CHALMERS |  UNIVERSITY OF GOTHENBURG

PREPRINT 2011:16

A Minimax Strategy for Global Optimization

STEFAN JAKOBSSON
PETER LINDROTH
ANN-BRITH STRÖMBERG

Department of Mathematical Sciences
Division of Mathematics
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg Sweden 2011

Preprint 2011:16

A Minimax Strategy for Global Optimization

Stefan Jakobsson, Peter Lindroth,
Ann-Brith Strömberg

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg, Sweden
Gothenburg, May 2011

Preprint 2011:16
ISSN 1652-9715

Matematiska vetenskaper
Göteborg 2011

A Minimax Strategy for Global Optimization

Stefan Jakobsson* Peter Lindroth†‡ Ann-Brith Strömberg‡

Abstract

A computationally expensive multi-modal optimization problem is considered. After an optimization loop it is desirable that the optimality gap, i.e., the difference between the best value obtained and the true optimum, is as small as possible. We define the concept of *maximum loss* as being the supremum of the optimality gaps over a set of functions, i.e., the largest possible optimality gap assuming that the unknown objective function belongs to a certain set of functions. The minimax strategy for global optimization is then to—at each iteration—choose a new evaluation point such that the maximum loss is decreased as much as possible. This strategy is in contrast to the maximum gain strategy, which is utilized in several common global optimization algorithms, and the relation between these strategies is described. We investigate how to implement the minimax strategy for the Lipschitz space of functions on box-constrained domains. Several problems are revealed. For example, to obtain uniqueness of the set of solutions to the minimax problem it is often necessary to decrease the domain such that the problem is more localized. We propose a number of algorithmic schemes, based on sequential linearization, to solve the different subproblems that appear. The algorithms are illustrated by numerical examples. We conclude that the minimax strategy is promising for global optimization when the main concern is to guarantee that the resulting solution is near-optimal.

Keywords: global optimization, minimax optimization, Lipschitz optimization

1 Introduction

Many real-world optimization problems in, e.g., engineering design, can be formulated as

$$f^* = \max_{\mathbf{x}\in\Omega} f(\mathbf{x}), \tag{1}$$

where $f : \Omega \to \mathbb{R}$ is a computationally expensive multi-modal function which is to be optimized over its domain $\Omega \subseteq \mathbb{R}^n$, and which is assumed to be continuous

*Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Chalmers Science Park, SE-412 88 Gothenburg, Sweden. E-mail: stefan.jakobsson@fcc.chalmers.se

†Volvo 3P, Chassis & Vehicle Dynamics, Chassis Strategies & Vehicle Analysis, SE-405 08, Gothenburg, Sweden. E-mail: peter.lindroth@volvo.com

‡Mathematical Sciences, Chalmers University of Technology, and Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: {peter.lindroth, anstr}@chalmers.se

and bounded from above. We consider the case when derivative information for the function is missing, usually then f is denoted as a *black-box* function. This can be the case, e.g., when $f(\mathbf{x})$ is the outcome of a computer simulation at $\mathbf{x}$ using a software without access to the source code, and / or when automatic differentiating techniques are inappropriate due to noise in the function or simply that the extra evaluations needed for the differentiation are better spent elsewhere in the domain.

Since, in a practical situation, one can afford to make only a finite number of function evaluations, only an approximate maximum value to the optimization problem (1) can be expected to be found. The difference between the true maximum value f^* and the obtained approximate value can be seen as a loss. In this article we apply concepts from statistical decision theory to global optimization and we begin by a classification of global optimization algorithms with respect to these concepts. We consider the choice of a new evaluation point in the optimization loop as a decision. Some common decision criteria in statistical decision theory are to

A. maximize the absolute gain,

B. maximize the expected gain,

C. minimize the expected loss, and

D. minimize the absolute loss.

This implicitly assumes a space of scenarios and a space of possible decisions and it also requires precise definitions of gain and loss. For criteria B and C some sort of probability model is also required. In Jones [9] and Jones et al. [10], this is implicitly imposed by using the maximum likelihood method to create a response surface by the kriging method. This allows to set up optimization algorithms for maximizing the probability for improvement and maximizing the expected gain.

In this article we consider the criterion A, to maximize the absolute gain, and the criterion D, to minimize the absolute loss, with a special emphasis on the latter. To our knowledge, the strategy of minimizing the absolute loss has not been applied to global optimization before which makes this study particularly interesting. The concern is to solve a minimax problem to find a new evaluation point for the outer optimization problem (1). In the sequel we drop the word 'absolute' in criterion A to simplify the terminology, and we replace the word 'absolute' with 'maximum' in criterion D to emphasize that we have the maximum over a set of scenarios, or, in our case a set of functions.

The ouline of the article is as follows: We begin in Section 2 by defining the maximum loss with respect to a function set and to subsets of the computational domain called the reference and evaluation sets. With this concept we present the two strategies for global optimization mentioned above. At this point we also classify some existing algorithms as maximum gain strategies. Upper and lower envelope functions for the function and evaluation sets are also introduced, and it is seen that the maximum gain solution coincides with with the maximum point of the upper envelope function. In Section 3 we apply this framework to function classes in the Lipschitz space. Since Lipschitz functions are very robust, e.g., the supremum of a

bounded set of Lipschitz functions is a Lipschitz function, the upper and lower envelope functions have explicit representations in this case, and it turns out that it is possible to formulate both criteria A and D as well-posed optimization problems using the upper envelope function. In one dimension for the Lipschitz space, criterion A is known as the Piyavskii–Shubert algorithm [17, 18] (the authors discovered the algorithm independently).

Because of the special form of the objective function in the minimax problem in the Lipschitz case we study in Section 4 the "two-functions minimax problem". We prove, under some mild conditions, that the function to be minimized in this problem is quasi-convex but not strictly quasi-convex, implying that the problem may not have a unique solution. The cure to enforce uniqueness is to reduce the reference domain for the minimax problem. A characterization of the minimax solution in terms of control points is also given. In Section 5 we present algorithms for solving the different subproblems required to find the new evaluation point according to the minimax strategy. To find the maximum gain solution we combine a sequential linear programming technique with a Branch-and-Bound scheme. The algorithm used to find the minimax solution is also based on sequential linear programming with the idea to iteratively improve the approximations of the control points and the solution itself. The algorithms are illustrated on some numerical examples in Section 6. The article ends with some conclusions and outlooks.

