Statistics and Its Interface Volume 10 (2017) 59–70

Lipschitz optimization methods for fitting a sum of damped sinusoids to a series of observations

J. W. Gillard* and D. E. Kvasov†

A general nonlinear regression model is considered in the form of fitting a sum of damped sinusoids to a series of non-uniform observations. The problem of parameter estimation in this model is important in many applications like signal processing. The corresponding continuous optimization problem is typically difficult due to the high multiextremal character of the objective function. It is shown how Lipschitz-based deterministic methods can be well-suited for studying these challenging global optimization problems, when a limited computational budget is given and some guarantee of the found solution is required.

AMS 2000 subject classifications: Primary 90C26, 93B30; secondary 90C56.

1. INTRODUCTION

1.1 Statement of optimization problem

Consider the general nonlinear regression problem. Assume that we have a series of real-valued observations $y_1, \ldots, y_T$ such that

$$y_t = \eta(\theta, t) + \varepsilon_t, \;\; t = 1, \ldots, T,$$

where θ is an n-dimensional vector of unknown parameters, $\eta(\theta, t)$ is a function nonlinear in θ and $\varepsilon_1, \ldots, \varepsilon_T$ is a series of noise terms (often assumed independently and identically distributed random variables, with zero mean and variance σ^2). The non-uniform sampling case $t = t_l$, $l = 1, \ldots, T$, can be studied similarly. Let $\Theta \subset \mathbb{R}^n$ be a parameter space so that $\theta \in \Theta$. Parameter estimation in the general nonlinear regression model can be reduced to solving the minimization problem

$$(1) \qquad F(\theta) = \sum_{t=1}^{T} (y_t - \eta(\theta, t))^2 \to \min_{\theta \in \Theta}, \;\; \Theta \subset \mathbb{R}^n,$$

with the estimator θ^* defined as

$$\theta^* = \arg\min_{\theta \in \Theta} F(\theta).$$

This problem can be often stated as a box-constrained global optimization problem, i. e.,

$$(2) \qquad \Theta = [\theta^-, \theta^+] = \{\theta \in \mathbb{R}^n : \theta^-(j) \le \theta(j) \le \theta^+(j),\ j = 1, \ldots, n\}.$$

We consider the case where the function $\eta(\theta, t)$ has the form

$$(3) \qquad \eta(\theta, t) = \sum_{i=1}^{q} a_i \exp(d_i t) \sin(2\pi\omega_i t + \phi_i), \;\; t = 1, \ldots, T.$$

Here, q is a given integer, $\theta = (a, d, \omega, \phi)$ with $a = (a_1, \ldots, a_q)$, $d = (d_1, \ldots, d_q)$, $\omega = (\omega_1, \ldots, \omega_q)$ and $\phi = (\phi_1, \ldots, \phi_q)$. Denote the true vector of parameters by $\theta^{(0)} = (a^{(0)}, d^{(0)}, \omega^{(0)}, \phi^{(0)})$ ($\theta^{(0)}$ coincides with θ^* in the case of noise-free observations). The ranges for parameters a_i and d_i is $(-\infty, \infty)$, whilst the ranges for ω_i and ϕ_i are $[0, 1)$ and $[0, \pi/2)$, respectively.

The model (3) is a very well-known model which in the signal processing literature is called the 'sums of damped sinusoids' (see, e. g., [1, 2, 6, 9, 15, 37] for some typical techniques used in the signal processing field). This model is also associated with the so-called Hankel structured low-rank approximation problem which is described as follows. Let $\mathbf{X} = (x_{i,j})$ be an $L \times K$ matrix such that $x_{i,j} = y_{i+j-1}$ and $T + 1 = L + K$. The matrix $\mathbf{X}$ is of Hankel structure. The Hankel structured low rank approximation problem is that of finding another Hankel matrix 'close' to $\mathbf{X}$ (the most common instance of the problem uses the Frobenius norm; see, e. g., [7, 39]) which is of some pre-specified rank $r < \min(L, K)$. Full details describing the formal connection are available, e. g., in [19, 20, 21, 40].

1.2 Description of optimization problem

In this paper, we will show that optimization problem (1) is challenging if we consider the general model (3). We also will consider the following two particular cases of the model (3):

$$(4) \qquad \eta(a, \omega, \phi; t) = \sum_{i=1}^{q} a_i \sin(2\pi\omega_i t + \phi_i), \;\; t = 1, \ldots, T,$$

with $d = (0, 0, \ldots, 0)^T$, and

$$(5) \qquad \eta(a, d, \phi; t) = \sum_{i=1}^{q} a_i \exp(d_i t) \sin(2\pi\omega_i^{(0)} t + \phi_i), \;\; t = 1, \ldots, T,$$

*Corresponding author.

†The research executed to complete the revised version of the paper was supported by the Russian Science Foundation, project number 15-11-30022 "Global optimization, supercomputing computations, and applications".

where it is assumed that the vector ω is known: $\omega = \omega^{(0)}$. We will thus consider the following sets of parameters:

(i) $\theta = (a, d, \omega, \phi)$, the general case.
(ii) $\theta = (a, \omega, \phi)$, so that it is assumed that the vector d is known: $d = d^{(0)}$. In many applications, $d = (0, 0, \ldots, 0)^T$.
(iii) $\theta = (a, d, \phi)$, so that it is assumed that the vector ω is known: $\omega = \omega^{(0)}$.

The problem (i) is a classical problem and is important in spectral analysis (see, e. g., [18, 22, 46]) and many applications, such as magnetic resonance spectroscopy, radioastronomy, antenna theory, prospecting seismology, and so on (see, e. g., [10, 33, 29, 46]). The problem (ii) with $d = 0$ can be considered as a simple extension of the Fourier expansion and hence can be applied in many different fields (see, e. g., [2, 15]). As extension of the Fourier model, it would be especially valuable when the frequencies ω_i are not expected to necessarily be $1/k$ (this includes the case of the so-called quasi-periodic signals). The problem (iii) naturally appears in models where the frequencies are known (see, e. g., [46]). A typical example would be provided by a monthly economic activity time series where the most dominant frequency is 12 supplemented with fractions $12/k$ with $k = 2, 3, 4, 5$ and 6.

We are aware of a few papers that contain discussion about the behaviour of the objective function (1) (see, e. g., [34, 20, 21]). In [19, 20], the fact that the objective function F is multiextremal has been observed; the function F was decomposed into three different components and it was numerically demonstrated that the part of the objective function with the observation noise removed dominates the shape of the objective function. The optimization problem (1) is very difficult (even in the case $q = 1$ in (3)–(5)) with the objective function possessing many local minima (see also the related discussion in [29]). Although the objective function is Lipschitz-continuous (see, e. g., [21, 29, 55, 59, 61]), it has very high Lipschitz constants which increase with T, the number of observations. Additionally, increasing T leads to more erratic objective functions with a higher number of local minima. Adding noise to the observed data (see, e. g., [4, 11]) increases the complexity of the objective function (see, e. g., [5, 63]) and moves the global minimizer away from the vector of true parameters. Thus, efficient global optimization techniques should be used to tackle the stated problem.

Deterministic algorithms can be well-suited for studying the considered global optimization problems, since they are often able to provide a solution to the problem together with some guaranteed gap. Among them, Lipschitz-based methods can behave particularly well with respect to other state-of-the-art deterministic methods when a limited computational budget is given and some guarantee of the found solution is required (see, e. g., [13, 28, 29, 30, 43, 45, 54, 58, 59, 61]). We will show in this paper how these methods can be applied for solving the problems (1)–(5).

1.3 Structure of paper

The rest of the paper is structured as follows. In the next Section 2, some simple but important examples of the cost functions (1) are reported and studied in order to illustrate the complexity of the stated identification problem (1). In Section 3, Lipschitz-based deterministic techniques are briefly reviewed. Their application to solving the problems described in Section 2 is shown in Section 4. Finally, conclusions and future research directions are drawn in Section 5.

