A Modified DIviding RECTangles Algorithm for a Problem in Astrophysics*

D. di Serafino[†], G. Liuzzi[‡], V. Piccialli[§], F. Riccio[†], G. Toraldo[¶]

Communicated by Ryan P. Russell

Abstract

We present a modification of the DIRECT (DIviding RECTangles) algorithm, called DIRECT-G, to solve a box-constrained global optimization problem arising in the detection of gravitational waves emitted by coalescing binary systems of compact objects. This is a hard problem, since the objective function is highly nonlinear and expensive to evaluate, has a huge number of local extrema and unavailable derivatives. DIRECT performs a sampling of the feasible domain over a set of points, that becomes dense in the limit, thus ensuring the everywhere dense convergence; however, it becomes ineffective on significant instances of the problem under consideration, because it tends to produce a uniform coverage of the feasible domain, by oversampling regions that are far from the optimal solution. DIRECT has been modified by embodying information provided by a suitable discretization of the feasible domain, based on the signal theory, which takes into account the variability of the objective function. Numerical experiments show that DIRECT-G largely outperforms DIRECT and the grid search, the latter being the reference algorithm in the astrophysics community. Furthermore, DIRECT-G is comparable with a genetic algorithm specifically developed for the problem. However, DIRECT-G inherits the convergence properties of DIRECT, whereas the genetic algorithm has no guarantee of convergence.

Keywords: global optimization, DIRECT algorithm, detection of gravitational waves.

2010 Mathematics Subject Classification: 65K05, 90C26, 8508.

1 Introduction

DIRECT (DIviding RECTangles) is a deterministic algorithm for box-constrained global optimization problems with Lipschitz-continuous objective function [1, 2, 3]. It is a modification of standard Lipschitzian approaches, that eliminates the need to specify a Lipschitz

*We thank Stefano Lucidi for useful discussions on DIRECT and on strategies to accelerate its convergence. We are also thankful to the anonymous Referees and the journal Editors for their careful reading of the paper and their insightful comments, that helped us improving the paper. This work was supported by the Italian Ministry for Education, University and Research, under the PRIN Project *Nonlinear Optimization, Variational Inequalities and Equilibrium Problems.*

[†]Dipartimento di Matematica, Seconda Università degli Studi di Napoli, via Vivaldi 43, 81100 Caserta, Italy, {daniela.diserafino,filippo.riccio}@unina2.it

[‡]Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", CNR, viale Manzoni 30, 00185 Roma, Italy, giampaolo.liuzzi@iasi.cnr.it

[§]Dipartimento di Informatica, Sistemi e Produzione, Università degli Studi di Roma "Tor Vergata", viale del Politecnico 1, 00133 Roma, Italy, piccialli@disp.uniroma2.it

[¶]Dipartimento di Ingegneria Agraria e Agronomia del Territorio, Università degli Studi di Napoli "Federico II", via Università 100, 80055 Portici (NA), Italy, toraldo@unina.it (corresponding author)

constant. DIRECT does not exploit any a-priori knowledge on the objective function, including first derivative information, so that it is suitable for black-box optimization [4, 5, 6, 7]. The idea is to carry out simultaneous searches using all possible constants from zero to infinity and to regard the Lipschitz constant as a weighting parameter to balance global and local search. DIRECT performs a sampling of the feasible domain on a set of points, that becomes dense in the limit, thus guaranteeing strong theoretical convergence properties [1]. On the other hand, these strong properties are balanced by a quite slow convergence. To deal with such drawback, different strategies have been tried. In [8, 9] some modifications of the original version of DIRECT have been studied, which try to exploit either a priori information on the objective function or information gathered during the algorithm. Furthermore, in [10, 11] local searches have been introduced within the DIRECT framework in an attempt to speed up the convergence.

We investigate the application of DIRECT to a global optimization problem, that arises in the detection of Gravitational Waves (GWs). The detection of GWs is a very ambitious goal in modern astrophysics, because it will provide a validation of the General Relativity Theory of Einstein and a new type of information on the Universe [12]. Networks of detectors have been deployed, but GWs have not yet been observed because of many difficulties in the detection process. Such difficulties are related to the weakness of the GWs, the rarity of the events that generate them, and the instrumental noise of the detectors. For these reasons, sophisticated data analysis techniques are required. Furthermore, because of the huge quantity of data acquired by the detectors twenty-four hours a day, a real-time on-line data analysis is needed.

We focus on the detection of GWs emitted by coalescing binary systems, which are very promising sources. The technique generally used for this problem is the matched filter, which exploits the waveform of the gravitational signal and assumes that the instrumental noise be a stationary Gaussian stochastic process. The computational kernel of the matched filter is the solution of a hard box-constrained global optimization problem, in which the objective function is a stochastic process because of the presence of noise. In practice, the output data of a detector contain a sample of noise, thus the optimization problem corresponding to that sample must be solved. The objective function is highly nonlinear, with unavailable derivatives and many local maxima; furthermore, its evaluation has non-negligible computational cost.

The method most widely used by the astrophysicists to solve this global optimization problem is the grid search, which evaluates the objective function over a suitable discretization of the feasible domain [13]. The main motivation for using the grid search is that it provides information on the accuracy of the solution; on the other hand, it needs a large number of objective function evaluations and hence has a high computational cost. In order to reduce this cost, in [14, 15] a real-coded genetic algorithm specifically tailored to the problem was proposed, which is able to compute reasonably accurate solutions with a much smaller number of function evaluations. Although this algorithm has demonstrated its effectiveness on significant test problems, its convergence to the solution is not guaranteed. In this work, we present a modified version of DIRECT, which exploits the feasible domain discretization of the grid search to significantly reduce the number of objective function evaluations, while retaining the theoretical convergence properties of DIRECT.

The paper is organized as follows. A description of the GWs detection problem is given in Section 2. The grid search procedure is outlined in Section 3. The modification of DIRECT is presented in Section 4 and the results of its application to different test problems are analysed in Section 5. Finally, some conclusions are drawn in Section 6.

