to compare the performance of black-box optimization algorithms. For this class of problems, we generated 100 different starting set of points and then ran the proposed algorithm. Table 1 reports some information on the test set.

name	# variables	domain	#global optima	$f^\star$
Branin	2	$[-5, 10] \times [0, 15]$	3	0.397887
Hartman3	3	$[0, 1]^3$	2	-3.86278
Hartman6	6	$[0, 1]^6$	2	-3.32237
Goldstein-Price	2	$[-2, 2]^2$	1	0
Shekel5	4	$[0, 10]^4$	1	-10.1532
Shekel7	4	$[0, 10]^4$	1	-10.4029
Shekel10	4	$[0, 10]^4$	1	-10.5364

Table 1 Standard test functions.

In Figure 1 we report the percentage of instances in which each of the three variants of the method found the best objective function value with respect to the other methods. The apparently less significative results obtained for the Shekel10 test function might be due to the higher dimension of this problem. Indeed, as it has been observed by many authors, many methods based on the bumpiness function tend, in the first iterations, to place observations at the vertices of the feasible box; it is thus likely that 200 iterations are too few, for methods of this kind, to display significative results.

A second, much larger, test suite has been generated using the Shepard's function as defined in equation (7). Shepard functions (see [13]) allow to select not only the problem dimension but also a set of parameters that can tune function shape. Given a set of points $\{z_j, j = 1, \ldots, l\}$, a set of corresponding function values $\{f_1, \ldots, f_l\}$ and a set of reals $\{\alpha_j, j = 1, \ldots, l\}$, then the function is defined as

$$f(x) = \frac{\sum_{i=1}^{l} f_i \prod_{j \neq i} \|x - z_j\|^{\alpha_j}}{\sum_{i=1}^{l} \prod_{j \neq i} \|x - z_j\|^{\alpha_j}} \tag{7}$$

It can be shown that stationary points are located at points $\{x_j\}$ and that the minimum and maximum of the set $\{f_j\}$ are the global minimum and maximum value for the function. Moreover, if $\alpha_j \geq 1$, the function is continuously differentiable.

In the experiments, we generated random instances of Shepard functions in the following way:

- $d \in \{2, 5\}$;
- $l \in \{20, 50, 100\}$;
- the l stationary points $z_j, j = 1, \ldots, l$ were uniformly generated in the box $[0, 1]^d$;
- each function value $f(z_i)$ was uniformly generated in $[0, 10], [0, 100], [0, 1000]$ respectively, except $f(z_1)$ which was chosen to be 0, i.e. z_1 is the global minimum of $f(x)$;

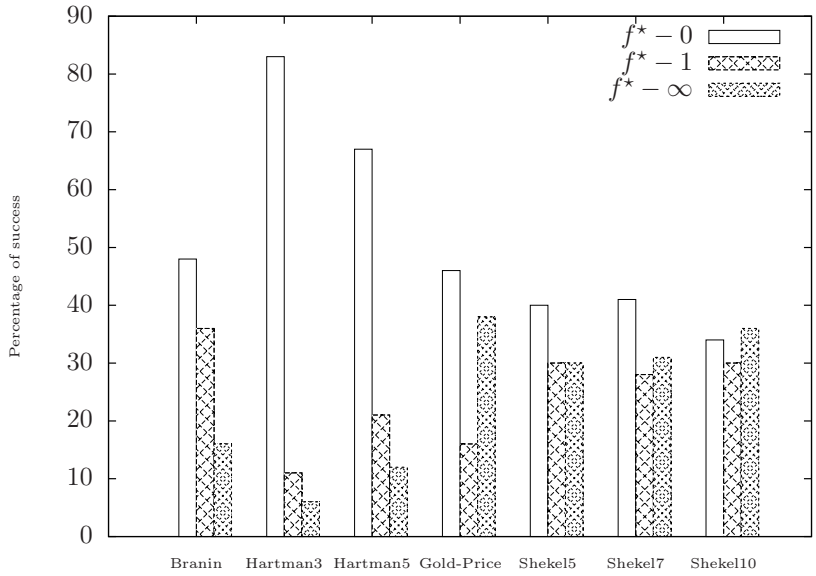

Fig. 1 Percentage of instances in which, respectively, the method based on the exact lower bound, the one based on a strict lower bound and the original one with no lower bounds found a better function value than the others.