2 Problem formulation

The purpose of this section is to formalize the concept of maximum loss for an optimization problem and to relate it to the concept of maximum gain. In a first subsection we define the maximum loss in a general setting and in the subsequent subsection we present strategies for the selection of new points utilizing a maximum loss or a maximum gain strategy, respectively.

2.1 The maximum loss in a general setting

We assume that any objective function f considered for the optimization problem (1) belong to some Banach space $\mathcal{B}(\Omega)$ of continuous functions over the domain Ω. Furthermore we assume that Ω is a compact set so that the maximum value of f is attained. The maximum loss is defined according to the following.

DEFINITION 2.1 *Let $\mathcal{F}$ be a closed subset of $\mathcal{B}(\Omega)$, called the* function set. *Let D be a closed subset of Ω, called the* reference set *(possibly the entire Ω), and let E be a closed subset of Ω, called the* evaluation set. *The* maximum loss *is a function defined for a triplet $(\mathcal{F}, D, E)$ given by*

$$\mathcal{ML}(\mathcal{F}, D, E) = \sup_{s \in \mathcal{F}} \left\{ \max_{\mathbf{x} \in D} s(\mathbf{x}) - \max_{\mathbf{x}' \in E} s(\mathbf{x}') \right\}. \tag{2}$$

There are some immediate consequences of this definition. The maximum loss increases monotonically with the function and the reference sets and decreases when the evaluation set is increased. In addition, the maximum loss is non-negative whenever $E \subseteq D$. We summarize these properties in a proposition.

PROPOSITION 2.2 *The maximum loss has the following properties*

1. *If* $\mathcal{F} \subseteq \mathcal{F}'$ *then*
$$\mathcal{ML}(\mathcal{F}, D, E) \leq \mathcal{ML}(\mathcal{F}', D, E). \tag{3}$$
2. *If* $D \subseteq D'$ *then*
$$\mathcal{ML}(\mathcal{F}, D, E) \leq \mathcal{ML}(\mathcal{F}, D', E).$$
3. *If* $E \subseteq E'$ *then*
$$\mathcal{ML}(\mathcal{F}, D, E) \geq \mathcal{ML}(\mathcal{F}, D, E'). \tag{4}$$
4. *If* $E \subseteq D$ *then*
$$\mathcal{ML}(\mathcal{F}, D, E) \geq 0.$$

As will be shown, it is convenient to rewrite the expression of maximum loss in the following way by interchanging to order of the $\max$ operators:

$$\mathcal{ML}(\mathcal{F}, D, E) = \max_{\mathbf{x} \in D} \sup_{s \in \mathcal{F}} \left\{ s(\mathbf{x}) - \max_{\mathbf{x}' \in E} s(\mathbf{x}') \right\}. \tag{5}$$

For (5) to hold true we have to assume that the function set $\mathcal{F}$ is such that the supremum over $\mathcal{F}$ is itself a continuous function, so that its maximum value over D is attained. This assumption is made from now on. At each point $\mathbf{x} \in \Omega$, there exist upper and lower bounds for the functions in the set $\mathcal{F}$. We introduce the upper and lower envelope functions representing these bounds.

DEFINITION 2.3 *The* upper envelope function $\mathcal{U}_{\mathcal{F}} : \Omega \to \mathbb{R}$ *and the* lower envelope function $\mathcal{L}_{\mathcal{F}} : \Omega \to \mathbb{R}$ *for the subset* $\mathcal{F} \subseteq \mathcal{B}(\Omega)$ *are defined by*

$$\mathcal{U}_{\mathcal{F}}(\mathbf{x}) = \sup_{s \in \mathcal{F}} s(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{6}$$
$$\mathcal{L}_{\mathcal{F}}(\mathbf{x}) = \inf_{s \in \mathcal{F}} s(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{7}$$

Suppose that we are solving (1) and that N evaluations of the objective function f have been made at the points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \Omega$ with the objective values $f(\mathbf{x}_k) = f_k$, $k = 1, \ldots, N$. The evaluation set is the union of all points at which the function has been evaluated;
$$E := E_N = \cup_{k=1}^{N} \mathbf{x}_k. \tag{8}$$
For the function set $\mathcal{F}$ it must hold that
$$\mathcal{F} \subseteq \{s \in \mathcal{B}(\Omega) : s(\mathbf{x}_k) = f_k, k = 1, \ldots, N\}. \tag{9}$$

Since all $s \in \mathcal{F}$ are completely determined on the evaluation set E_N, the following relation between the maximum loss (5) and the upper envelope function holds:

$$\mathcal{ML}(\mathcal{F}, D, E_N) = \max_{\mathbf{x}\in D} \mathcal{U}_{\mathcal{F}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k. \tag{10}$$

If the optimization of (1) was terminated at this stage, the best point in E_N would be chosen as an approximation of the optimum. The maximum loss is thus the difference between the best value that can be obtained (the maximum of the upper envelope function) and the best value obtained so far, i.e., the worst case optimality gap assuming that f is a function in $\mathcal{F}$. A termination criterion for the optimization of (1) is therefore that the maximum loss is sufficiently close to zero.

To choose an appropriate function set it is reasonable to assume that the objective function has a norm that is bounded by some constant $c \in \mathbb{R}_+$ (as a comment, without putting any restriction on f, it is impossible to guarantee convergence within any $\varepsilon > 0$ in finite time). Therefore we choose the function set as

$$\mathcal{F}_{N,c} = \{s \in \mathcal{B}(\Omega) : \|s\| \leq c,\, s(\mathbf{x}_k) = f_k,\, k = 1, \ldots, N\}, \tag{11}$$

i.e., all functions in $\mathcal{B}(\Omega)$ interpolating f at the points $\{\mathbf{x}_k\}_{k=1}^N$ and having a norm that is less than or equal to c. The value of c has a large impact on the value of the maximum loss and also on the optimization strategies to be studied, this is further discussed in Section 6.