2 A Hard Global Optimization Problem in the Detection of Gravitational Waves

The problem of detecting GWs consists in deciding if the output of a detector contains a gravitational signal or it is made of noise only. This output is generally modeled as

$$x(t) := r(t) + h(t; \boldsymbol{\theta}),$$

where t is the time, $r(t)$ is the noise, $h(t; \boldsymbol{\theta})$ is the gravitational signal, and $\boldsymbol{\theta}$ is a vector of parameters. We assume that $r(t)$ be strictly white noise, i.e. a wide-sense stationary Gaussian stochastic process with mean 0 and variance 1. We are interested in gravitational signals emitted by coalescing binary systems of compact objects (neutron stars and/or black holes), which are one of the most promising sources for ground-based laser interferometric detectors. In this case, a relative large number of events per year is expected (tens per year within a few hundred Mpc) [16], and a model of the emitted waves is available, i.e. the so-called *chirp* signal in which both the frequency and the amplitude increase with time until the two objects of the system are close enough to merge [17]. The corresponding vector of parameters is $\boldsymbol{\theta} := (A, \varphi_0, t_0, m_1, m_2)$, where A is the amplitude, φ_0 the initial phase, and t_0 the arrival time of the signal, while m_1 and m_2 are the masses of the binary system; all of them are unknown. In practice, the output of the detector is sampled with a certain time step, thus a segment of data is available, which is an N-dimensional vector $\boldsymbol{x} := (x[0], \ldots, x[N-1])$; the corresponding sampled gravitational signal, if present, is an M-dimensional vector $\boldsymbol{h} := (h[0], \ldots, h[M-1])$, with $M < N$ (the dependence on $\boldsymbol{\theta}$ has been neglected for simplicity).

As noted in Section 1, the most widely used detection technique is the matched filter, which is an optimal linear filter for detecting signals of known shape in stationary Gaussian noise [13]. It basically consists of three steps:

1. correlating the output of the detector with a family of templates, consisting of chirp signals $\boldsymbol{h}(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ varying in a suitable set;

2. finding the maximum of the correlation with respect to all the parameters;

3. comparing this maximum with a suitable threshold to decide if the output of the detector contains a gravitational signal (a detection is announced if the maximum exceeds the threshold).

This procedure is based on the observation that the highest Signal-to-Noise Ratio (SNR) of the filter output is achieved when the values of the parameters identifying the template are the same as in the signal, and that this SNR is equal to the maximum of the mean value of the correlation [18]. The choice of the threshold is related to the probability of false alarm, i.e. of stating a detection in the absence of a signal, and to the probability of detection, i.e. of stating a detection when the output contains a gravitational signal [18].

We note that the amplitude and the initial phase corresponding to the maximum in step 2 can be expressed in terms of the optimal values of the remaining parameters; therefore the maximization of the correlation must be actually performed with respect to the masses (m_1, m_2) and the index $n_0 \in I = \{0, \ldots, N-M\}$ of the sample corresponding to the arrival time of the signal. Furthermore, given (m_1, m_2), the maximum with respect to n_0 can be easily computed [13]. Thus, step 2 can be formulated as the following box-constrained global optimization problem:

$$\underset{(m_1, m_2) \in \Omega}{\text{maximize}}\, F(m_1, m_2), \tag{1}$$

with

$$\Omega := \{(m_1, m_2) \in \Re^2 \ : \ l \leq m_1, m_2 \leq u\}$$

and

$$F(m_1, m_2) := \sqrt{\max_{n_0 \in I} \left(C_0^2(n_0, m_1, m_2) + C_{\pi/2}^2(n_0, m_1, m_2) \right)},$$

where $C_0(n_0, m_1, m_2)$ and $C_{\pi/2}(n_0, m_1, m_2)$ are the correlations between $\boldsymbol{x}$ and the so-called normalized quadrature components of the template, $\hat{\boldsymbol{h}}_0(m_1, m_2)$ and $\hat{\boldsymbol{h}}_{\pi/2}(m_1, m_2)$ [18], i.e.

$$C_0(n_0) := \sum_{k=n_0}^{n_0+M-1} x[k]\hat{h}_0[k-n_0] \qquad C_{\pi/2}(n_0) := \sum_{k=n_0}^{n_0+M-1} x[k]\hat{h}_{\pi/2}[k-n_0],$$

where the dependence on m_1 and m_2 has been neglected for simplicity of notations. We note that computing the maximum of F with accuracy is crucial in the detection process, because a large error may correspond to a value of F which is under the threshold, thus preventing the detection of the signal [19].

The solution of problem (1) is a difficult task, because the objective function F is highly nonlinear, has many local maxima, and its derivatives are not available (see Figure 1). Furthermore, the evaluation of F is quite expensive, since it requires the solution of two ordinary differential equations (ODEs) to generate the quadrature components of each template [20], and the execution of three FFTs of length N to compute the correlations of $\boldsymbol{x}$ with them [18]. Common values of N are $\mathrm{O}(10^5)$; the time for solving the ODEs depends on the values of the masses, is highly variable and increases as the masses decrease.

3 Grid Search

The reference algorithm for the solution of problem (1) is the grid search, which performs an exhaustive search on a suitable discretization of the feasible domain Ω. This algorithm provides information on the accuracy of the computed maximum in a probabilistic sense, as explained next.

For each pair of masses (m_1, m_2), $F(m_1, m_2)$ is a random variable, because of the presence of noise. If we assume that the gravitational signal $h(t; \bar{\boldsymbol{\theta}})$ contained in the output of the detector has the parameter values $\bar{\boldsymbol{\theta}} = (\bar{A}, \bar{\varphi}_0, \bar{t}_0, \bar{m}_1, \bar{m}_2)$, then the mean value of $F(m_1, m_2)$ is

$$\mathcal{F}(m_1, m_2; \bar{m}_1, \bar{m}_2) := E[F(m_1, m_2)] = \sqrt{\max_{n_0 \in \{0,\ldots,N-M\}} (\bar{C}_0^2 + \bar{C}_{\pi/2}^2)},$$

where $\bar{C}_0$ and $\bar{C}_{\pi/2}$ are the discrete correlations between the sampled version of the gravitational signal, $h(t, \bar{\boldsymbol{\theta}})$, and the normalized quadrature components of the template, $\hat{\boldsymbol{h}}_0(m_1, m_2)$ and $\hat{\boldsymbol{h}}_{\pi/2}(m_1, m_2)$. As previously noted, from the matched filter technique it follows that [18]

$$\max_{m_1, m_2 \in \Omega} \mathcal{F}(m_1, m_2; \bar{m}_1, \bar{m}_2) = \mathcal{F}(\bar{m}_1, \bar{m}_2; \bar{m}_1, \bar{m}_2) = \bar{\mathcal{S}},$$

where $\bar{\mathcal{S}}$ is the maximum SNR. In other words, the maximum of $\mathcal{F}$ is achieved when the masses of the template are equal to the masses of the gravitational signal. We observe also that $\bar{\mathcal{S}}$ concides with the amplitude $\bar{A}$ of the gravitational signal in the output of the detector [18].