2. EXAMPLES ILLUSTRATING THE COMPLEXITY OF THE PROBLEM

2.1 Benchmark problems

In this Section, we provide a number of examples to illustrate the complexity of the problem (1) with the function given by (3)–(5). Figure 1 contains plots of the objective function for the following cases:

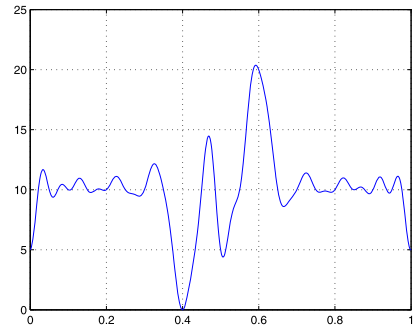
(a) Sine with unknown frequency only: we take $q = 1$ (the simplest case) and $T = 10$ in (3) and consider a particular objective function of the form

$$F(\omega_1) = \sum_{t=1}^{T} \left(y_t - \sin(2\pi\omega_1 t)\right)^2, \tag{6}$$

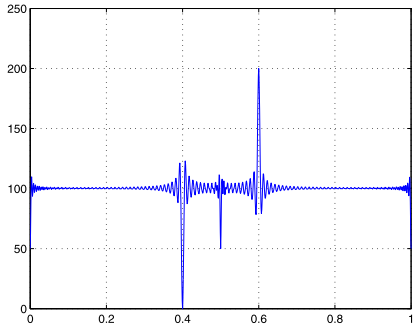
where $\omega^{(0)} = \omega_1^{(0)} = 0.4$ (see Figure 1(a)). This one-dimensional function $F(\omega_1)$ is periodic with period 1 and the minimum value $F^* = 0$ (noise-free observations are considered for simplicity) is attained at the points $\omega_1^* = \omega_1^{(0)} + p$ ($p = 0, \pm 1, \pm 2, \ldots$). The feasible domain for ω_1 can therefore be chosen as $[0, 1) \subset \mathbb{R}^1$; in this interval, the function $F(\omega_1)$ has one global minimizer at $\omega_1^* = \omega_1^{(0)}$ and many local minimizers.

(b) We repeat the above but take a higher $T = 100$ (see Figure 1(b)).
In both the cases (a) and (b), it can be easily seen that the objective function $F(\omega_1)$ is highly multiextremal and very irregular (for the parameter settings in these examples, the number of local minimizers increases linearly in T). For $T = 10$, the Lipschitz constant of $F(\omega_1)$ (estimated over 10^{-7}-grid) is approximately 432.0. For $T = 100$, it becomes equal to approximately 28690.8. The global minimizer has a very narrow attraction region and it moves away from the true parameters vector $\theta^{(0)}$ in (3) when the observations y_t are measured with noise. Therefore, already in these relatively simple cases ($q = 1$ in (3)) particular care should be taken in the choice of global optimization algorithms to find this global minimizer.

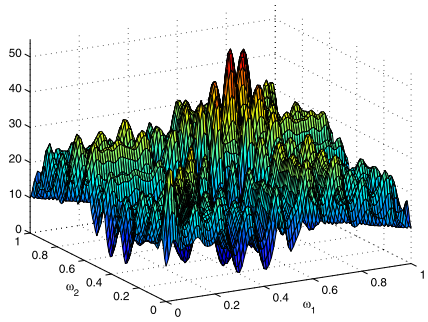
(c) Two sines with unknown frequencies only: we take $q = 2$ in (3), $T = 10$ noise-free observations in (1) and consider the following objective function over the admissible domain $\omega = (\omega_1, \omega_2) \in [0, 1) \times [0, 1)$:



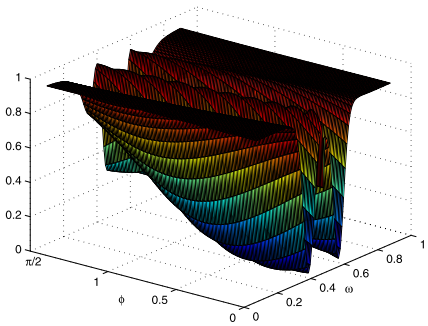
(a) $F(\omega_1)$ from (6) with $T = 10$



(b) $F(\omega_1)$ from (6) with $T = 100$



(c) $F(\omega_1, \omega_2)$ from (7) with $T = 10$



(d) Cross-section (ω, ϕ) of $f(d, \omega, \phi)$ from (11)

Figure 1. Objective functions as described in subsection 2.1.

$$(7)\quad F(\omega_1, \omega_2) = \sum_{t=1}^{T} \left(y_t - \sin(2\pi\omega_1 t) - \sin(2\pi\omega_2 t)\right)^2,$$

where $\omega_1^{(0)} = 0.3$, $\omega_2^{(0)} = 0.4$ (see Figure 1(c)).
The two-dimensional objective function $F(\omega_1, \omega_2)$ is again highly multiextremal (note its symmetry with respect to permutation of the parameters). Although two global minimizers are well-separated (in terms of the objective function values too) from the multitude of local minimizers, the effect of having two close frequencies can be seen in Figure 1(c).

(d) Sine with unknown frequency, phase, and amplitude: we consider again the optimization problem (1) defined by the model (3) with $q = 1$. As (3) is quadratic in the parameter a, it is possible (and sometimes can be useful) to obtain an explicit estimator for a based on the remaining parameters d, ω, and ϕ (as shown in the next subsection 2.2), thus defining the three-dimensional objective function $f(d, \omega, \phi)$ for $d \in [-2, 2] \subset \mathbb{R}^1$, $\omega \in (0, 1)$, and $\phi \in (0, \pi/2)$. In Figure 1(d), its cross-section (ω, ϕ) with $d = d^{(0)} = -0.2$ in the case of $T = 10$ noise-free observations is plotted, with true values $a^{(0)} = 1.0$, $d^{(0)} = -0.2$, $\omega^{(0)} = 0.4$, and $\phi^{(0)} = 0.3$. A highly multiextremal behaviour of this three-dimensional objective function can be again observed, especially with respect to the frequency ω.

The described objective functions are Lipschitz-continuous with high Lipschitz constants, essentially multiextremal and derivable (see the next subsection 2.2). Their evaluation is often associated (especially when a high number T of noisy observations are given) with performing computationally expensive numerical experiments. Moreover, noisy observations in (1) furthermore increase the problem complexity by both shifting the global minimizer away from the true parameters values and making unknown the desired global minimum value F^* (which is equal to $F^* = 0$ in the case of noise-free observations).

2.2 Derivatives and simplifications of the benchmark objective functions

Let us consider the optimization problem (1) defined by the model (3) with different unknown parameters, as described in the previous subsection 2.1.

If $q = 1$ (the simplest case of one sine function in (3)) and only one frequency ω_1 is unknown (the cases (a) and (b) in subsection 2.1), the first derivative of the objective function $F(\omega_1)$ from (6) is explicitly written over $\omega_1 \in [0, 1)$ as

$$(8)\quad F'(\omega_1) = -4\pi \sum_{t=1}^{T} t \cos(2\pi\omega_1 t)(y_t - \sin(2\pi\omega_1 t)).$$

If two sine functions ($q = 2$ in (3)) with two unknown frequencies ω_1 and ω_2 are considered (the case (c) in sub-

section 2.1 with $F(\omega_1, \omega_2)$ from (7)), then

(9)
$$\frac{\partial F}{\partial \omega_1} = -4\pi \sum_{t=1}^{T} t \cos(2\pi\omega_1 t)(y_t - \sin(2\pi\omega_1 t) - \sin(2\pi\omega_2 t)),$$
$$\frac{\partial F}{\partial \omega_2} = -4\pi \sum_{t=1}^{T} t \cos(2\pi\omega_2 t)(y_t - \sin(2\pi\omega_1 t) - \sin(2\pi\omega_2 t)).$$

Let us finally consider the case (d) from subsection 2.1, with $q = 1$ in (3) but unknown frequency, phase, and amplitude parameters in the dumped sinusoid (3). For brevity, let $x_t = \exp(dt)\sin(2\pi\omega t + \phi)$. Equation (1) may be written

$$F(a, d, \omega, \phi) = \sum_{t=1}^{T} (y_t - a x_t)^2 . \tag{10}$$

Since

$$\frac{\partial F(a, d, \omega, \phi)}{\partial a} = -2 \sum_{t=1}^{T} (y_t - a x_t) x_t ,$$

then we may obtain ($x_t \neq 0$ for all $t = 1, \ldots, T$) an explicit estimator for a, which we denote $\hat{a}$. This estimator is a function of the remaining parameters d, ω and ϕ:

$$\hat{a} = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} .$$

Substituting $\hat{a}$ into (10) (such a transformation can be used to decrease the number of independent parameters) gives a new objective function, which we denote $f(d, \omega, \phi)$:

$$f(d, \omega, \phi) = \sum_{t=1}^{T} \left(y_t - x_t \frac{\sum_{k=1}^{T} y_k x_k}{\sum_{k=1}^{T} x_k^2} \right)^2 . \tag{11}$$