2.2 Strategies for choosing new evaluation points

In the optimization of (1), new points $\mathbf{y} \in \Omega$ are included in the evaluation set in each iteration. An iteration consists of the two steps to

1. choose a new evaluation point $\mathbf{y} \in \Omega$, and
2. evaluate the objective function at $\mathbf{y}$, i.e., compute $f(\mathbf{y})$.

In each step the maximum loss decreases but for different reasons. In Step 1 a new point is added to the evaluation set, $E \to E \cup \{\mathbf{y}\}$, and thereby the maximum loss is reduced according to (4). In Step 2 we impose a new restriction on the set of functions, $\mathcal{F} \to \mathcal{F} \cap \{s : s(\mathbf{y}) = f(\mathbf{y})\}$, reducing the maximum loss according to (3). What can be controlled by an optimization algorithm is Step 1. Symbolically we describe it as follows. The function and evaluation sets $\mathcal{F}_k$ and E_k at iteration k satisfy the respective inclusion relations

$$\mathcal{F}_k \supseteq \mathcal{F}_{k+1} \quad \text{and} \quad E_k \subseteq E_{k+1}. \tag{12}$$

The properties (4) and (3) imply the inequalities

$$\mathcal{ML}(\mathcal{F}_k, D, E_k) \geq \mathcal{ML}(\mathcal{F}_k, D, E_{k+1}) \geq \mathcal{ML}(\mathcal{F}_{k+1}, D, E_{k+1}). \tag{13}$$

We define two alternative strategies for global optimization of which at least the first is well-known and used, but, however, not described in this terminology. The main idea of this article is to analyze and to develop methods for the second one.

DEFINITION 2.4 (Maximum gain strategy) *In the* maximum gain *strategy, i.e., criterion A on page 2, the new evaluation point* $\hat{\mathbf{x}} \in \Omega$ *is chosen such that*

$$\hat{\mathbf{x}} \in \arg\max_{\mathbf{x}\in\Omega} \mathcal{U}_{\mathcal{F}}(\mathbf{x}). \tag{14}$$

DEFINITION 2.5 (Minimize the maximum loss strategy (minimax)) *In the* minimize the maximum loss *strategy (the* minimax *strategy from now on), i.e., criterion D on page 2, the new evaluation point* $\mathbf{y}^* \in D \subseteq \Omega$ *is chosen such that*

$$\mathbf{y}^* \in \arg\min_{\mathbf{y}\in D} \mathcal{ML}(\mathcal{F}, D, E \cup \{\mathbf{y}\}). \tag{15}$$

This strategy can be extended to parallel evaluation by choosing several new evaluation points $\mathbf{y}_k^* \in D \subseteq \Omega$*,* $k = 1, \ldots, K$*, such that*

$$\{\mathbf{y}_k^*\}_{k=1}^K \in \arg\min_{\mathbf{y}_k\in D, k=1,\ldots,K} \mathcal{ML}(\mathcal{F}, D, E \cup \{\mathbf{y}_k\}_{k=1}^K). \tag{16}$$

In the maximum gain strategy a new evaluation point is chosen such that the new function value potentially can be as large as possible in each step. This strategy is used by several global optimization algorithms. For example, Piyavskii's and Shubert's Lipschitz optimization algorithm [17, 18] is of this type with function sets of the form (11). This algorithm was later extended to multiple dimensions by [15]. See also [8, 13, 14]. In [7], Gutmann introduces a global optimization algorithm based on interpolation with radial basis functions (RBF). A Hilbert space of functions, called the native space, is associated to each RBF. One variant of Gutmann's algorithm can be interpreted in the maximum gain setting with function sets of the form (11) in the Native space norm (the standard variant of the algorithm is similar but then so called target values are used instead).

In economics, the maximum gain strategy might be called the gamblers choice because of its potential high reward but also high risks (the span between the lower and upper envelope functions might be wide). In optimization the drawback of this strategy is that new evaluation points are often chosen on the boundary of the domain which—especially in higher dimensions—can imply that evaluations are wasted through excessive exploration of the boundary.

The minimax strategy is more conservative. It is less likely to put new evaluation points on the boundary since the strategy tries to balance the areas with high loss. For the Lipschitz case with convex domains, which is explored in later sections, new evaluation points may be found on the boundary, but never at extreme points of the domain (points which do not lie in any open line segment joining two points in the domain).

The starting point for the minimax strategy is to find the maximum gain solution (14), since this is also the point where the current maximum loss is attained. The introduction of the reference set D, which might be a strict subset of Ω, in the definitions of the maximum loss and of the minimax strategy is made for technical reasons. A suitable choice of D may avoid non-uniqueness of the optimal solution in the resulting minimax problem, which otherwise may occur. Examples of non-uniqueness will be presented in Section 4.1.

In some situations it is interesting to compute the objective function at several evaluation points simultaneously, e.g., if several simulations can be run in parallel. This can be handled in the minimax strategy by minimizing over a fixed number of points (see (16)) , although this obviously increases the complexity of the strategy. It is in the end of the optimization procedure that the maximum loss should be a small as possible, not in each particular iteration, which also motivates the inclusion of more points in the minimax problem. To consider only one additional point at a time can be seen as a greedy version of this general strategy.

The main drawback of the minimax strategy is its complexity. In some situations the resulting minimax problem (15) might even be a harder problem to solve than the original problem (1). This will, however, not be the case when the evaluation of the objective function in (1) is sufficiently computationally expensive.

3 Maximum loss in the Lipschitz space

In the former section the maximum loss was defined for a general function class. A common assumption in global optimization (particularly in simulation-based optimization) is that the objective function is Lipschitz continuous. In this section we analyze the case when the function class consists of Lipschitz continuous function. We will see that this simplifies the general expressions considerably.