Of course, the values of the parameters of the gravitational signal are unknown, but it is possible to build a finite set $G \subset \Omega$ (the *grid*) such that, for any vector

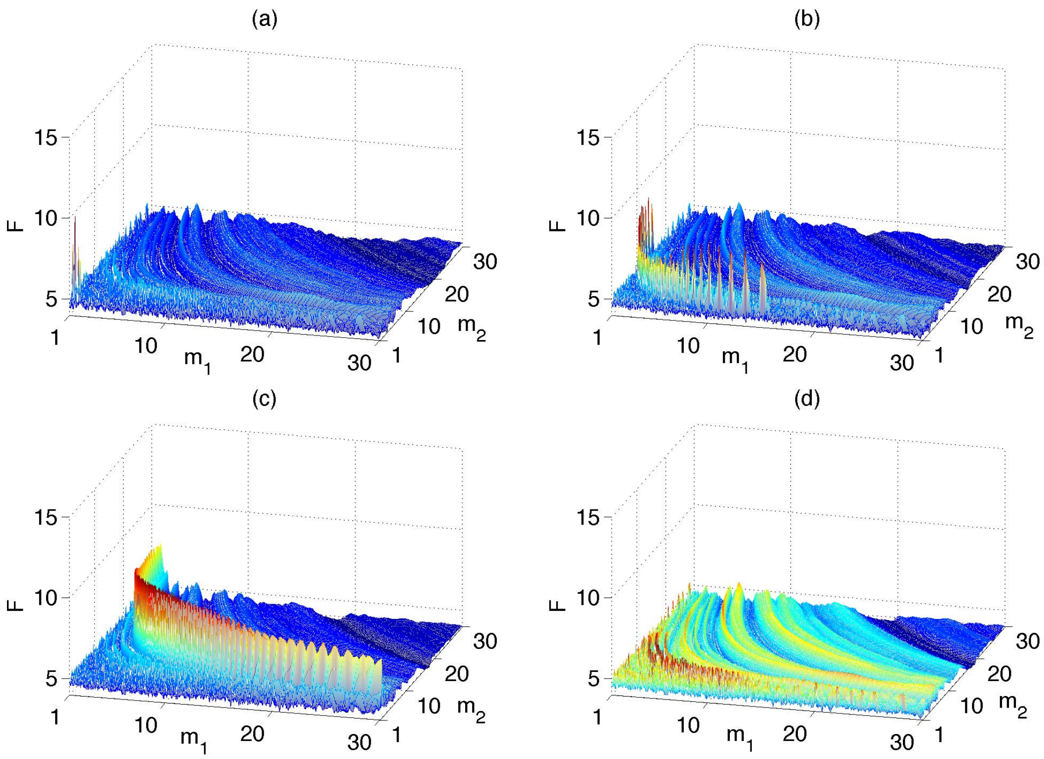

Figure 1: 3D plot of the objective function F, corresponding to noise plus gravitational signal from different pairs of masses $(\bar{m}_1, \bar{m}_2)$, or to noise only: (a) $\bar{m}_1 = \bar{m}_2 = 1.4\ M_\odot$, (b) $\bar{m}_1 = 1.4\ M_\odot$, $\bar{m}_2 = 10\ M_\odot$, (c) $\bar{m}_1 = 5\ M_\odot$, $\bar{m}_2 = 10\ M_\odot$, (d) noise only. $M_\odot$ is the solar mass; the SNR is equal to 10.

$\bar{\boldsymbol{\theta}} = (\bar{A}, \bar{m}_1, \bar{m}_2, \bar{\varphi}_0, \bar{t}_0)$ corresponding to a gravitational signal in the output of the detector, the maximum of $\mathcal{F}$ over G is greater than a fixed percentage $MM \in [0, 1]$ of $\bar{\mathcal{S}}$ (*minimal match*) [13, 21], i.e.

$$\max_{m_1, m_2 \in G} \mathcal{F}(m_1, m_2; \bar{m}_1, \bar{m}_2) \geq MM \cdot \bar{\mathcal{S}}. \tag{2}$$

This is achieved by determining each point of the grid in such a way that the corresponding value of $\mathcal{F}$ is representative of a region of the feasible domain, and the set of all such regions is a covering of this domain. More precisely, by using a second-order Taylor series expansion of $\mathcal{F}$ around the maximum point $\bar{\boldsymbol{m}} = (\bar{m}_1, \bar{m}_2)$,

$$\mathcal{F}(m_1, m_2; \bar{m}_1, \bar{m}_2) \approx \mathcal{F}(\bar{m}_1, \bar{m}_2; \bar{m}_1, \bar{m}_2) + \frac{1}{2} \Delta \boldsymbol{m}^T H_{\mathcal{F}}(\bar{\boldsymbol{m}}) \Delta \boldsymbol{m},$$

where $H_{\mathcal{F}}(\bar{\boldsymbol{m}})$ is the Hessian of $\mathcal{F}$ in $\bar{\boldsymbol{m}}$ and $\Delta \boldsymbol{m} := \boldsymbol{m} - \bar{\boldsymbol{m}}$, we can define a positive-definite metric tensor $T(\bar{\boldsymbol{m}}) := -H_{\mathcal{F}}(\bar{\boldsymbol{m}})/(2\bar{\mathcal{S}})$ and express the variability of the function $\mathcal{F}$ as

$$M(m_1, m_2; \bar{m}_1, \bar{m}_2) := \frac{\bar{\mathcal{S}} - \mathcal{F}(m_1, m_2; \bar{m}_1, \bar{m}_2)}{\bar{\mathcal{S}}} = \Delta \boldsymbol{m}^T T(\bar{\boldsymbol{m}}) \Delta \boldsymbol{m}.$$

We note that $T(\bar{\boldsymbol{m}})$ does not depend on $\bar{\mathcal{S}}$, since $H_{\mathcal{F}}(\bar{\boldsymbol{m}}) = \bar{\mathcal{S}} H_{\mathcal{L}}(\bar{\boldsymbol{m}})$, where $H_{\mathcal{L}}(\bar{\boldsymbol{m}})$ is the Hessian in $\bar{\boldsymbol{m}}$ of a function $\mathcal{L}(m_1, m_2; \bar{m}_1, \bar{m}_2)$ independent of $\bar{\mathcal{S}}$ [19].