– α_j uniformly chosen in $[2, 3]$;

For each combination of problem dimension, number of stationary points and their range, 100 random instances have been generated, obtaining this way a set of $2 \times 3 \times 3 \times 100 = 1\,800$ test functions.

For each test function the 3 different variants of the algorithm were run 10 times, accounting for a total of 54 000 runs of the global optimization method. Such a large computational test could be performed in a reasonable CPU time thanks to the fact that these functions are indeed quite cheap to evaluate. It goes without saying that this does not affect the basis of the method, which is indeed designed for expensive problems.

To visualize and compare so many tests we use the performance profiles introduced in [2]. In Figures 2 and 3 we report the performance profile of the best function value among all instances for the two-dimensional case after 50 and 200 function evaluations respectively. The results of the tests on the five dimensional problem instances, after 200 function evaluations, are summarized in Figure 4.

In these figures it can be easily seen that not only the new method is capable of finding better function values more frequently than the standard one, but also that this improved behavior is robust, as it can be seen from the fact that the curves relative to the new method remain always above that of the standard one. Moreover, while choosing the best lower bound usually helps, even with a strict underestimation the performance is still very good.

More in detail, we summarize in Table 2 and 4, for the 2D and 5D instances respectively, the results for each combination of number of stationary points and their range. In order to better appreciate the above statistics, it is worth to report some basic facts about how often the procedure introduced in this paper

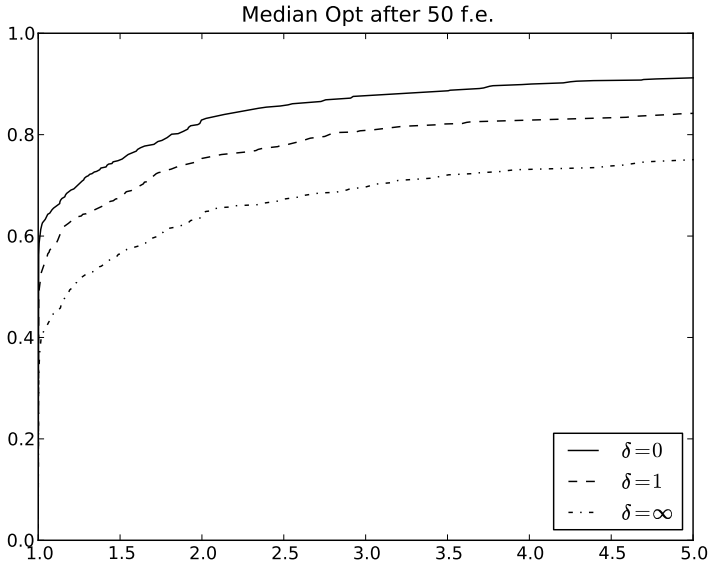

Fig. 2 Performance profile for the 2D instances: the median of the best function value after 50 function evaluations is reported.

$f(x)$ range	# of stationary points			total
	20	50	100	
$[0, 10]$	64/19/17	44/35/21	45/28/27	153/82/65
$[0, 100]$	47/40/13	43/41/16	40/29/31	130/110/60
$[0, 1000]$	38/51/11	40/39/21	33/38/29	111/128/61
total	149/110/41	127/115/58	118/95/87	394/320/186

Table 2 Number of 2D instances with the best final objective function value, after 50 function evaluations, considering different number of stationary points and their range (with $\delta = \{0, -1, \infty\}$).

was indeed used during numerical experiments. Tables 5-6 report, for the test at dimension 2 and 5 respectively, and for each combination of parameters for the generation of test functions, the average number of times the procedure for the addition of inequalities in (6) was called at least once.