It is possible to compute the derivatives of the objective function with respect to each of the unknown parameters, although they cannot be always written in a neat form. Here, we state only the first derivatives of $f(d, \omega, \phi)$ with respect to each of the unknown parameters:

$$\frac{\partial f}{\partial d} = -2 \sum_{t=1}^{T} \left\{ \left[y_t - \frac{x_t \sum_{k=1}^{T} y_k x_k}{\sum_{k=1}^{T} x_k^2} \right] \times \left[\frac{x_t \sum_{k=1}^{T} k y_k x_k}{\sum_{k=1}^{T} x_k^2} + \frac{x_t \sum_{k=1}^{T} y_k x_k}{\left(\sum_{k=1}^{T} x_k^2 \right)^2} \left(t \sum_{k=1}^{T} x_k^2 - 2 \sum_{k=1}^{T} k x_k^2 \right) \right] \right\}. \tag{12}$$

Let $c_1^{(t)} = t \exp(dt)\cos(2\pi\omega t + \phi) \sum_{k=1}^{T} x_k^2 - 2 x_t \sum_{k=1}^{T} k x_k \exp(dk)\cos(2\pi\omega k + \phi)$. Then,

$$\frac{\partial f}{\partial \omega} = -4\pi \sum_{t=1}^{T} \left\{ \left[y_t - \frac{x_t \sum_{k=1}^{T} y_k x_k}{\sum_{k=1}^{T} x_k^2} \right] \times \left[\frac{x_t \sum_{k=1}^{T} y_k k \exp(dk)\cos(2\pi\omega k + \phi)}{(\sum_{k=1}^{T} x_k^2)^2} + \frac{\sum_{k=1}^{T} y_k x_k}{\left(\sum_{k=1}^{T} x_k^2 \right)^2} c_1^{(t)} \right] \right\}. \tag{13}$$

Let $c_2^{(t)} = \exp(dt)\cos(2\pi\omega t + \phi) \sum_{k=1}^{T} x_k^2 - 2 x_t \sum_{k=1}^{T} x_k \exp(dk)\cos(2\pi\omega k + \phi)$. Then,

$$\frac{\partial f}{\partial \phi} = -2 \sum_{t=1}^{T} \left\{ \left[y_t - \frac{x_t \sum_{k=1}^{T} y_k x_k}{\sum_{k=1}^{T} x_k^2} \right] \times \left[\frac{x_t \sum_{k=1}^{T} y_k \exp(dk)\cos(2\pi\omega k + \phi)}{(\sum_{k=1}^{T} x_k^2)^2} + \frac{\sum_{k=1}^{T} y_k x_k}{\left(\sum_{k=1}^{T} x_k^2 \right)^2} c_2^{(t)} \right] \right\}. \tag{14}$$

It can be shown that all the derivatives (8), (9), and (12)–(14) are Lipschitz-continuous functions (generally, with very high but unknown Lipschitz constants) over suitably defined domains of the parameters.

3. LIPSCHITZ-BASED DETERMINISTIC METHODS FOR SOLVING THE STATED GLOBAL OPTIMIZATION PROBLEM

3.1 Deterministic and stochastic global optimization

In the global optimization literature, numerical methods are often divided into deterministic and stochastic. Deterministic algorithms (see, e. g., the references in [13, 45, 58, 59]) can ensure (assuming exact computations and a sufficient run time) that after a finite time an approximation of a global minimizer will be found within a preset tolerance. Stochastic methods (see, e. g., the references in [41, 42, 60, 61, 62]) usually offer some probabilistic guarantees of determining the global solution. Their convergence theory can state that the global minimizer is found in an infinitely long run time with the probability one. In both cases, derivative-free or derivative-based algorithms can be constructed depending on whether the objective function is differentiable and its derivatives can or cannot be calculated or estimated during the search for the best combination of the unknown parameters. Of course, the availability of the information on derivatives (like for our problems (1)–(5) particularly described in Section 2) allows one to apply faster and more accurate global optimization methods with respect to the situations when such information is unavailable.

Adaptive stochastic search strategies are frequently used for solving practical global optimization problems and are mainly based on random sampling in the search domain. Such techniques as adaptive random search, simulated annealing, tabu search, evolution and genetic algorithms can be cited here (see, e. g., [13, 42] for details). Stochastic techniques can often deal with the problems described in the Introduction in a simpler manner than the deterministic approaches (being also suitable for the case of noisy observations). However, some difficulties with these meth-

ods can occur. For example, solutions obtained by many popular stochastic algorithms (among them are heuristic nature-inspired methods like evolutionary algorithms, simulated annealing, etc., frequently used in practice; see, e. g., [13, 42, 31]) can be only local solutions to the problems, located far from the global ones. Several restarts can also be involved without providing an acceptable guarantee on the found solution. This can preclude such methods from their practical usage when accurate and guaranteed estimates of the unknown parameters are required under some assumptions on the problem, yet they can be useful to tackle some black-box problems with a lack of a priori suppositions on the objective functions.

Several general frameworks for describing computational schemes of global optimization methods (of both deterministic and stochastic types) and for studying their convergence properties in a unified manner have been proposed. The so-called 'divide-the-best' approach DBA (see [54, 49]) can be successfully used for this aim. In the DBA scheme, an adaptive partition of the parameter domain $\Theta \subset \mathbb{R}^n$ from (1) (which is a hyperinterval (2) in our problems) into subdomains Θ_i^k is considered at each iteration k. The 'merit' (called characteristic; see, e. g., [52, 59]) R_i of each subdomain for performing a more detailed exploration is calculated on the basis of the information about the objective function obtained by evaluating the function (and possibly its derivatives) at several parameter vectors $\theta \in \Theta_i^k$. The best (in some computational sense) characteristic estimated over a sub-domain Θ_h^k corresponds to a higher possibility to locate the global minimizer within Θ_h^k. Sub-domain Θ_h^k is, thus, partitioned at the next iteration of the algorithm and new function evaluations are executed, trying to improve the current approximation of the solution to problem (1) (see, e. g., [29, 49, 56] for details).

3.2 Lipschitz global optimization methods

To produce guaranteed estimates of the global solution within a finite number of function evaluations, as it is required in many applied design problems (including the considered problem (1)–(5)), some assumptions on the objective function structure should be made. These suppositions would allow one to develop a global search algorithm, able to outperform the simple uniform grid techniques for solving multiextremal problems (this is a well-known result in global optimization: if no particular assumptions are made on the objective function, any finite number of function evaluations cannot guarantee getting close to the global minimizer).

Efficient deterministic global optimization algorithms belonging to the DBA class can be constructed in the framework of Lipschitz global optimization (LGO). The Lipschitz condition is realistic for many practical simulation-based problems since in technical systems the energy of change is always limited and, thus, the corresponding objective function (or its derivatives) has (have) bounded slopes. The LGO algorithms work with the objective functions $F(\theta)$ from (1) that either satisfies the Lipschitz condition over the parameter space Θ with the Lipschitz constant $0 < L < \infty$ (usually unknown) or have the Lipschitz-continuous gradient with some unknown Lipschitz constant $0 < K < \infty$, i. e.,

$$\|\nabla F(\theta') - \nabla F(\theta'')\| \le K\|\theta' - \theta''\|, \ \theta', \theta'' \in \Theta \subset \mathbb{R}^n. \quad (15)$$

In what follows, condition (15) will be used in the LGO methods for solving the problem (1) since the objective function $F(\theta)$ from (1) is derivable and its gradient is Lipschitz-continuous (see subsection 2.2).