The *mismatch* function M represents the SNR fractional loss when the function $\mathcal{F}$ is evaluated at the point $\boldsymbol{m} = (m_1, m_2)$ and the masses of the gravitational signal in the output of the detector are $\bar{\boldsymbol{m}} = (\bar{m}_1, \bar{m}_2)$. In terms of the mismatch function, inequality (2) can be written as

$$\max_{m_1, m_2 \in G} M(m_1, m_2; \bar{m}_1, \bar{m}_2) \leq 1 - MM, \tag{3}$$

i.e. the relative error in the computed SNR is lower than a given percentage $1 - MM$. Then, for any point $\boldsymbol{m} = (m_1, m_2) \in \Omega$, we can define

$$B_{\boldsymbol{m}} := \{\bar{\boldsymbol{m}} \in \Omega \mid \Delta \bar{\boldsymbol{m}}^T T(\boldsymbol{m}) \Delta \bar{\boldsymbol{m}} \leq 1 - MM, \ \Delta \bar{\boldsymbol{m}} = \bar{\boldsymbol{m}} - \boldsymbol{m}\},$$

which is a ball of radius $R = \sqrt{1 - MM}$ in the manifold (Ω, T), and we can cover the feasible domain Ω with a fixed number of $B_{\boldsymbol{m}}$'s in such a way that their centers satisfy (3). These centers are taken as the points of the grid G.

We note that covering a manifold with balls is generally a difficult problem, but in this case it can be simplified by using a nonlinear map that transforms the space of masses (m_1, m_2) into the so-called space of chirp times (τ_0, τ_3). In the latter space, the centers form a rectangular lattice which is easier to build [13]. We note also that in the space of chirp times the feasible domain is a nonconvex set and the unfeasible points may have no physical meaning, therefore in this work we consider the feasible domain in the space of masses. Examples of grids corresponding to the two spaces are shown in Figure 2; since the objective function is symmetric with respect to m_1 and m_2, the grids have been built only in the subdomain where $m_1 > m_2$ and in the corresponding subdomain of the space of chirp times. We note that in the space of masses the grid is highly nonuniform, with more points in the regions where the objective function may have greater variability.

4 A Modified DIRECT Algorithm

DIRECT is a global optimization algorithm for box-constrained problems with Lipschitz-continuous objective function. It is based on a space-partitioning scheme, designed to

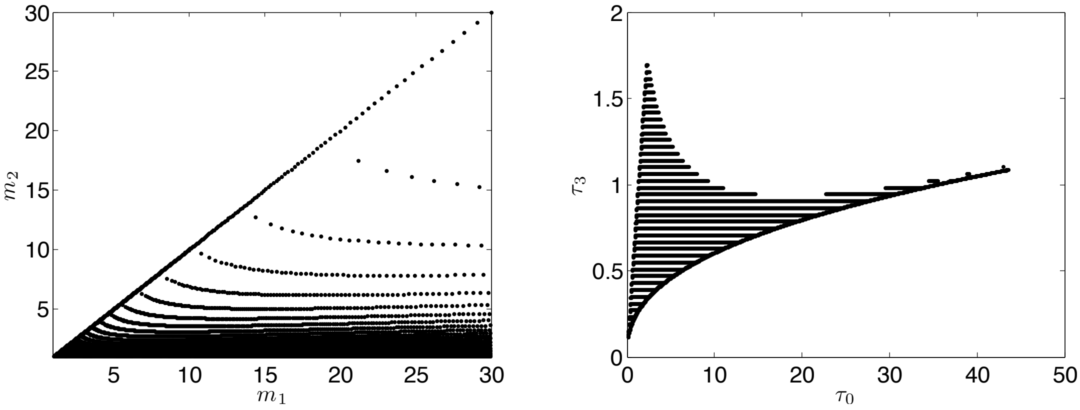

Figure 2: A grid in the space of the masses (left) and in the space of the chirp times (right).

automatically balance local and global search at each iteration. We give a brief description of DIRECT, to better explain our modification of it. We refer to the problem

$$\underset{x \in D}{\text{minimize}}\, f(x), \tag{4}$$

where $D = \{x \in \Re^n : 0 \le x_i \le 1, i = 1, \ldots, n\}$, to which every box-constrained problem can be reduced. For a detailed presentation of the algorithm and for thorough convergence analysis, the reader is referred to [1, 2, 10].

At the first step of DIRECT, $f(x)$ is evaluated at the center of D; the hypercube is then partitioned into a set of smaller hyperrectangles and $f(x)$ is evaluated at their centers. At the generic $k-$th iteration, a partition $\mathcal{H}_k$ of D into hyperrectangles is built, by subdividing a set of *potentially optimal* hyperrectangles of the previous partition $\mathcal{H}_{k-1}$. The identification of a potentially optimal hyperectangle is based on some measure of the hyperrectangle itself and on the value of f at its center. The refinement of the partition continues until a prescribed number of function evaluations has been performed, or another stopping criterion is satisfied. The minimum of f over all the centers of the final partition, and the corresponding centers, provide an approximate solution to the problem. The structure of DIRECT is outlined in Figure 3.

The DIRECT algorithm is specified once the following issues are defined:

- the way to measure the hyperrectangles (size);
- the way to select the potentially optimal hyperrectangles (selection strategy);
- the way to divide the hyperrectangles (partitioning strategy).

In the original version of the algorithm, the size of a hyperrectangle is set as the distance d from its center to a vertex, i.e. half the diameter of the hyperrectangle [1]. Different choices have been also considered, see, e.g., [8, 9] and the references therein.

Concerning the partition strategy, DIRECT divides a potentially optimal interval by trisection along the directions of the longest edges, as shown in Figure 4. A rule based on the objective function value is proposed in [1] to choose among multiple longest sides; this rule ensures that smaller function values are in larger hyperectangles, which, according to the selection strategy described below, are more likely to be divided at the next iterations.