A function $f : \Omega \to \mathbb{R}$ is said to be *Lipschitz* continuous if there exists a constant $c \geq 0$ such that

$$|f(\mathbf{x}') - f(\mathbf{x}'')| \leq c\|\mathbf{x}' - \mathbf{x}''\|, \qquad \mathbf{x}', \mathbf{x}'' \in \Omega. \tag{17}$$

The smallest such constant c is called the *Lipschitz norm* of the function and is denoted by $\|f\|_{\text{Lip}(\Omega)}$. Returning to the optimization setting where the function f has been evaluated at the points $\{\mathbf{x}_k\}_{k=1}^N \subseteq \Omega$, the Lipschitz norm can be estimated by the lower bound given by the inequality

$$\|f\|_{\text{Lip}(\Omega)} \geq \max_{k,\ell=1,\ldots,N,\, k\neq\ell} \frac{|f_k - f_\ell|}{\|\mathbf{x}_k - \mathbf{x}_\ell\|}. \tag{18}$$

The right hand side of (18) can be interpreted as the Lipschitz norm of f on the discrete evaluation set $E_N = \cup_{k=1}^N \mathbf{x}_k$. One important result on Lipschitz functions is Kirszbraun's Theorem ([11]), which states that it is always possible to extend a Lipschitz function on a subdomain to a larger domain, containing the subdomain, without increasing the Lipschitz norm. In particular, this implies that the set $\mathcal{F}_{N,c}$ defined in (11) is non-empty whenever

$$c \geq \max_{k,\ell=1,\ldots,N,\, k\neq\ell} \frac{|f_k - f_\ell|}{\|\mathbf{x}_k - \mathbf{x}_\ell\|}. \tag{19}$$

The next proposition proves a representation formula for the upper envelope function (6) in the Lipschitz case, showing that the upper envelope function then has an explicit expression.

PROPOSITION 3.1 *For the Lipschitz space we can write the upper envelope function for the function set $\mathcal{F}_{N,c}$ as*

$$\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) = \min_{k=1,\ldots,N} \{f_k + c\|\mathbf{x} - \mathbf{x}_k\|\}, \quad \mathbf{x} \in \Omega. \tag{20}$$

Proof From the definitions of $\mathcal{U}_{\mathcal{F}}$ in (6) and of $\mathcal{F}_{N,c}$ in (11) it follows that

$$\begin{aligned}
\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) &= \sup_{s\in\mathcal{F}_{N,c}} s(\mathbf{x}) \\
&= \sup_{s\in\mathcal{B}(\Omega)} \{s(\mathbf{x}) : \|s\| \le c,\, s(\mathbf{x}_k) = f_k,\, k = 1,\ldots,N\} \\
&= \sup_{s\in\mathcal{B}(\Omega)} \{s(\mathbf{x}) : s(\mathbf{x}) \le f_k + c\|\mathbf{x} - \mathbf{x}_k\|,\, k = 1,\ldots,N\} \\
&= \sup_{s\in\mathcal{B}(\Omega)} \{s(\mathbf{x}) : s(\mathbf{x}) \le \min_{k=1,\ldots,N}\{f_k + c\|\mathbf{x} - \mathbf{x}_k\|\}\} \\
&= \sup_{s\in\mathcal{B}(\Omega)} \{\min_{k=1,\ldots,N}\{f_k + c\|\mathbf{x} - \mathbf{x}_k\|\}\} \\
&= \min_{k=1,\ldots,N} \{f_k + c\|\mathbf{x} - \mathbf{x}_k\|\}.
\end{aligned}$$

∎

The right hand side of (20) is also included, e.g., in [17, 18]. We can now express the objective function of the minimax problem explicitly in terms of the upper envelope function instead of considering the whole space of functions $\mathcal{F}$. This is a main result and is the basis for our implementation of the algorithm.

THEOREM 3.2 *Define h to be the objective function in the minimax problem* (15)*, i.e.,*

$$h(\mathbf{y}) := \mathcal{ML}(\mathcal{F}_{N,c}, D, E_N \cup \{\mathbf{y}\}), \quad \mathbf{y} \in D,$$

for the evaluation set E_N and the function set $\mathcal{F}_{N,c}$ as defined by (8) *and* (11)*, respectively, with the Lipschitz norm c. This objective function also has the explicit representation*

$$h(\mathbf{y}) = \max_{\mathbf{x}\in D} \min \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x} - \mathbf{y}\| \right\}, \quad \mathbf{y} \in D. \tag{21}$$

Moreover, unless $h(\mathbf{y})$ is constant over D, the set of optimal points for the minimax problem (15) *satisfies*

$$\arg\min_{\mathbf{y}\in D} \mathcal{ML}(\mathcal{F}_{N,c}, D, E_N \cup \{\mathbf{y}\}) \subseteq \left\{ \mathbf{x} \in D : \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) > \max_{k=1,\ldots,N} f_k \right\}. \tag{22}$$

The proof of Theorem 3.2 relies on the following lemma, whose proof is postponed to Appendix A.

LEMMA 3.3 *Let E_N and $\mathcal{F}_{N,c}$ be defined by* (8) *and* (11)*, respectively, with the Lipschitz norm. Then, for all $\mathbf{x}, \mathbf{y} \in \Omega$ it holds that*

$$\max_{s\in\mathcal{F}_{N,c}} \left\{ s(\mathbf{x}) - \max_{\mathbf{x}'\in E_N\cup\{\mathbf{y}\}} s(\mathbf{x}') \right\} = \min \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x} - \mathbf{y}\| \right\}. \tag{23}$$

Proof of Theorem 3.2 The first part of the theorem, i.e., the equivalence (21), follows immediately from Lemma 3.3 by utilizing the reformulation (5) of the maximum loss.

For the second part, i.e., the inclusion (22), suppose that we have $\mathbf{y}' \in D$ such that $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}') \leq \max_{k=1,\ldots,N} f_k$. The intention is to show that such a $\mathbf{y}'$ cannot be optimal for the problem on the left hand side of (22). Let $\hat{\mathbf{x}} \in \arg\max_{\mathbf{x}\in D} \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x})$. From the Lipschitz property of $\mathcal{U}_{\mathcal{F}_{N,c}}$, it then follows that

$$c\|\hat{\mathbf{x}} - \mathbf{y}'\| \geq \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}') \geq \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_{k=1,\ldots,N} f_k.$$

Thus,

$$\begin{aligned} h(\mathbf{y}') &= \max_{\mathbf{x}\in D} \min \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x} - \mathbf{y}'\| \right\} \\ &\geq \min \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_{k=1,\ldots,N} f_k,\, c\|\hat{\mathbf{x}} - \mathbf{y}'\| \right\} = \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_{k=1,\ldots,N} f_k. \end{aligned}$$