For what concerns the failure in (6) to generate an interpolant which satisfies the lower bound even after as many as 200 additional points were added, this is quite a rare event. In all dimension 2 test cases the median number of cases in which such a failure occurred is 0 when the exact threshold is chosen and it is one when the strict lower bound was employed. For 5–dimensional test cases, the

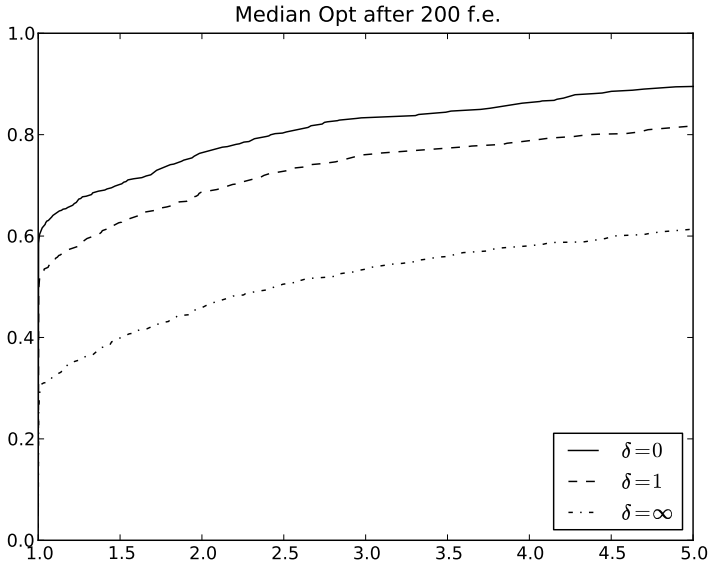

Fig. 3 Performance profile for the 2D instances: the median of the best function value after 200 function evaluations is reported.

$f(x)$ range	# of stationary points			total
	20	50	100	
$[0, 10]$	67/22/11	70/12/18	60/14/26	197/48/55
$[0, 100]$	43/49/8	54/36/10	40/48/12	137/133/30
$[0, 1000]$	34/50/16	36/49/15	35/53/12	105/152/43
total	144/121/35	160/97/43	135/115/50	439/333/128

Table 3 Number of 2D instances with the best final objective function value, after 200 function evaluations, considering different number of stationary points and their range (with $\delta = \{0, -1, \infty\}$).

median was 1 both when using the exact lower bound and when using the struck one.

5 Conclusions

In this paper we introduced a modification of the algorithm proposed by Gutmann in [5] for the global optimization of costly black–box functions. It is of interest to consider the availability of a lower bound on the objective function (situation that arises for instance in the parameter estimation problem): this information

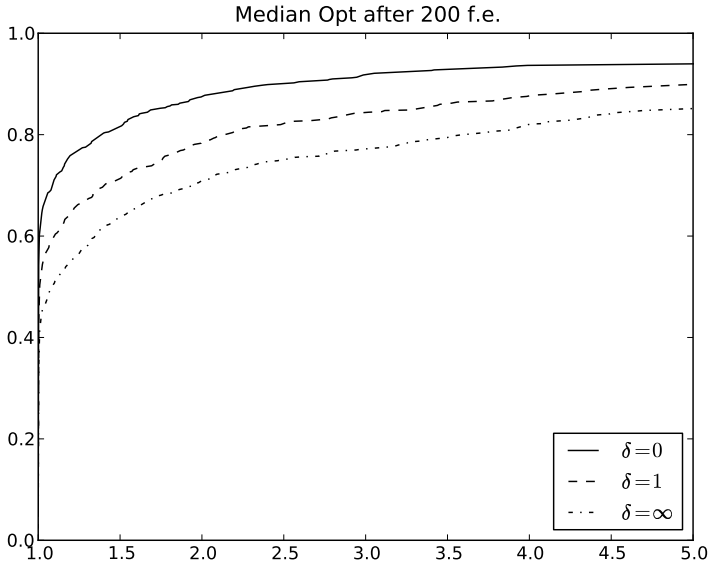

Fig. 4 Performance profile for 5D problem instances: the median of the best function value after 200 function evaluations is reported.