$\mathcal{H}_1 = \{D\}$, $c =$ center of D, $f_{min} = f(c)$, $X_{min} = \{c\}$, $k = 1$

repeat

identify the potentially optimal hyperrectangles in $\mathcal{H}_k$

subdivide the potentially optimal hyperrectangles to build a new partition $\mathcal{H}_{k+1}$

evaluate f in the centers of the new hyperrectangles

$f_{min} = \min\{f(c) : c \in C_k\}$, $X_{min} = \{c \in C_k : f(c) = f_{min}\}$, $k = k + 1$
($C_k = \{$centers of the hyperrectangles in $\mathcal{H}_k\}$)

until (stopping criterion satisfied)

return f_{min}, X_{min}

Figure 3: Structure of DIRECT.

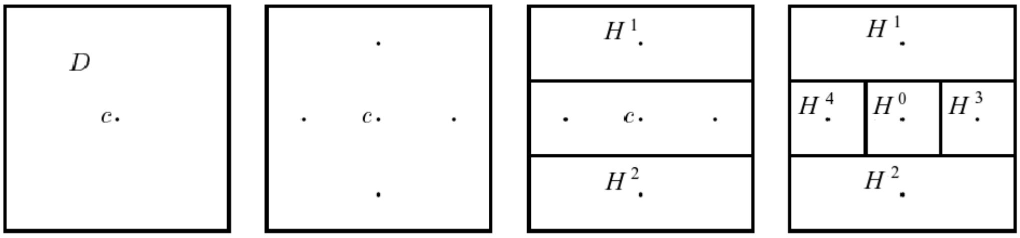

Figure 4: Partitioning of the hyperrectangles in DIRECT.

A simpler rule is described in [2]. However, a discussion on the partitioning strategies and of their impact on the algorithm is beyond the scope of this paper, and the reader is referred to [9] for details on this issue. In the selection strategy, a hyperrectangle $\hat{H} \in \mathcal{H}_k$ is defined potentially optimal if there exists $\alpha > 0$ such that

$$f(\hat{c}) - \alpha\hat{d} \leq f(c) - \alpha d \quad \forall\, H \in \mathcal{H}_k, \tag{5}$$

$$f(\hat{c}) - \alpha\hat{d} \leq f_{min} - \varepsilon|f_{min}|, \tag{6}$$

where f_{min} is the smallest value of f found so far, and ε is a parameter whose value is set at the beginning of the algorithm. A graphical interpretation of the selection criteria (5)-(6) can be given, as shown in Figure 5. Each hyperrectangle $H \in \mathcal{H}_k$ is represented as a black dot with coordinates (d, f); criteria (5)-(6) select hyperrectangles on the lower-right convex hull of the points in the graph. In other words, the hyperrectangles are grouped according to their sizes, and the best ones (i.e. with smallest function values) in some of these groups are selected. This partitioning strategy allows a balance between local and global search, since it selects both small and large hyperectangles, by weighting their function values with respect to the center-vertex distance. The parameter ε is a safeguard against excessive local search [8]; the larger it is, the higher the probability to exclude hyperrectangles with very good function values, but rather small sizes. Conversely, a small value of ε biases the algorithm to select small hyperrectangles as potentially optimal, thus forcing local search. In that sense, ε can be interpreted as a tuning parameter, to further control the balance between local and global search [8].

At each iteration DIRECT samples a set of points of D, i.e. the centers of the hyperrectangles of the domain partition; its convergence is based on the fact that, as the

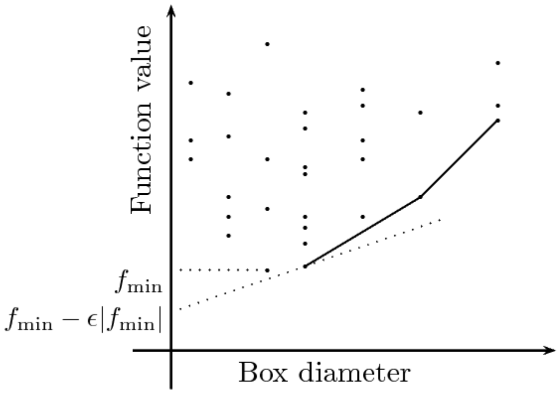

Figure 5: Potentially optimal hyperrectangles.

number of iterations goes to infinity, the set of sampled points becomes everywhere dense in D [1].

In this work we consider the original version of DIRECT, as described in [1]. When applied to problem (4), it produces a sampling of the feasible domain Ω which is highly uniform, because of the presence of a huge number of local solutions. This may result in a very slow convergence. On the other hand, the grid described in Section 3 contains significant information about the behaviour of the objective function, which could be exploited to drive the partitioning strategy of DIRECT. Therefore, we propose a modification to DIRECT aimed to embody the information of the grid.

We start from the observation that when DIRECT is applied to our problem, if, at some iteration, two hyperrectangles H_i and H_j have the same diameter, they are grouped together and discriminated against each other just on the basis of their representative function values. Nevertheless, given the highly nonuniform distribution of the points of G in Ω (see Figure 2), it may happen that

$$|G \cap H_i| >> |G \cap H_j|,$$

where $|A|$ denotes the cardinality of a set A. In this case, if the representative function values $f(c_i)$ and $f(c_j)$ are close, it seems reasonable giving higher priority to H_i for further investigation, and hence subdivision, in the next iterations. This simple observation suggests that, in selecting the potentially optimal hyperrectangles, the algorithm should consider also the number of grid points that fall into them. Since the selection criteria are based on the representative function values and the sizes of the hyperrectangles, the most straightforward way to embody grid-driven information in the selection process is to modify the definition of hyperrectangle size. Specifically, in the selection criteria (5)-(6) we substitute the center-vertex distance d with

$$d^* := \begin{cases} |G \cap H| & \text{if } |G \cap H| > 1, \\ d & \text{otherwise.} \end{cases} \tag{7}$$

We note that, for every hyperractangle, $d^* \geq d$ and, in particular, for those hyperractangles having $d^* = d$, it results $d^* < 1$ (since $n = 2$).