A classical geometric interpretation of the Lipschitz condition (ensuring the boundedness of the functions changes) allows one to develop global optimization algorithms and to prove their convergence. This interpretation lies at the basis of the so-called geometric LGO methods (see, e. g., [26, 56, 50, 54]). As in surrogate-based optimization (see, e.g., [13] for details), a geometric LGO algorithm iteratively constructs and updates a piecewise auxiliary function built by using an estimate of the Lipschitz constant K from (15); it performs an evaluation of $F(\theta)$ and $\nabla F(\theta)$ (the operation of evaluating $F(\theta), \nabla F(\theta)$ at a parameter vector $\theta \in \Theta \subset \mathbb{R}^n$ is often called trial) at a vector corresponding to a minimum of this auxiliary function. In certain cases, this point in $\mathbb{R}^n$ is easy to find (see, e.g., [59, 55, 26]).

Since it is possible to evaluate both the objective function $F(\theta)$ and its gradient $\nabla F(\theta)$ at a trial point $\theta \in \Theta$ from (2), more information about the problem is available (especially, the information on its local properties expressed by the gradient values). The usage of this information allows us to obtain auxiliary functions that fit closely the goal function and to generate a more accurate solution to the problem.

Different geometric LGO methods based on constructing auxiliary functions with the usage of various estimates of the Lipschitz constant K from (15) have been proposed (see, e. g., [13, 56, 14, 26, 27, 43, 54, 55, 58, 59]). They can be distinguished either by the mode in which information about the Lipschitz constant K from (15) is obtained or by the strategy of exploration of the parameter space Θ from (2).

In particular, several ways to specify the Lipschitz constant K can be considered: this constant can be given a priori (see, e. g., [3, 12]); its adaptive estimates (local or global) can be obtained during the search (see, e. g., [14, 26, 27, 35, 50, 54, 56, 59]); multiple estimates of the Lipschitz constant can be also used (see, e. g., [25, 27]). While for the first two groups of methods convergence to only global minimizers can be often ensured, the third group of methods (frequently used in practice) normally manifests the so-called everywhere dense convergence without the possibility to obtain guaranteed bounds on the solutions. It should be noted that either the Lipschitz constant is known and an LGO method is constructed correspondingly, or it is not known but there exist a sufficiently large number of parameters of the considered algorithm ensuring its convergence.

A smart balancing of local and global information during the search, as for example, by means of the so-called local tuning approach to estimate local Lipschitz constants, can be useful in this connection to accelerate the algorithm execution without compromising the problem solution (see, e. g., [26, 32, 35, 48, 50, 54, 58, 59]).

For exploring the multidimensional parameter space, various adaptive partitioning strategies can be applied. For example, one-point-based algorithms subsequently subdivide the search region in smaller ones and evaluate the objective function and its gradient at one point within each subregion (see, e. g., [12, 27, 38]). Diagonal partitions that evaluate $F(\theta)$ and $\nabla F(\theta)$ at two points within each subregion are very interesting for practical applications with expensive black-box functions (see, e. g., [45, 53, 54, 56]). More complex partitions, based, for example, on simplices can be also used (see, e. g., [8, 13, 43, 44, 61]), that can be useful to tackle symmetries in the objective function (see Section 2).

Particularly, in diagonal algorithms for solving problem (1)–(2) both the objective function $F(\theta)$ and its gradient $\nabla F(\theta)$ are evaluated (independently of the problem dimension) only at the vertices corresponding to the main diagonal $[\theta_i^-, \theta_i^+]$ of each hyperinterval Θ_i (either at both the vertices, see, e. g., [16, 32, 45, 51, 54, 56], or at only one of them, see, e. g., [27, 54]). Results of these trials are then used to calculate the characteristics of the generated hyperintervals and, thus, to estimate the function behavior over these hyperintervals. In order to calculate the characteristic R_i of a multidimensional hyperinterval Θ_i, some one-dimensional characteristics can be successfully used as prototypes (see, e. g., [17, 23, 26, 45, 54, 58, 59, 61]). They can be applied to the one-dimensional segment being the main diagonal $[\theta_i^-, \theta_i^+]$ of the hyperinterval Θ_i. A hyperinterval having the 'best' characteristic (e. g., the smallest lower bound of $F(\theta)$ over all the hyperintervals) is subdivided (by hyperplanes passing through some point on the main diagonal) by means of a diagonal partition strategy (as, e. g., bisection, partition 2^n, and non-redundant partition from [24, 51, 54]; see the recent review [57] for details) and new trials are performed.

For example, in Figure 2, function trials are performed at points θ_i^- and θ_i^+ of a hyperinterval Θ_i. The function $F(\theta)$ is approximated along the diagonal $[\theta_i^-, \theta_i^+]$ of Θ_i by means of a specially constructed auxiliary function $\varphi_i(\tilde{\theta})$, $\tilde{\theta} \in [\theta_i^-, \theta_i^+]$. The minimum value of this approximating function estimates the lower bound of $F(\theta)$ over the segment $[\theta_i^-, \theta_i^+]$. This estimate is multiplied (as graphically indicated by arrows in Figure 2) by a coefficient in order to be a lower estimate for $F(\theta)$ not only over the diagonal but also over the whole multidimensional hyperinterval Θ_i. This modified lower estimate is accepted as the characteristic R_i of the hyperinterval Θ_i (see, e. g., [24, 51, 54, 57] for more details).

The diagonal algorithms belong to the class of divide-the-best methods [50] and have proved to be efficient in solving practical problems (see, e. g., [28, 30, 29]). Two specific diagonal LGO methods, used in numerical experiments, are briefly presented in the next subsection.

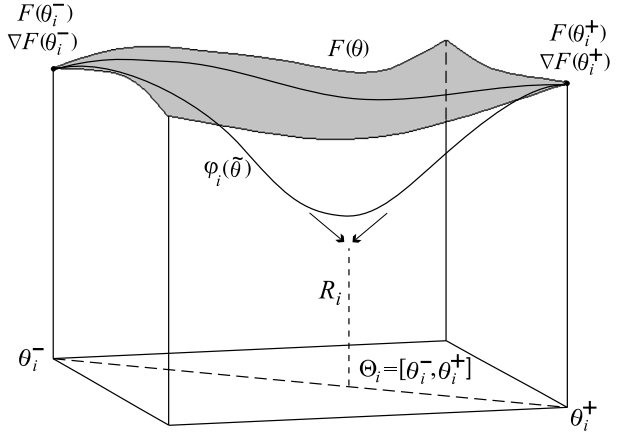


Figure 2. Obtaining the lower bound R_i for the objective function $F(\theta)$ with the Lipschitz gradient $\nabla F(\theta)$ over hyperinterval $\Theta_i = [\theta_i^-, \theta_i^+]$ by using smooth auxiliary function $\varphi_i(\tilde{\theta})$ along the main diagonal $[\theta_i^-, \theta_i^+]$ of Θ_i.

3.3 MultK and SmoothD methods

Among different LGO methods reviewed in the previous subsection, a particular attention in this work is given to two diagonal geometric LGO algorithms. They differ both in the used partition strategies and estimates of the Lipschitz constant for the gradient of $F(\theta)$.

The first algorithm (denoted hereafter as MultK) is from [27]. It extends the one-dimensional LGO method [25] using multiple estimates of the Lipschitz constant K by means of the non-redundant one-point-based diagonal strategy (see, e. g., [54, 57]). The second method (denoted hereafter as SmoothD) is from [56]. It uses smooth auxiliary functions to estimate $F(\theta)$ over main diagonals of hyperintervals (as in its one-dimensional prototype from [50]) and is based on the non-redundant diagonal partition strategy (see, e. g., [24, 51, 53, 54]). The first algorithm is also characterized by a refined usage of the local information during its work (enhanced by the so-called two-phase approach; see, e. g., [53, 47, 44]), whilst the latter one is taken in its basic version that can be further improved (for example, by adopting the local tuning technique from [32, 48, 50] or the two-phase approach as in [27, 53, 44]).

In both the methods, the construction of a piecewise quadratic auxiliary function along the main diagonal of a hyperinterval is based on the Taylor expansion of $F(\theta)$ limited to the second order term. Once a trial at vertex $\theta_i \in \Theta_i = [\theta_i^-, \theta_i^+]$ is executed, the following inequality holds for an estimate $\tilde{K} \geq K$ of the Lipschitz constant K from (15):

(16)
$$F(\theta) \geq F(\theta_i) + \langle \nabla F(\theta_i), (\theta - \theta_i) \rangle - 0.5\tilde{K}\|\theta - \theta_i\|^2, \quad \theta \in \Theta_i,$$

where $\langle \cdot, \cdot \rangle$ is the scalar product, $\| \cdot \|$ is the Euclidean norm in $\mathbb{R}^n$.