$f(x)$ range	# of stationary points			total
	20	50	100	
$[0, 10]$	71/15/14	44/30/26	35/35/30	150/80/70
$[0, 100]$	44/40/16	44/25/31	41/32/27	129/97/74
$[0, 1000]$	43/37/20	44/31/25	35/40/25	122/108/70
total	158/92/50	132/86/82	111/107/82	401/285/214

Table 4 Number of 5D instances with the best final objective function value, after 200 function evaluations, considering different number of stationary points and their range (with $\delta = \{0, -1, \infty\}$).

is used to improve the aspiration level selection and to improve the quality of the surrogate model including possibly artificial function observations. Extensive numerical results have shown that the proposed method is capable of significantly improving the performance of the original algorithm on a large class of small dimensional, yet challenging, test problems.

	$\delta = 0$		
$f(x)$ range	# of stationary points		
	20	50	100
$[0, 10]$	19.8	6.4	8.8
$[0, 100]$	9.6	12.0	13.7
$[0, 1000]$	68.0	41.0	22.8
total	18.47		
	$\delta = 1$		
$f(x)$ range	# of stationary points		
	20	50	100
$[0, 10]$	3.46	1.58	2.15
$[0, 100]$	2.29	3.22	4.67
$[0, 1000]$	7.50	5.61	7.19
total	3.61		

Table 5 Average number of times at least one inequality constraint was added (2D test cases).

	$\delta = 0$		
$f(x)$ range	# of stationary points		
	20	50	100
$[0, 10]$	3.45	3.61	4.26
$[0, 100]$	8.13	4.68	3.66
$[0, 1000]$	0.50	2.07	1.34
total	4.47		
	$\delta = 1$		
$f(x)$ range	# of stationary points		
	20	50	100
$[0, 10]$	1.07	0.68	0.64
$[0, 100]$	1.37	1.06	1.05
$[0, 1000]$	0	1.46	1.10
total	1.15		

Table 6 Average number of times at least one inequality constraint was added (5D test cases).

References

1. Byrd, R.H., Nocedal, J., Lu, P., Zhu, C.: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. Tech. rep., Northwestern University, Department of Electrical Engineering and Computer Science (1995)
2. Dolan, E.D., Mor, J.J.: Benchmarking optimization software with performance profiles. Mathematical Programming **91**, 201–213 (2002)
3. Forrester, A.I.J., Keane, A.J.: Recent advances in surrogate-based optimization. Progress in Aerospace Sciences **45**(1-3), 50 – 79 (2009). DOI doi:10.1016/j.paerosci.2008.11.001
4. Gill, P.E., Murray, W., Saunders, M.A.: User's guide for SNOPT version 7: Software for large-scale nonlinear programming. Systems Optimization Laboratory - University of Stanford (2007)
5. Gutmann, H.M.: A radial basis function method for global optimization. Journal of Global Optimization **19**, 201–227 (2001)
6. Holmström, K.: An adaptive radial basis algorithm (arbf) for expensive black-box global optimization. Journal of Global Optimization **41**(3), 447–464 (2008). DOI 10.1007/s10898-007-9256-8
7. Jakobsson, S., Patriksson, M., Rudholm, J., Wojciechowski, A.: A method for simulation based optimization using radial basis functions. Optimization and Engineering p. on line (2010). DOI 10.1007/s11081-009-9087-1
8. Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization **21**(4), 345–383 (2001)
9. McKay, M., Beckman, R., Conover, W.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **42**(1), 55–61 (2000)

10. Regis, R.G., Shoemaker, C.A.: Constrained global optimization of expensive black box functions using radial basis functions. Journal of Global Optimization **31**, 153–171 (2005)
11. Regis, R.G., Shoemaker, C.A.: Improved strategies for radial basis function methods for global optimization. Journal of Global Optimization **37**(1), 113–135 (2007)
12. Regis, R.G., Shoemaker, C.A.: Parallel radial basis function methods for the global optimization of expensive functions. European Journal of Operational Research **182**(2), 514 – 535 (2007). DOI doi:10.1016/j.ejor.2006.08.040
13. Schoen, F.: A wide class of test functions for global optimization. Journal of Global Optimization **3**, 133–138 (1993)