With this choice, the hyperrectangles are grouped according to the number of grid points they contain, if this number is greater than one, otherwise according to the diameter length. By using the graphical interpretation in Figure 5, the dots representing hyperrectangles that contain at most one grid point are on the left side.

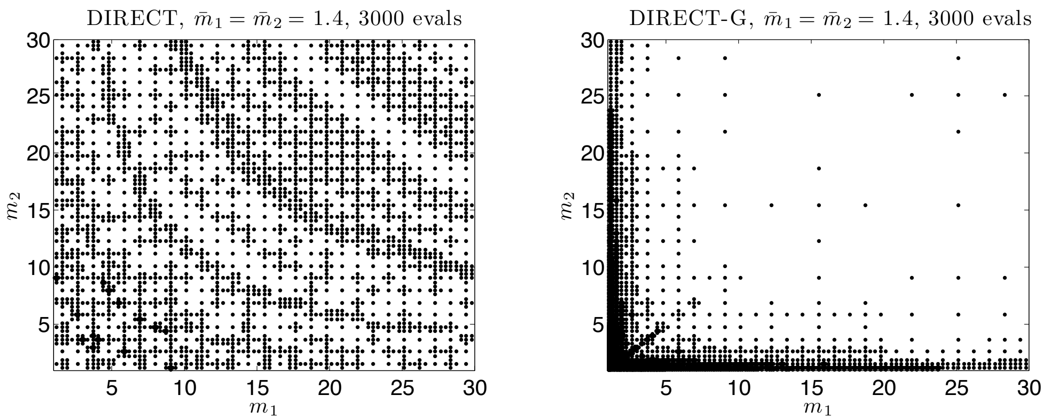

Figure 6: Points of the feasible domain generated by DIRECT (left) and DIRECT-G (right) after 3000 objective function evaluations, for a problem with solution $(\bar{m}_1, \bar{m}_1) = (1.4, 1.4)$.

It is possible to prove that the proposed modification of DIRECT, henceforth called DIRECT-G, still retains the everywhere dense convergence of the original algorithm. To this aim, we show that an iteration index $\bar{k}$ exists such that for every $k \geq \bar{k}$ DIRECT-G acts as DIRECT.

Proposition 4.1 *Let $\{\mathcal{H}_k\}$ be the sequence of domain partitions generated by DIRECT-G. Then an iteration index $\bar{k}$ exists such that, for every $k \geq \bar{k}$ and $H \in \mathcal{H}_k$, $d^* = d$.*

Proof. Since the grid G contains a finite number of points, a constant $b > 0$ exists such that the Euclidean distance between any two points of G is bounded below by b.

Assume by contradiction that for all k a hyperrectangle $H \in \mathcal{H}_k$ exists such that its size is

$$d^* = |G \cap H| > 1 \tag{8}$$

(we recall that, for every hyperractangle having $d^* = d$, it results $d^* < 1$). At a generic iteration k, let $\widetilde{H}$ be a hyperrectangle having the largest value of d^* among all the hyperrectangles in $\mathcal{H}_k$; the selection strategy implies that $\widetilde{H}$ is chosen for further subdivision at some iteration $\tilde{k} \geq k$. By repeating this argument, we can conclude that an iteration index $\bar{k}$ exists such that every $H \in \mathcal{H}_{\bar{k}}$ has diameter strictly smaller than b. Hence, none of the hyperrectangles can contain more than one grid point, thus contradicting (8) and concluding the proof. $\triangleleft$

From the above proposition it follows that Algorithm DIRECT-G inherits the everywhere dense convergence of the original DIRECT method whose proof can be found, e.g., in [1, 10].

In order to show the different behaviour of DIRECT and DIRECT-G, we show in Figure 6 the set of points generated, after 3000 function evaluations, by the two algorithms applied to an instance of problem (1) with optimal solution $(\bar{m}_1, \bar{m}_2) = (1.4, 1.4)$. We see that DIRECT tends to uniformly cover the feasible domain, by oversampling regions where G has very few points (see Figure 2), and therefore $F(x)$ has very little variability. Conversely, DIRECT-G forces the partitioning strategy to follow the spatial distribution of the points of G, thus speeding up the progress toward the optimal point.

5 Numerical Experiments

We performed numerical esperiments to evaluate the effectiveness of DIRECT-G versus DIRECT and the grid search, in solving problem (1). We also compared DIRECT-G with the Genetic Algorithm in [14], specifically developed for this problem.

We generated three sets of test problems where the data to be analysed contain white noise plus a gravitational signal with SNR $\bar{\mathcal{S}} = 10$. Each set corresponds to a specific configuration of the binary system, i.e. to a pair of masses:

- $\bar{m}_1 = \bar{m}_2 = 1.4 M_\odot$ (two neutron stars);
- $\bar{m}_1 = 1.4 M_\odot$, $\bar{m}_2 = 10 M_\odot$ (a neutron star and a black hole);
- $\bar{m}_1 = 5 M_\odot$, $\bar{m}_2 = 10 M_\odot$ (two black holes).

The three configurations are representative of coalescing binary systems of compact objects. For each pair of masses we considered 30 instances of noise, thus obtaining 30 data streams to be analysed. We note that the smaller the masses, the more difficult is the set of test problems, according to the spatial distribution of the grid points in the space of masses (see Figure 2). Furthermore, the test set corresponding to $\bar{m}_1 = \bar{m}_2 = 1.4$ is the most significant, since binary systems of neutron stars are known to exist and, for some of them, general relativistic effects in the binary orbits have been accurately measured [22]. The length of the data streams to be analysed is $N = 131072$, whereas the length of the signal depends on the masses: it is $M = 51207$ for $\bar{m}_1 = \bar{m}_2 = 1.4$, $M = 10823$ for $\bar{m}_1 = 1.4$ and $\bar{m}_2 = 10$, and $M = 3216$ for $\bar{m}_1 = 5$ and $\bar{m}_2 = 10$. The feasible domain is $\Omega = [1, 30] \times [1, 30]$. The grid G used in the experiments corresponds to a minimal match of 97%, i.e. allows a relative error of at most 3% in the maximum of the mean value of F; it consists of 27379 points and is independent of the gravitational signal parameters. All the data were generated by using the LAL package [23], which is considered a reference tool for GWs data analysis.