Furthermore, for any $\mathbf{y} \in D$ it holds that

$$\begin{aligned} h(\mathbf{y}) &= \max_{\mathbf{x}\in D} \min \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x} - \mathbf{y}\| \right\} \\ &\leq \max_{\mathbf{x}\in D} \left\{ \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k \right\} = \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_{k=1,\ldots,N} f_k, \end{aligned}$$

meaning that any $\mathbf{y} \in D$ is at least as good as $\mathbf{y}'$. Since h is non-constant over D, there must exist a $\mathbf{y} \in D$ with $h(\mathbf{y}) < h(\mathbf{y}')$, implying that

$$\mathbf{y}' \notin \arg\min_{\mathbf{y}\in D} \mathcal{ML}(\mathcal{F}_{N,c}, D, E_N \cup \{\mathbf{y}\}),$$

and the result follows. ∎

The upper envelope function and the objective function in the minimax problem are illustrated in Figure 1 for a one-dimensional case, in which the minimax function is computed for the reference set $D = [0, 1]$. One can see in the right subfigure that h in this case does not have a unique minimum. We will study this uniqueness issue further in the next section.

4 Optimality conditions for the minimax problem

The purpose of this section is to study optimality conditions for minimax problems where the objective function is a generalization of the resulting maximum loss from the Lipschitz case; these are later exploited in the proposed solution methods.

Theorem 3.2 shows that the loss function h in the minimax problem (15) for the Lipschitz case has a special and simple representation. In the following definition we make a generalization of this type of minimax problem.

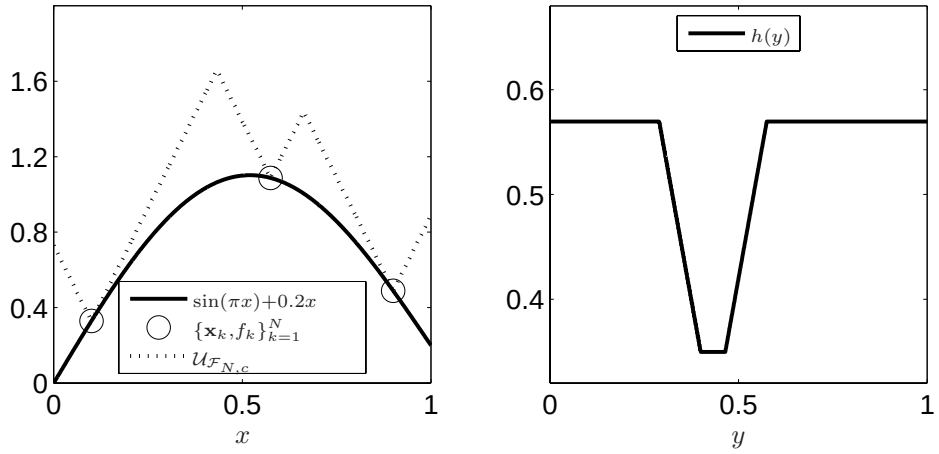


Figure 1: An illustration of the upper envelope function for the univariate function $f(x) = \sin(\pi x) + 0.2x$ sampled at three points, $E_N = \{0.1, 0.6, 0.9\}$, and with the value $c = 4$ of the Lipschitz constant. The right subfigure shows the function $h(y)$ as represented in Equation (21) to be minimized in the minimax problem (15).

DEFINITION 4.1 (Two-functions minimax problem) *Let $g : D \to \mathbb{R}$ and $\widetilde{g} : \mathbb{R}^n \to \mathbb{R}$ be two continuous functions. We denote*

$$\min_{\mathbf{y}\in D} h(\mathbf{y}), \tag{24}$$

where

$$h(\mathbf{y}) := \max_{\mathbf{x}\in D} \min\left\{g(\mathbf{x}), \widetilde{g}(\mathbf{y}-\mathbf{x})\right\}, \tag{25}$$

the two-functions minimax problem.

With

$$g(\mathbf{x}) = \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k \quad \text{and} \quad \widetilde{g}(\mathbf{x}) = c\|\mathbf{x}\|,$$

the objective function h characterized in (21) is recovered.

Recall that a function $\widetilde{g}$ is said to be *quasiconvex* on a convex set D if

$$\widetilde{g}((1-\lambda)\mathbf{x}' + \lambda\mathbf{x}'') \le \max\left\{\widetilde{g}(\mathbf{x}'), \widetilde{g}(\mathbf{x}'')\right\}, \tag{26}$$

for all $\mathbf{x}', \mathbf{x}'' \in D$ and $\lambda \in [0,1]$. A function is *strictly quasiconvex* if strict inequality in (26) holds whenever $\lambda \in (0,1)$, provided that $\widetilde{g}(\mathbf{x}') \neq \widetilde{g}(\mathbf{x}'')$. The following is an immediate result for quasiconvex functions (see, e.g., [2]).

THEOREM 4.2 *Every strict local minimum of a quasiconvex function is a global minimum, and the set of minimizing points for a quasiconvex function is convex.*

THEOREM 4.3 *Let h be defined by* (25) *with a convex domain D. Then h is quasiconvex if $\widetilde{g}$ is quasiconvex. As a consequence, the set of optimal points*

$$\arg\min_{\mathbf{y}\in D} h(\mathbf{y}) \tag{27}$$

is convex (if non-empty).