In the context of the diagonal methods, the directional derivative $F'(\theta_i)$ evaluated at a vertex θ_i along the main diagonal $[\theta_i^-, \theta_i^+]$ of Θ_i (i. e., either $\theta_i = \theta_i^-$ or $\theta_i = \theta_i^+$) can be also used instead of the gradient vector $\nabla F(\theta_i)$:

$$F'(\theta_i) = \left(\sum_{j=1}^{n} \frac{\partial F(\theta_i)}{\partial \theta(j)} (\theta_i^+(j) - \theta_i^-(j)) \right) / \Delta_i, \tag{17}$$

where

$$\Delta_i = \|\theta_i^+ - \theta_i^-\| = \sqrt{\sum_{j=1}^{n} (\theta_i^+(j) - \theta_i^-(j))^2}$$

is the length of the main diagonal of hyperinterval $\Theta_i \subset \Theta$, n is from (2).

Since the exact Lipschitz constant K for the gradient of $F(\theta)$ (or its valid overestimate) can be unknown in the stated global optimization problem, a set of possible estimates for K from zero to infinity is used in the MultK method, thus introducing a set of nondominated hyperintervals (i. e., hyperintervals with the smallest characteristics for a particular estimate $\tilde{K}$ of the Lipschitz constant K in (16)). In terms of the geometric LGO approach, this means that the MultK method uses during its work a series of nonsmooth (discontinuous) piecewise quadratic auxiliary functions. The usage of multiple estimates of K simplifies the method scheme (in fact, due to this simplicity, algorithms based on the similar idea of using a variety of estimates of the Lipschitz constant for a Lipschitz objective function are very popular in derivative-free global optimization; see, e. g., the references in [13, 36]) but does not allow one to obtain a meaningful stopping criterion for this method. The MultK method stops when a preset trials budget is depleted and no guarantee on the goodness of a solution is provided unless a very high number of trials is allowed (similar behavior is intrinsic to some widely used nature-inspired metaheuristics; see, e. g., [13, 42, 31]).

The following theorem from [27] establishes the everywhere dense convergence of trial points generated by the MultK method.

Theorem 3.1. *For any vector $\theta \in \Theta$ and any $\delta > 0$ there exist an iteration number $k(\delta) \geq 1$ and a trial point $\theta' = \theta_{i(k)}$, $k > k(\delta)$, generated by the* MultK *method with the infinite trial budget, such that $\|\theta - \theta'\| < \delta$.*

In the SmoothD method, the unique estimate $\tilde{K}$ of the Lipschitz constant K from (15) is used at each iteration to construct smooth piecewise quadratic auxiliary functions as it follows from (16) (see the function $\varphi_i(\tilde{\theta})$ in Figure 2):

$$\tilde{K} = \begin{cases} r \max_i \hat{K}_i, & \text{if } \max_i \hat{K}_i > \xi = 10^{-6}, \\ r\,\xi, & \text{otherwise,} \end{cases} \tag{18}$$

where $r > 1$ is the reliability parameter of the method, ξ is a small positive value (it ensures the correct algorithm execution when the values $\hat{K}_i$ are too small), and $\hat{K}_i$ are calculated as

$$\hat{K}_i = \frac{|2(F(\theta_i^-) - F(\theta_i^+)) + (F'(\theta_i^-) + F'(\theta_i^+))\Delta_i| + D_i}{\Delta_i^2},$$

with $F'(\theta_i^-)$, $F'(\theta_i^+)$ and Δ_i from (17) and

$$D_i = \{[2(F(\theta_i^-) - F(\theta_i^+)) + (F'(\theta_i^-) + F'(\theta_i^+))\Delta_i]^2 + (F'(\theta_i^+) - F'(\theta_i^-))^2 \Delta_i^2\}^{\frac{1}{2}}$$

(the local tuning technique can be also used to estimate local Lipschitz constants for the objective gradient in different sub-regions; see, e. g., [50, 59, 35, 58]).

The reliability parameter r from (18) controlling the estimate $\tilde{K}$ of the Lipschitz constant K influences the algorithm performance. Its higher values increase the method reliability. As this parameter decreases, the search rate increases, but the probability of convergence to a point different from a global minimizer of $F(\theta)$ grows as well. Theoretical and experimental suggestions on an appropriate choice of this parameter are given in [56].

The SmoothD method has its internal probative stopping criterion: the method stops when the volume (or the length of the main diagonal as in [56]) of a hyperinterval with the smallest characteristic becomes smaller than the required accuracy related to a predefined constant $\varepsilon > 0$. In this case, a guarantee on the obtained solution can be given as confirmed by the following theoretical result from [56]:

Theorem 3.2. *For any function $F(\theta)$ with the gradient satisfying the Lipschitz condition (15) with the constant K, $0 < K < \infty$, there exists a value r^* of the reliability parameter r from (18) such that for any $r \geq r^*$ the infinite ($\varepsilon = 0$ in the method stopping criterion) sequence of trial points, generated by the* SmoothD *method during minimization of $F(\theta)$, will converge only to the global minimizers of $F(\theta)$.*

In the next Section, the performance of both the methods while tackling the problem (1)–(5) is illustrated on the benchmark objective functions from Section 2.

4. NUMERICAL EXAMPLES AND SIMULATION STUDY

In this section, we present results of numerical experiments performed with the methods MultK and SmoothD on the benchmark functions from subsection 2.1 to illustrate the performance of LGO algorithms on problems (1)–(5).

Closed admissible domains Θ from (2) were considered for all problems. Since in the case of noise-free benchmark problems the solutions are known ($\theta^* = \theta^0$ in (1)), a particular problem from subsection 2.1 was considered to be solved

Table 1. Solutions to benchmark problems from subsection 2.1 obtained by the MULTK *and* SMOOTHD *methods*

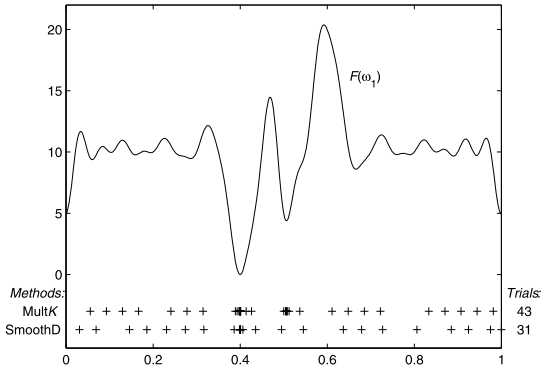
$F(\theta)$	MULTK	SMOOTHD
(a)	$\omega_1^* \approx 0.400000$ # trials = 43	$\omega_1^* \approx 0.400000$ # trials = 31 (12)
(b)	$\omega_1^* \approx 0.400000$ # trials = 170	$\omega_1^* \approx 0.400000$ # trials = 153 (63)
(c)	$\omega_1^* \approx 0.299040$ $\omega_2^* \approx 0.400549$ # trials = 204	$\omega_1^* \approx 0.300412$ $\omega_2^* \approx 0.399177$ # trials = 321 (231)
(d)	$d^* \approx -0.222222$ $\omega^* \approx 0.395062$ $\phi^* \approx 0.310123$ # trials = 1449	$d^* \approx -0.222222$ $\omega^* \approx 0.395062$ $\phi^* \approx 0.290741$ # trials = 4478 (1511)

by the MULTK method if it generated a trial point θ' in an ε-neighborhood of θ^* (precisely, θ' should be inside a hyperinterval with a vertex θ^* having the volume smaller than the volume of the initial hyperinterval Θ from (2) multiplied by ε; see, e. g., [27, 29] for more details). In the case of real-life noisy problems, this method can be stopped (as often adopted in popular metaheuristic algorithms, see, e. g., [31]) when a prescribed number of trials is reached: the current best value is thus taken as a solution to the problem.