Fortran 90, double precision implementations of DIRECT-G and DIRECT were used in the experiments; both codes do not exploit the symmetry of the objective function. A maximum number of function evaluations equal to the number of grid points, i.e. to the number of function evaluations of the grid search, was considered as stopping criterion. The experiments were carried out using $\varepsilon = 10^{-4}, 10^{-6}, 10^{-8}$. We observed that both DIRECT-G and DIRECT were rather insensitive to the choice of ε when applied to the problems corresponding to $\bar{m}_1 = 1.4$ and $\bar{m}_2 = 10$ and to $\bar{m}_1 = 5$ and $\bar{m}_2 = 10$. On the test problems with $\bar{m}_1 = \bar{m}_2 = 1.4$ DIRECT continued to be insensitive to ε (at least within the maximum number of iterations), while DIRECT-G resulted faster with $\varepsilon = 10^{-8}$, according to the high density of the grid around the optimal point (1.4,1.4). Therefore, we show only the results obtained with $\varepsilon = 10^{-8}$.

In Figure 7, for each set of test problems, we plot the number of function evaluations performed by DIRECT-G and DIRECT against the best function value obtained with this number of function evaluations. Actually, the best function value is the mean of the best function values computed for each problem in the test set. We also report, in Table 1, the number of function evaluations needed by DIRECT-G to achieve the reference value $\bar{\mathcal{S}}$. The results show that DIRECT-G largely outperforms DIRECT for $\bar{m}_1 = \bar{m}_2 = 1.4$ and for $\bar{m}_1 = 1.4$ and $\bar{m}_2 = 10$. For the former set of problems, DIRECT is not able to achieve $\bar{\mathcal{S}}$ within the maximum number of iterations; furthermore, the best function value is far below $\bar{\mathcal{S}}$. Conversely, DIRECT-G achieves $\bar{\mathcal{S}}$ with 5191 function evaluations. For $\bar{m}_1 = 1.4$ and $\bar{m}_2 = 10$ both the algorithms are able to compute accurate solutions, but DIRECT-G is more efficient, since it achieves $\bar{\mathcal{S}}$ with 23% of the function evaluations required by

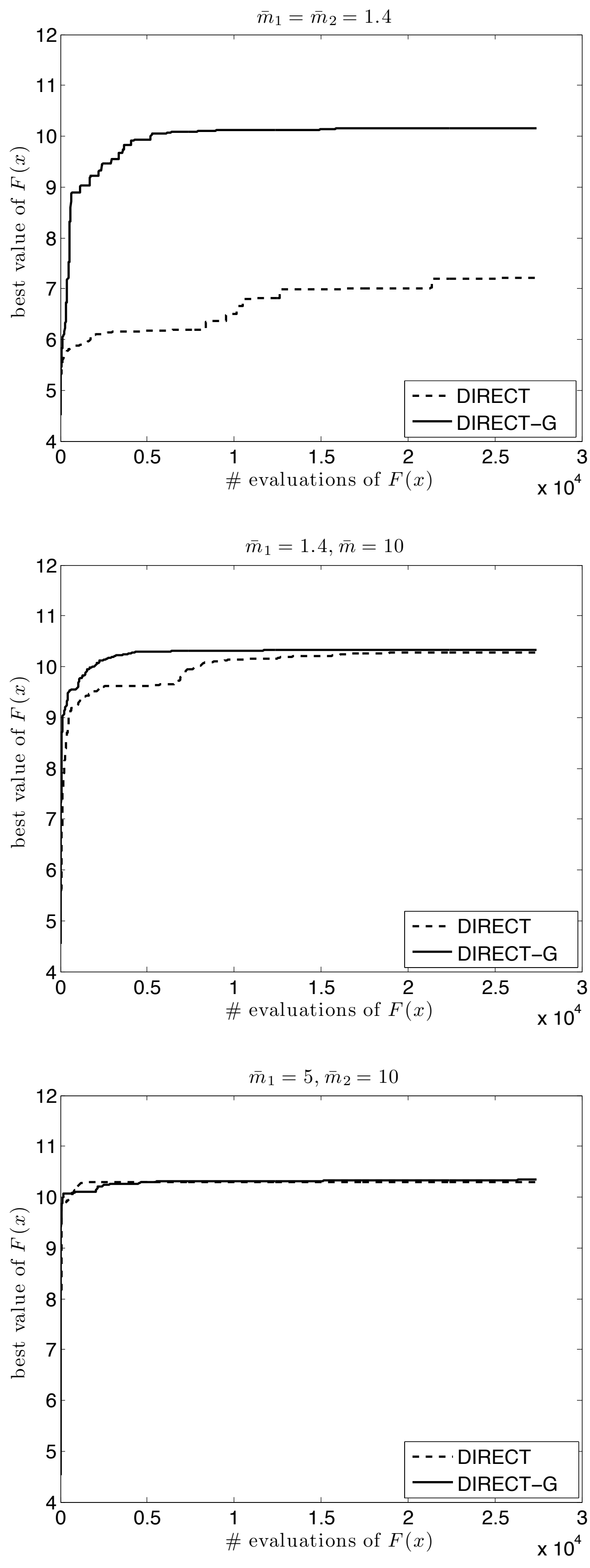

Figure 7: Comparison of DIRECT-G and DIRECT.

$\bar{m}_1$	$\bar{m}_2$	DIRECT-G	DIRECT
1.4	1.4	5191	—
1.4	10	1845	7907
5	10	155	575

Table 1: Comparison of DIRECT-G and DIRECT: number of function evaluations to achieve the reference value $\bar{\mathcal{S}}$.

$\bar{m}_1$	$\bar{m}_2$	GA	DIRECT-G	evals
1.4	1.4	9.8617	9.9215	4202
1.4	10	10.2529	10.2588	3992
5	10	10.2669	10.2545	3638

Table 2: Comparison of DIRECT-G and GA: mean of best objective function values obtained with the same number of function evaluations.

DIRECT. For $\bar{m}_1 = 5$ and $\bar{m}_2 = 10$ DIRECT and DIRECT-G show very close behaviours. The reference value $\bar{\mathcal{S}}$ is achieved by DIRECT-G with few function evaluations, which amount to 27% of those required by DIRECT; on the other hand, if we consider the best function value achieved within the maximum number of function evaluations, we see that DIRECT gets it faster than DIRECT-G. The latter result agrees with the fact that the partitioning strategy of DIRECT-G is biased to favour hyperrectangles containing many grid points. We note also that, for all the test sets, the number of function evaluations performed by DIRECT-G to achieve $\bar{\mathcal{S}}$ is much smaller than the number of function evaluations required by the grid search.