Proof Let $\mathbf{y}_1, \mathbf{y}_2 \in D$ and $\lambda \in [0,1]$. The quasiconvexity of $\widetilde{g}$ yields that

$$h\left((1-\lambda)\mathbf{y}_1 + \lambda\mathbf{y}_2\right)) \leq \max_{\mathbf{x}\in D} \min\left\{g(\mathbf{x}), \max\left\{\widetilde{g}(\mathbf{y}_1 - \mathbf{x}), \widetilde{g}(\mathbf{y}_2 - \mathbf{x})\right\}\right\}.$$

Since $\min\left\{a, \max\{b,c\}\right\} = \max\left\{\min\{a,b\}, \min\{a,c\}\right\}$, for all $a, b, c \in \mathbb{R}$, we have that

$$\begin{aligned} h((1-\lambda)\mathbf{y}_1 + \lambda\mathbf{y}_2) &\leq \max_{\mathbf{x}\in D} \max\left\{\min\{g(\mathbf{x}), \widetilde{g}(\mathbf{y}_1 - \mathbf{x})\}, \min\{g(\mathbf{x}), \widetilde{g}(\mathbf{y}_2 - \mathbf{x})\}\right\} \\ &= \max\left\{h(\mathbf{y}_1), h(\mathbf{y}_2)\right\}, \end{aligned}$$

where we have interchanged the order of the $\max$ operators in the last equality, and the first result follows. The convexity of the optimal set then follows from Theorem 4.2 ∎

In general, the function h is neither convex nor strictly quasiconvex (with no locally constant regions) even if $\widetilde{g}$ is. Therefore, the minimax problem does not necessarily have a unique optimal solution. The cure for this is to modify the reference set D. In a typical situation, we begin by calculating the maximum gain solution $\hat{\mathbf{x}}$ according to (14) and choose the reference set D such that $\hat{\mathbf{x}} \in D \subseteq \Omega$. A key feature of a two-functions minimax problem with a unique solution is balance in the sense that no matter which direction we move $\mathbf{y}$ from the optimal solution $\mathbf{y}^*$, an increased value of h emerges for an $\mathbf{x} \in D$ in the opposite direction. The next section contains optimality conditions for the minimax problem with respect to D, and also gives some suggestions on how to choose this domain.

4.1 Optimality conditions and domains of uniqueness for the two-functions minimax problem

We now restrict ourselves to the case when $\widetilde{g}(\mathbf{x}) = \phi(\|\mathbf{x}\|)$, for some continuous and strictly increasing function $\phi : \mathbb{R}_+ \to \mathbb{R}$ with $\phi(0) = 0$. For this special case it turns out that the two-functions minimax problem is closely related to following geometrical problem: Let $P \subset \mathbb{R}^n$ be a closed and bounded set. The *smallest enclosing ball* problem for P is to find a center $\mathbf{y} \in \mathbb{R}^n$ and a radius r by solving the optimization problem

$$\min\left\{r : \mathbf{y} \in \mathbb{R}^n,\, r \in \mathbb{R}_+,\, B(\mathbf{y}, r) \supseteq P\right\}, \tag{28}$$

where $B(\mathbf{y}, r)$ is the closed ball in $\mathbb{R}^n$ with center at $\mathbf{y} \in \mathbb{R}^n$ and radius $r \in \mathbb{R}_+$. This problem is treated by, e.g., Nielsen and Nock in [16] when P is a finite point set, but was considered already by Sylvester in [19]. The following fact, which is a consequence of the optimality of the smallest enclosing ball, will be useful in the further analysis.

LEMMA 4.4 *Let $B(\mathbf{y}^*, r^*)$ be the smallest enclosing ball for the set P. Denote the points in the set $\partial B(\mathbf{y}^*, r^*) \cap P$ control points. Then, the center $\mathbf{y}^*$ belongs to the convex hull of the control points. Equivalently, every closed half-space containing $\mathbf{y}^*$ also contains at least one control point.*

Let r be the function mapping a closed set $P \subset \mathbb{R}^n$ to the radius r^* of the smallest ball enclosing P according to (28);

$$r(P) = r^*.$$

A trivial property of the function r is that if $P \subseteq Q$, then $r(P) \leq r(Q)$. The following proposition connects the smallest enclosing ball problem and the two-functions minimax problem.

PROPOSITION 4.5 *The two-functions minimax problem* (24) *has a unique solution if and only if there exists a solution d^* (which is necessarily unique) to the equation*

$$\phi(r(\{\mathbf{x} \in D : g(\mathbf{x}) \geq d\})) = d. \tag{29}$$

Suppose that $d = d^*$ solves (29) *and let $B(\mathbf{y}^*, r^*)$ be the smallest ball enclosing $\{\mathbf{x} \in D : g(\mathbf{x}) \geq d^*\}$. Then $\mathbf{y}^*$ is also the solution to the minimax problem* (24) *and $h(\mathbf{y}^*) = \phi(r^*) = d^*$.*

Proof Suppose that $d = d^*$ solves (29). We first estimate $h(\mathbf{y}^*)$:

$$\begin{aligned}
h(\mathbf{y}^*) &= \max_{\mathbf{x}\in D} \min\left\{g(\mathbf{x}), \phi(\|\mathbf{y}^* - \mathbf{x}\|)\right\} \\
&= \max\left\{\max_{\mathbf{x}\in \overline{D\setminus B(\mathbf{y}^*,r^*)}} \min\left\{g(\mathbf{x}), \phi(\|\mathbf{y}^*-\mathbf{x}\|)\right\}, \max_{\mathbf{x}\in D\cap B(\mathbf{y}^*,r^*)} \min\left\{g(\mathbf{x}), \phi(\|\mathbf{y}^*-\mathbf{x}\|)\right\}\right\} \\
&\leq \max\left\{\max_{\mathbf{x}\in \overline{D\setminus B(\mathbf{y}^*,r^*)}} g(\mathbf{x}), \max_{\mathbf{x}\in D\cap B(\mathbf{y}^*,r^*)} \phi(\|\mathbf{y}^*-\mathbf{x}\|)\right\} \leq \max\{d^*, d^*\} = d^*.
\end{aligned}$$

This shows that $\mathbf{y}^*$ is a candidate for an optimal solution to the minimax problem (24) with an objective value less than or equal to d^*. Now, take any $\mathbf{y} \neq \mathbf{y}^*$. Since there exists an $\mathbf{x} \in D$ with $\|\mathbf{y} - \mathbf{x}\| > r^*$ such that $g(\mathbf{x}) > d^*$ and since ϕ is strictly increasing, we have that $\phi(\|\mathbf{y} - \mathbf{x}\|) > \phi(r^*) = d^*$. Thus, $h(\mathbf{y}) > d^*$ which proves that the unique minimum is attained at $\mathbf{y}^*$. By continuity, $h(\mathbf{y}^*) = d^*$.