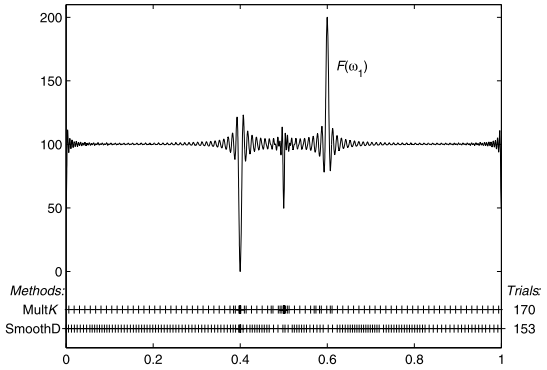
On the contrary, the SMOOTHD method was stopped by its internal stopping rule related to an accuracy coefficient ε, as described in subsection 3.3 (although, observations of the first successful trial were performed for this algorithm too, as for the MULTK method). In order to provide a guarantee on the found solution to a problem, the SMOOTHD method executed a number of iterations after a 'good' (successful) estimate of the global minimum F^* was obtained during the search (see, e. g., [56, 54] for more details).

In the experiments performed, the accuracy coefficient ε was taken at least equal to 10^{-6}, that can be often considered an acceptable accuracy in practice (see, e. g., [29, 31]). The balancing parameter equal to 10^{-4} was used in the MULTK method (this parameter prevents the algorithm from subdividing too small hyperintervals; see, e. g., [13, 44]). The reliability parameter r in the SMOOTHD method was set close to 1.1 (see [56] for the related discussion).

Results of numerical experiments with the two methods on the problems from subsection 2.1 are reported in Table 1 where for all the cases both the found solution and the number of generated trial points are given. For the SMOOTHD method, the numbers of the first successful trials are also indicated in parentheses. It can be seen from this Table that the found parameters were quite accurate and were determined by both the methods within a very limited budget of trials. It is important to highlight that the values in the third column of Table 1 were obtained by the SMOOTHD method after having verified its internal probative stopping criterion, what is significant from the practical point of view.



(a) $F(\omega_1)$ from (6) with $T = 10$

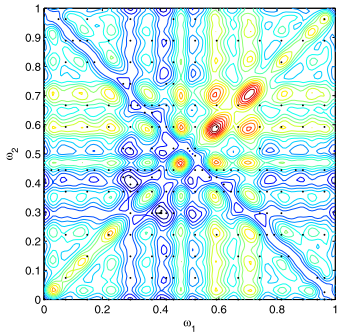


(b) $F(\omega_1)$ from (6) with $T = 100$

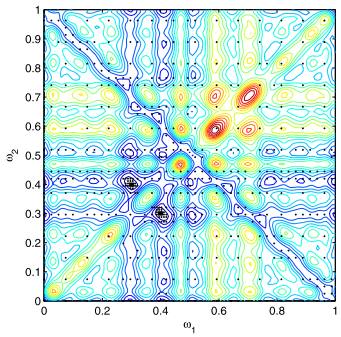
Figure 3. Distribution of trial points when solving one-dimensional problems (a) and (b) $F(\omega_1)$ from subsection 2.1.

Trial points generated by both the methods when solving the benchmark problems from subsection 2.1 are reported in Figures 3–5 together with the objective function contours. Convergence of the sequences of trial points generated by the methods to the global minimizers can be thus evidenced and differences between the two considered stopping criteria can be observed.

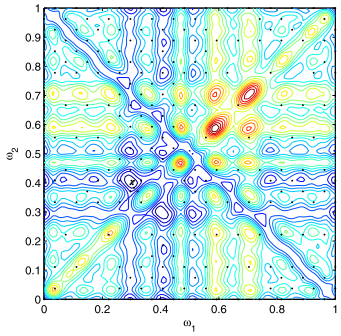
For example, the situations when the MULTK and the SMOOTHD methods generated the first trial point in an ε-neighborhood of the known global minimizer of the two-dimensional function $F(\omega_1, \omega_2)$ from (7) are illustrated in Figures 4(a) and 4(c), respectively. Recall that such an artificial stopping rule (applicable for benchmark problems) cannot be used in practical real-life problems. In order to evaluate the real behaviour of the methods, more trials are needed. For the MULTK, this can be done either by allowing the method to run until achieving some computational budget, or (equivalently) by decreasing the accuracy coefficient ε. For the SMOOTHD method, it is enough to allow the method to stop due to its internal criterion. Both the situations are reported in Figures 4(b) and 4(d), respectively.



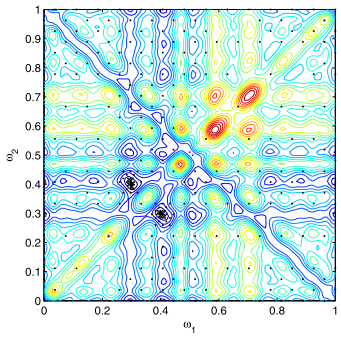
(a) MULTK method with accuracy $\varepsilon = 10^{-6}$, number of the first successful trial is 204



(b) MULTK method with accuracy $\varepsilon = 10^{-7}$, number of the first successful trial is 540

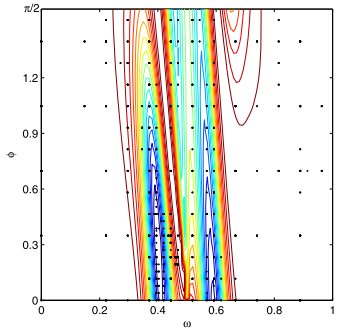


(c) SMOOTHD method with accuracy $\varepsilon = 10^{-6}$, number of the first successful trial is 231

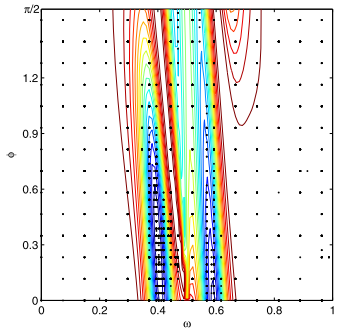


(d) SMOOTHD method with accuracy $\varepsilon = 10^{-6}$, total number of trials is 321

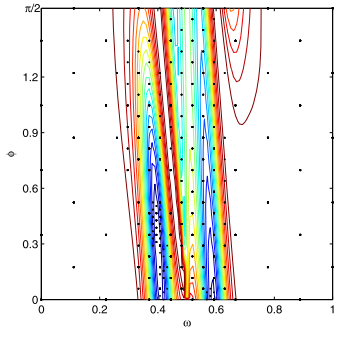
Figure 4. Distribution of trial points when solving two-dimensional problem (c) $F(\omega_1, \omega_2)$ from (7).



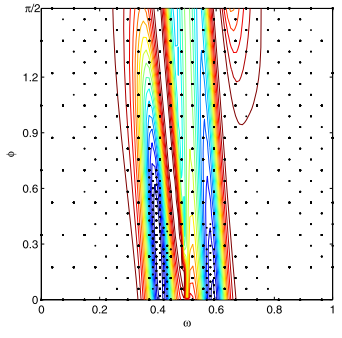
(a) MULTK method with accuracy $\varepsilon = 10^{-6}$, number of the first successful trial is 1449



(b) MULTK method with accuracy $\varepsilon = 10^{-7}$, number of the first successful trial is 4015



(c) SMOOTHD method with accuracy $\varepsilon = 10^{-6}$, number of the first successful trial is 1511



(d) SMOOTHD method with accuracy $\varepsilon = 10^{-6}$, total number of trials is 4478

Figure 5. Distribution of trial points when solving three-dimensional problem (d) $f(d, \omega, \phi)$ from (11).

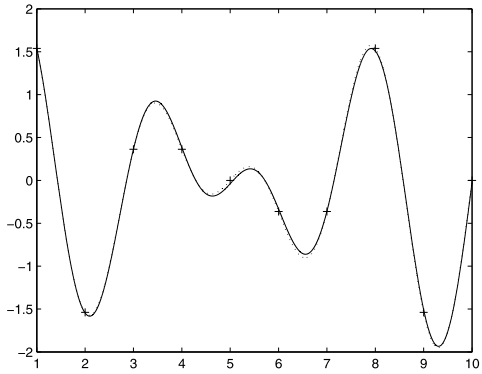
Note that trial points were concentrated around global minimizers $(0.3, 0.4)$ and $(0.4, 0.3)$ of this two-dimensional function $F(\omega_1, \omega_2)$ which is symmetric in this case with respect to permutation of its arguments ω_1 and ω_2. Closer were the frequencies $\omega_1^{(0)}$ and $\omega_2^{(0)}$, less trials were needed to locate the best parameters. LGO algorithms that are able to make use of the symmetries in the objective function (as, e. g., simplicial partitioning methods; see [44, 43, 61]) could be more advantageous for optimizing such particular functions.