Finally, we compare DIRECT-G with the Genetic Algorithm (GA) in [14], applied to the same set of test problems considered here. GA was run with a population of 100 individuals randomly taken from G and a number of generations (iterations) equal to 50. For each problem in a test set, GA was executed 30 times, choosing different seeds to initialize the pseudo-random number generator used in the algorithm. We note that GA was designed to exploit the symmetry of F. To make a comparison with the best results reported in [14], we consider the best function values obtained by DIRECT-G with the same number of function evaluations performed by GA to achieve its best value (see Table 2). We see that the two algorithms show about the same performance. However, GA has no guarantee of convergence and requires a suitable setting of several parameters, whereas DIRECT-G inherits the convergence properties of DIRECT and requires only the setting of ε.

6 Conclusions

We proposed a modified version of the DIRECT algorithm, called DIRECT-G, for solving a box-constrained global optimization problem arising in the detection of GWs emitted by coalescing binary systems. DIRECT-G embodies the information given by the grid, a discretization of the feasible domain that provides an approximation of the optimal solution with a fixed accuracy in a probabilistic sense. Numerical experiments on different test problems showed that DIRECT-G largely outperforms DIRECT and the grid search, the latter being the reference algorithm for the astrophysics community; in particular, DIRECT-G resulted very efficient on the hardest test problems. DIRECT-G resulted also comparable with a genetic algorithm specifically developed for the optimization problem under consideration. However, DIRECT-G has strong theoretical convergence properties, while the convergence of the genetic algorithm is not guaranteed. Finally, we expect that a modification of DIRECT-G, to consider only the hyperrectangles included in a half of the feasible domain ($m_1 \leq m_2$), will further improve its performance.

References

1. D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.

2. D. R. Jones. DIRECT global optimization. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of optimization*, pages 725–735. Springer, 2009.

3. R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to global optimization.* Kluwer Academic Publishers, second edition, 2000.

4. M. C. Bartholomew-Biggs, S. C. Parkhurst, and S. P. Wilson. Using DIRECT to solve an aircraft routing problem. *Computational Optimization and Applications*, 21(3):311–323, 2002.

5. R. G. Carter, J. M. Gablonsky, A. Patrick, C. T. Kelley, and O. J. Eslinger. Algorithms for noisy problems in gas transmission pipeline optimization. *Optimization and Engineering*, 2(2):139–157, 2001.

6. S. E. Cox, R. T. Haftka, C. A. Baker, B. Grossman, W. H. Mason, and L. T. Watson. A comparison of global optimization methods for the design of a high-speed civil transport. *Journal of Global Optimization*, 21(4):415–433, 2001.

7. J. He, L. T. Watson, N. Ramakrishnan, C. A. Shaffer, A. Verstak, J. Jiang, K. Bae, and W. H. Tranter. Dynamic data structures for a direct search algorithm. *Computational Optimization and Applications*, 23(1):5–25, 2002.

8. J. M. Gablonsky and C. T. Kelley. A locally-biased form of the DIRECT algorithm. *Journal of Global Optimization*, 21(1):27–37, 2001.

9. G. Liuzzi, S. Lucidi, and V. Piccialli. A partition-based global optimization algorithm. *Journal of Global Optimization*, 2010.

10. G. Liuzzi, S. Lucidi, and V. Piccialli. A DIRECT-based approach exploiting local minimizations for the solution of large-scale global optimization problems. *Computational Optimization and Applications*, 2008.

11. T. D. Panning, L. T. Watson, N. A. Allen, K. C. Chen, C. A. Shaffer, and J. J. Tyson. Deterministic parallel global parameter estimation for a model of the budding yeast cell cycle. *Journal of Global Optimization*, 40:719–738, 2008.

12. K. S. Thorne. Gravitational radiation. In S. W. Hawking and W. Israel, editors, *300 Years of Gravitation*, pages 330–458. Cambridge University Press, 1987.

13. S. Babak, R. Balasubramanian, D. Churches, T. Cokelaer, and B. S. Sathyaprakash. A template bank to search for gravitational waves from inspiralling compact binaries I: physical models. *Classical and Quantum Gravity*, 23:5477–5504, 2006.

14. D. di Serafino, S. Gomez, L. Milano, F. Riccio, and G. Toraldo. A genetic algorithm for a global optimization problem arising in the detection of gravitational waves. *Journal of Global Optimization*, 48(1):41–55, 2010.

15. D. di Serafino and F. Riccio. On the application of multiple-deme parallel genetic algorithms in astrophysics. *Proceedings of the 18th Euromicro International Conference on Parallel Distributed and Network-Based Processing*, pages 231–237, 2010.

16. L. Milano, F. Barone, and M. Milano. Time domain amplitude and frequency detection of gravitational waves from coalescing binaries. *Physical Review D*, 55(8):4537–4554, 1997.

17. S. D. Mohanty. Hierarchical search strategy for the detection of gravitational waves from coalescing binaries: extension to post-newtonian waveforms. *Physical Review D*, 57(2):630–658, 1998.

18. S. D. Mohanty and S. V. Dhurandhar. Hierarchical search strategy for the detection of gravitational waves from coalescing binaries. *Physical Review D*, 54(12):7108–7128, 1996.

19. B. J. Owen. Search templates for gravitational waves from inspiraling binaries: choice of template spacing. *Physical Review D*, 53(12):6749–6761, 1996.

20. L. Blanchet, B. Rlyer, and A. G. Wiseman. Gravitational waveforms from inspiralling compact binaries to second-post-Newtonian order. *Classical and Quantum Gravity*, 13:575–584, 1996.

21. R. Prix. Template-based searches for gravitational waves: efficient lattice covering of flat parameter spaces. *Classical and Quantum Gravity*, 24:S481–S490, 2007.

22. A. R. Rasio and S. L. Shapiro. Coalescing binary neutron stars. *Classical and Quantum Gravity*, 16(6):R1–R29, 1999.

23. B. Allen et al. *LAL Software Documentation.* Revision 1.44, 2005.