For the converse implication, assume that $\mathbf{y}^*$ is the unique solution to the minimax problem with value $d^* = h(\mathbf{y}^*)$. Then there exists a number $r^* > 0$ such that $\phi(r^*) = d^*$. Since ϕ is strictly increasing it follows that $g(\mathbf{x}) \leq d^*$ for $\mathbf{x} \in D\setminus B(\mathbf{y}^*, r^*)$, otherwise $h(\mathbf{y}^*) > d^*$ which is a contradiction. Therefore $Q := \{\mathbf{x} \in D : g(\mathbf{x}) \geq d^*\} \subseteq B(\mathbf{y}^*, r^*)$. Now let $B(\widetilde{\mathbf{y}}, \widetilde{r})$ be the smallest enclosing ball for Q. The aim is to show that $B(\widetilde{\mathbf{y}}, \widetilde{r}) = B(\mathbf{y}^*, r^*)$. Assume on the contrary that the balls are not equal. Then, since $Q \subseteq B(\mathbf{y}^*, r^*)$ it must hold that $\widetilde{r} < r^*$. Now let r' be such that $\widetilde{r} < r' < r^*$. Then it follows that there exists a $\delta > 0$ such that $g(\mathbf{x}) \leq d^* - \delta$ for $\mathbf{x} \in D \setminus B(\widetilde{\mathbf{y}}, r')$ and $\phi(\|\widetilde{\mathbf{y}} - \mathbf{x}\|) \leq d^* - \delta$ for $\mathbf{x} \in B(\widetilde{\mathbf{y}}, r')$. Therefore $h(\widetilde{\mathbf{y}}) \leq d^* - \delta < h(\mathbf{y}^*)$. This contradicts the optimality of $\mathbf{y}^*$ for the minimax problem which yields that $B(\mathbf{y}^*, r^*)$ is the smallest enclosing ball for Q. Now since $\phi(r^*) = d^*$ it follows that d^* solves (29). ∎

Since

$$\{\mathbf{x} \in D : g(\mathbf{x}) \geq d\} \subseteq \{\mathbf{x} \in D : g(\mathbf{x}) \geq d'\}$$

whenever $d \geq d'$ it follows that

$$d \mapsto r(\{\mathbf{x} \in D : g(\mathbf{x}) \geq d\}) \tag{30}$$

is a decreasing (not necessarily strictly) function of d. By imposing a continuity assumption, which is related to a suitable D, we can prove an existence result for the solution of the equation (29).

COROLLARY 4.6 *Suppose that g and $D \subseteq \Omega$ are such that the function* (30) *is continuous for $d \in (a, b) \subset [-\infty, \max_{\mathbf{x}\in D} g(\mathbf{x})]$ for some $a,\ b \in \mathbb{R}$ such that*

$$\phi\left(r(\{\mathbf{x} \in D : g(\mathbf{x}) \geq a\})\right) > a,$$
$$\phi\left(r(\{\mathbf{x} \in D : g(\mathbf{x}) \geq b\})\right) < b.$$

Then there exists a unique solution $d^* \in (a, b)$ to the equation (29) *and therefore to the minimax problem* (25).

Proof From the assumptions on ϕ and from (30), it follows that the function

$$q(d) := \phi(r(\{\mathbf{x} \in D : g(\mathbf{x}) \geq d\})) - d$$

is continuous and strictly decreasing on (a, b), and that $q(a) < 0$ and $q(b) > 0$. Therefore there exists a unique $d^* \in (a, b)$ such that $q(d^*) = 0$ which is equivalent to (29). ∎

The aim is to construct a D such that the assumptions of Corollary 4.6 are fulfilled.

The following result is clear from the definition of the minimax problem and shows that for certain (small) domains, the minimax problem has a particularly simple solution.

COROLLARY 4.7 *Suppose that D and g are such that*

$$\phi\left(r(D)\right) \leq \min_{\mathbf{x}\in D} g(\mathbf{x}). \tag{31}$$

Then the minimax solution $\mathbf{y}^*$ is the center of the smallest enclosing ball of D.

Proposition 4.5 and Corollary 4.6 are illustrated in Figure 2 with the example from Figure 1, for which the minimax solution is non-unique. This non-uniqueness is a consequence of the rightmost local maximum which induces a jump in the radius function (30) and therefore in the left hand side of (29). In the two lower subfigures the reference set D is decreased so that it only includes one of the local maxima, making the radius function continuous and therefore yielding a unique minimax solution. The conclusion to be made from Proposition 4.5 is that for a suitable choice of D, the optimal solution to the minimax problem (24) is unique. A strategy for

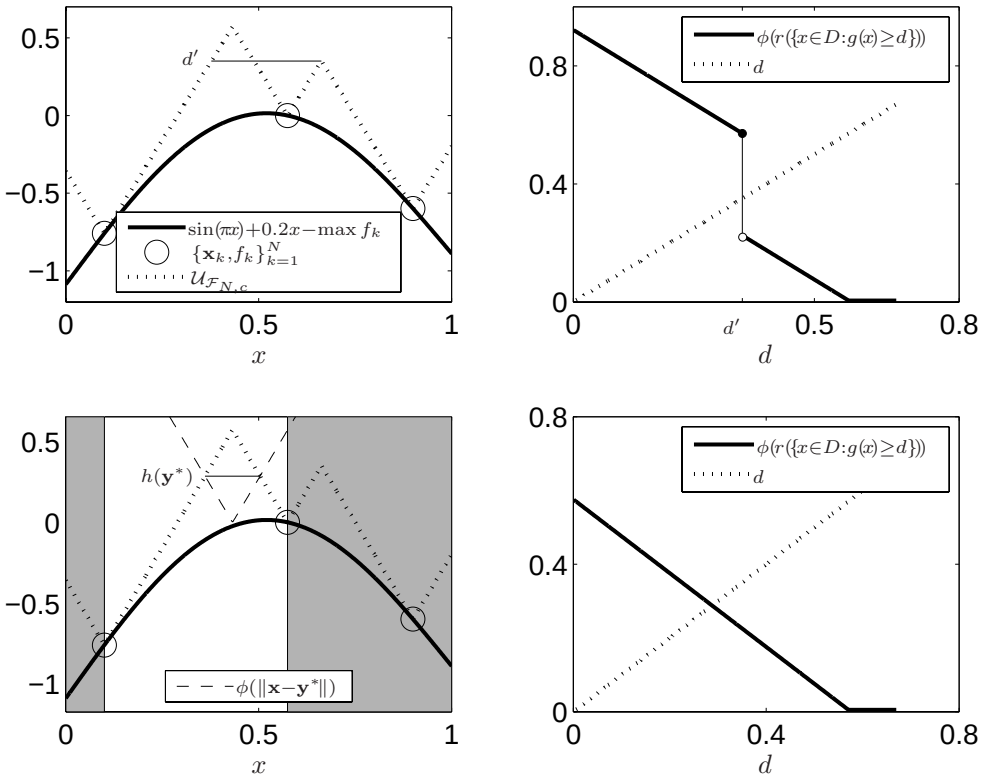


Figure 2: By decreasing the domain D in the lower set of subfigures (excluding the gray region) so that the upper envelope function only contains one local maximum over the domain, a unique solution to (29) and hence to the minimax problem is obtained.