Similar conclusions with regard to the methods' performance can be also given in the case of the three-dimensional benchmark function $f(d, \omega, \phi)$ from (11). Its two-dimensional cross-section (ω, ϕ) for the true parameter $d^{(0)} = -0.2$ with trial points generated (and projected to the taken cross-section) by the two methods at their different stages is reported in Figure 5. The importance of the demonstrative phase of the SmoothD method (resulted therefore more reliable algorithm than the Mult$K$ method on the considered problems) can be particularly seen from Figures 5(c) and 5(d).

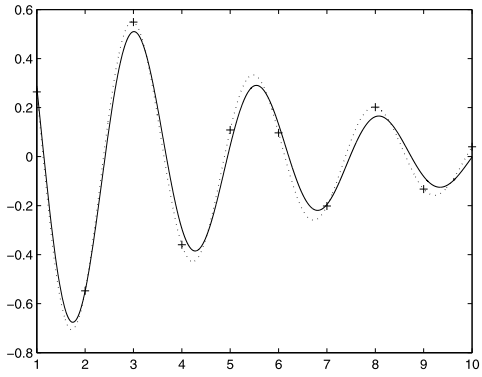
Finally, Figures 6(a) and 6(b) report realizations of observations y_t and the corresponding true and reconstructed signals from (3) in the case of benchmark problems (c) and (d) from subsection 2.1, respectively (the case of benchmark problems (a) and (b) is trivial since their optimal value ω_1^* was determined by the methods exactly up to the seventh decimal point). It can be observed how the reconstructed signal fitted closely the true one. All the optimal parameters produced by the two LGO methods can be (and should be) further improved by a specific local optimization procedure (see, e. g., [13]), thus approaching the true signals even better.

5. CONCLUSIONS

In this paper, a classical parameter estimation problem in nonlinear regression models has been considered as a global optimization problem. Several examples of the objective functions stated to fit sums of dumped sinusoids to series of observations have been particularly examined to illustrate the complexity of this identification problem. For its study, the usage of deterministic global optimization techniques has been proposed, as they can often provide solutions to these difficult problems together with some guaranteed gaps. Particularly, some promising Lipschitz-based methods have been briefly described and successfully applied to determine the solutions to the analyzed problems. A more detailed numerical comparison of the considered deterministic methods with other state-of-the-art techniques (including heuristic nature inspired algorithms widely used in practice) on extended general instances of the parameter estimation problem could be a useful and interesting direction of future research. The problem considered is closely related to Hankel structured low rank approximation, which is known to be a difficult problem. Hence, efficient and robust optimization methods which can tackle problems in this area



(a) True signal $\eta(\omega_1^{(0)}, \omega_2^{(0)})$ from (7), $\omega_1^{(0)} = 0.3, \omega_2^{(0)} = 0.4$, and the estimated reconstructed signal $\eta(\omega_1^*, \omega_2^*)$ with $\omega_1^* \approx 0.300412, \omega_2^* \approx 0.399177$



(b) True signal $\eta(d^{(0)}, \omega^{(0)}, \phi^{(0)})$ from (11), $d^{(0)} = -0.2, \omega^{(0)} = 0.3, \phi^{(0)} = 0.4$, and the estimated reconstructed signal $\eta(d^*, \omega^*, \phi^*)$ with $d^* \approx -0.222222, \omega^* \approx 0.395062, \phi^* \approx 0.290741$

Figure 6. Plots of observations y_t (signed by $+$), true signals (dotted lines), and the reconstructed signals (solid lines) estimated by the SmoothD *method for benchmark problems.*

are desired. It is hoped that this paper has gone some way to describe potential options.

ACKNOWLEDGEMENTS

The idea of this work has arisen during the visit of the second author at the Cardiff School of Mathematics in the period November–December 2014. The authors are grateful to Anatoly Zhigljavsky for productive discussions and friendly attention to their work.

Received 9 June 2015

REFERENCES

[1] Bishop, W. B. and Djurić, P. M. (1996). Model order selection of damped sinusoids in noise by predictive densities. *IEEE Trans. Signal Process.* **44**(3) 611–619.

[2] Bloomfield, P. (2000). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, New York. MR1884963
[3] Breiman, L. and Cutler, A. (1993). A deterministic algorithm for global optimization. *Math. Program.* **58**(1-3) 179–199.
[4] Cadzow, J. (1988). Signal enhancement: a composite property mapping algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **36**(1) 49–62.
[5] Calvin, J. M. and Žilinskas, A. (2005). One-dimensional global optimization for observations with noise. *Comput. Math. Appl.* **50**(1-2) 157–169.
[6] Carnì, D. L. and Fedele, G. (2012). Multi-sine fitting algorithm enhancement for sinusoidal signal characterization. *Comput. Stand. Inter.* **34**(6) 535–540.
[7] Chu, M. T., Funderlic, R. E., and Plemmons, R. J. (2003). Structured low rank approximation. *Linear Algebra Appl.* **366** 157–172. MR1987719
[8] Clausen, J. and Žilinskas, A. (2002). Subdivision, sampling and initialization strategies for simplical branch and bound in global optimization. *Comput. Math. Appl.* **44**(7) 943–955.
[9] Coluccio, L., Eisinberg, A., and Fedele, G. (2007). A Prony-like method for non-uniform sampling. *Signal Process.* **87**(10) 2484–2490.
[10] Costanzo, S., Venneri, I., Di Massa, G., and Borgia, A. (2010). Benzocyclobutene as substrate material for planar millimeter-wave structures: dielectric characterization and application. *J. Infrared Milli Terahz Waves* **31** 66–77.
[11] De Moor, B. (1994). Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Trans. Signal Process.* **42**(11) 3104–3113.
[12] Evtushenko, Yu. G. and Posypkin, M. A. (2013). A deterministic approach to global box-constrained optimization. *Optim. Lett.* **7**(4) 819–829.
[13] Floudas, C. A. and Pardalos, P. M., editors (2009). *Encyclopedia of Optimization (6 Volumes)*. Springer, 2nd edition.
[14] Fowkes, J. M., Gould, N. I. M., and Farmer, C. L. (2013) A branch and bound algorithm for the global optimization of Hessian Lipschitz continuous functions. *J. Global Optim.* **56**(4) 1791–1815. MR3078333
[15] Garnier, H. and Wang, L., editors (2008). *Identification of Continuous-Time Models from Sampled Data*. Springer-Verlag, London.
[16] Gergel, V. P. (1997). A global optimization algorithm for multivariate function with Lipschitzian first derivatives. *J. Global Optim.* **10**(3) 257–281. MR1443089
[17] Gergel, V. P. and Sergeyev, Ya. D. (1997). Sequential and parallel algorithms for global minimizing functions with Lipschitzian derivatives. *Comput. Math. Appl.* **37**(4-5) 163–179.
[18] Gillard, J. W. (2010). Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Stat. Interface* **3**(3) 335–343. MR2720137
[19] Gillard, J. W. and Zhigljavsky, A. (2011). Analysis of structured low rank approximation as an optimisation problem. *Informatica* **22**(4) 489–505.
[20] Gillard, J. W. and Zhigljavsky, A. (2013). Optimization challenges in the structured low rank approximation problem. *J. Global Optim.* **57** 733–751.
[21] Gillard, J. W. and Zhigljavsky, A. (2015). Stochastic algorithms for solving structured low-rank matrix approximation problems. *Commun. Nonlinear Sci. Numer. Simulat.* **21** 70–88.
[22] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC, Boca Raton. MR1823012
[23] Grishagin, V. A. and Strongin, R. G. (1984). Optimization of multi-extremal functions subject to monotonically unimodal constraints. *Eng. Cybernetics* **22**(5) 117–122.
[24] Kvasov, D. E. (2008). Multidimensional Lipschitz global optimization based on efficient diagonal partitions. *4OR – Quart. J. Oper. Res.* **6**(4) 403–406.
[25] Kvasov, D. E. and Sergeyev, Ya. D. (2009). A univariate global search working with a set of Lipschitz constants for the first derivative. *Optim. Lett.* **3**(2) 303–318.
[26] Kvasov, D. E. and Sergeyev, Ya. D. (2012). Univariate geometric Lipschitz global optimization algorithms. *Numer. Algebra Contr. Optim.* **2**(1) 69–90.
[27] Kvasov, D. E. and Sergeyev, Ya. D. (2012). Lipschitz gradients for global optimization in a one-point-based partitioning scheme. *J. Comput. Appl. Math.* **236**(16) 4042–4054.
[28] Kvasov, D. E. and Sergeyev, Ya. D. (2013). Lipschitz global optimization methods in control problems. *Automat. Remote Contr.* **74**(9) 1435–1448.
[29] Kvasov, D. E. and Sergeyev, Ya. D. (2015). Deterministic approaches for solving practical black-box global optimization problems. *Adv. Eng. Softw.* **80** 58–66.
[30] Kvasov, D. E., Menniti, D., Pinnarelli, A., and Sergeyev, Ya. D. (2008). Tuning fuzzy power-system stabilizers in multi-machine systems by global optimization algorithms based on efficient domain partitions. *Electr. Power Syst. Res.* **78**(7) 1217–1229.
[31] Kvasov, D. E., Mukhametzhanov, M. S., and Sergeyev, Ya. D. (2014). A numerical comparison of some deterministic and nature-inspired algorithms for black-box global optimization. In: Topping, B. H. V. and Iványi, P., editors. *Proceedings of the Twelfth International Conference on Computational Structures Technology*. Civil-Comp Press, Stirlingshire, United Kingdom, article 169 (doi: 10.4203/ccp.106.169).
[32] Kvasov, D. E., Pizzuti, C., and Sergeyev, Ya. D. (2003). Local tuning and partition strategies for diagonal GO methods. *Numer. Math.* **94**(1) 93–106. MR1971214
[33] Kvasov, I. E. and Petrov, I. B. (2012). High-performance computer simulation of wave processes in geological media in seismic exploration. *Comput. Math. Math. Phys.* **52**(2) 302–313.
[34] Lemmerling, P. and Van Huffel, S. (2001). Analysis of the structured total least squares problem for Hankel/Toeplitz matrices. *Numer. Algorithms* **27**(1) 89–114.
[35] Lera, D. and Sergeyev, Ya. D. (2013) Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives. *SIAM J. Optim.* **23**(1) 508–529.
[36] Lera, D. and Sergeyev, Ya. D. (2015) Deterministic global optimization using space-filling curves and multiple estimates of Lipschitz and Hölder constants. *Commun. Nonlinear Sci. Numer. Simulat.* **23**(1-3) 328–342.
[37] Li, Y., Liu, K. J. R. and Razavilar, J. (1997). A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation. *IEEE Trans. Signal Process.* **45**(2) 481–486.
[38] Liuzzi, G., Lucidi, S., and Piccialli, V. (2010). A partition-based global optimization algorithm. *J. Global Optim.* **48**(1) 113–128. MR2678849
[39] Markovsky, I. (2008). Structured low-rank approximation and its applications. *Automatica* **44** 891–909. MR2530933
[40] Markovsky, I. (2010). Bibliography on total least squares and related methods. *Stat. Interface* **3**(3) 329–334. MR2720136
[41] Mockus, J. (1989). *Bayesian Approach to Global Optimization*. Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht. MR1111483
[42] Pardalos, P. M. and Romeijn, H. E., editors (2002). *Handbook of Global Optimization* **1**. Kluwer Academic Publishers, Dordrecht.
[43] Paulavičius, R. and Žilinskas, J. (2014). *Simplicial Global Optimization*. SpringerBriefs in Optimization, Springer, New York.
[44] Paulavičius, R., Sergeyev, Ya. D., Kvasov, D. E., and Žilinskas, J. (2014). Globally-biased Disimpl algorithm for expensive global optimization. *J. Global Optim.* **59**(2-3) 545–567. MR3216692
[45] Pintér, J. D. (1996). *Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementa-*

tions and Applications). Kluwer Academic Publishers, Dordrecht. MR1374104

[46] Pollock, D. S. G. (1999). *A Handbook of Time Series Analysis, Signal Processing, and Dynamics*. Academic Press Inc., London. MR1737528

[47] Schoen, F. (2002). Two-phase methods for global optimization. In: Pardalos, P. M. and Romeijn, H. E., editors. *Handbook of Global Optimization* **2**. Kluwer Academic Publishers, Dordrecht, 151–177. MR1919533

[48] Sergeyev, Ya. D. (1995). An information global optimization algorithm with local tuning. *SIAM J. Optim.* **5**(4) 858–870. MR1358808

[49] Sergeyev, Ya. D. (1998). On convergence of "Divide the Best" global optimization algorithms. *Optimization* **44**(3) 303–325. MR1775619

[50] Sergeyev, Ya. D. (1998). Global one-dimensional optimization using smooth auxiliary functions. *Math. Program.* **81** 127–146. MR1617756

[51] Sergeyev, Ya. D. (2000). An efficient strategy for adaptive partition of N-dimensional intervals in the framework of diagonal algorithms. *J. Optim. Theory Appl.* **107**(1) 145–168. MR1800933

[52] Sergeyev, Ya. D. and Grishagin, V. A. (1994). Sequential and parallel algorithms for global optimization. *Optim. Methods Softw.* **3**(1-3) 111–124.

[53] Sergeyev, Ya. D. and Kvasov, D. E. (2006). Global search based on efficient diagonal partitions and a set of Lipschitz constants. *SIAM J. Optim.* **16**(3) 910–937.

[54] Sergeyev, Ya. D. and Kvasov, D. E. (2008). *Diagonal Global Optimization Methods*. Fizmatlit, Moscow (in Russian).

[55] Sergeyev, Ya. D. and Kvasov, D. E. (2011). Lipschitz global optimization. In: Cochran, J. J., et al., editors. *Wiley Encyclopedia of Operations Research and Management Science* **4**. Wiley, New York, 2812–2828.

[56] Sergeyev, Ya. D. and Kvasov, D. E. (2015). A deterministic global optimization using smooth diagonal auxiliary functions. *Commun. Nonlinear Sci. Numer. Simulat.* **21** 99–111.

[57] Sergeyev, Ya. D. and Kvasov, D. E. (2015). On deterministic diagonal methods for solving global optimization problems with Lipschitz gradients. In: Migdalas, A. and Karakitsiou, A., editors. *Optimization, Control, and Applications in the Information Age*. Springer Proceedings in Matematics and Statistics **130**, Springer, Switzerland, 319–337.

[58] Sergeyev, Ya. D., Strongin, R. G., and Lera, D. (2013). *Introduction to Global Optimization Exploiting Space-Filling Curves*. SpringerBriefs in Optimization, Springer, New York. MR3113120

[59] Strongin, R. G. and Sergeyev, Ya. D. (2000). *Global Optimization with Non-Convex Constraints: Sequential and Parallel Algorithms*. Kluwer Academic Publishers, Dordrecht (2nd ed., 2012; 3rd ed., 2014, Springer, New York).

[60] Zhigljavsky, A. (1991). *Theory of Global Random Search*. Kluwer Academic Publishers, Dordrecht. MR1187048

[61] Zhigljavsky, A. and Žilinskas, A. (2008). *Stochastic Global Optimization*. Springer Optimization and Its Applications **9**, Springer, New York.

[62] Žilinskas, A. (1982). Axiomatic approach to statistical models and their use in multimodal optimization theory. *Math. Program.* **22**(1) 104–116. MR0637534

[63] Žilinskas, A. (2010). On similarities between two models of global optimization: statistical models and radial basis functions. *J. Global Optim.* **48**(1) 173–182. MR2678853

J. W. Gillard
Cardiff School of Mathematics
Cardiff University
United Kingdom
E-mail address: GillardJW@Cardiff.ac.uk

D. E. Kvasov
DIMES
University of Calabria
Rende (CS)
Italy
and
Lobachevsky State University of Nizhni Novgorod
Russia
E-mail address: kvadim@dimes.unical.it