decreasing the reference set D to obtain uniqueness in the minimax problem is presented in Section 4.2.

Proposition 4.8 involves a set of auxiliary control points which is shown to be *balanced* when corresponding to the optimal solution $\mathbf{y}^*$ to (24). Our proposed algorithm for solving the problem (24) seeks simultaneously for these points and for the optimal solution $\mathbf{y}^*$.

PROPOSITION 4.8 (optimality conditions with control points) *Let g and D be specified as in Corollary 4.6. A point $\mathbf{y}^* \in D$ is the unique solution to the two-functions minimax problem if and only if there is a set $\{\mathbf{z}_k\}_{k=1}^M \subset D$ of control points and a scalar $d^* \in \mathbb{R}$ such that*

1. $\mathbf{y}^* \in \operatorname{conv}\left(\{\mathbf{z}_k\}_{k=1}^M\right)$,

2. $\min\{g(\mathbf{z}_k), \phi(\|\mathbf{y}^* - \mathbf{z}_k\|)\} = d^*,\ k = 1, \ldots, M$ *(if* $\mathbf{z}_k \in \operatorname{int}(D)$*, then* $g(\mathbf{z}_k) = \phi(\|\mathbf{y}^* - \mathbf{z}_k\|) = d^*$*), and*

3. $h(\mathbf{y}^*) = \max_{\mathbf{x}\in D} \min\{g(\mathbf{x}), \phi(\|\mathbf{y}^* - \mathbf{x}\|)\} = d^*$.

Proof Assume that $\mathbf{y}^* \in D$ is the unique solution to the two-functions minimax problem (Definition 4.1) with value d^*. Let $B(\mathbf{y}^*, r^*)$ be the smallest ball enclosing the set $\{\mathbf{x} \in D : g(\mathbf{x}) \geq d^*\}$. According to Proposition 4.5, the equality $\phi(r^*) = d^*$ holds. By Lemma 4.4, there exists a set of control points $\{\mathbf{z}_k\}_{k=1}^M \subset D$ such that $\mathbf{y}^* \in \operatorname{conv}\left(\{\mathbf{z}_k\}_{k=1}^M\right)$. Since $\|\mathbf{y}^* - \mathbf{z}_k\| = r^*$, it follows that $\phi(\|\mathbf{y}^* - \mathbf{z}_k\|) = d^*$ and also that $g(\mathbf{z}_k) \geq d^*$. Item *3* is a consequence of the optimality of the minimax solution.

For the reverse implication, the idea is to prove that under the assumption that *1–3* hold, the value $h(\mathbf{y}^*)$ will be smaller than the function values for all points in a ball around $\mathbf{y}^*$. Then $\mathbf{y}^*$ is a strict local minimum, and from Theorem 4.2 the result will be proved.

Pick an arbitrary point $\mathbf{y}'$ in a ball $B(\mathbf{y}^*, \varepsilon)$ specified by $\mathbf{y}' = \mathbf{y}^* + \varepsilon'\mathbf{v}$ where $\mathbf{v} \in \mathbb{R}^n$ and $0 < \varepsilon' < \varepsilon$. Let $\mathbf{z}$ be one of the $\mathbf{z}_k$, $k = 1, \ldots, M$, that, from *1*, are contained in the closed half-space $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{v}^{\mathrm{T}}(\mathbf{x} - \mathbf{y}^*) \leq 0\}$.

Clearly, $\phi(\|\mathbf{y}^* - \mathbf{z}\|) \geq d^*$ from *2* and $\phi(\|\mathbf{y}' - \mathbf{z}\|) > \phi(\|\mathbf{y}^* - \mathbf{z}\|)$ since ϕ is strictly increasing with the norm of its argument. From continuity, there now exists a ball $B(\mathbf{z}, \delta)$, $\delta > 0$, such that for all $\mathbf{x} \in B(\mathbf{z}, \delta)$ it holds that

$$\phi(\|\mathbf{y}' - \mathbf{x}\|) > \phi(\|\mathbf{y}^* - \mathbf{z}\|) \geq d^*.$$

Furthermore, from *2* we have that $g(\mathbf{z}) \geq d^*$ holds. If $g(\mathbf{z}) > d^*$, then $g(\mathbf{x}) > d^*$ holds for all $\mathbf{x} \in B(\mathbf{z}, \delta)$ if δ is chosen sufficiently small. Otherwise, if $g(\mathbf{z}) = d^*$, then it can be concluded from the assumptions on D and g that $g(\mathbf{z})$ cannot be a local maximum over D (since $g(\mathbf{z}) \in (a, b)$ and a local maximum in this interval would lead to a non-continuous radius function, see Corollary 4.6). Therefore, we get

$$\max_{\mathbf{x}\in B(\mathbf{z},\delta)\cap D} g(\mathbf{x}) > g(\mathbf{z}) \geq d^*.$$

Hence, $h(\mathbf{y}') > h(\mathbf{y}^*)$, i.e., the arbitrarily chosen point $\mathbf{y}' \in B(\mathbf{y}^*, \varepsilon)$ results in a strictly larger function value and the result follows. ∎

In general, we want to have a domain D which results in a continuous radius function (30), but we also want the domain to be as large as possible in order to maximize the information provided from the function. However, the searched sample point $\mathbf{y}$ will affect the upper envelope function only locally, and therefore there is an upper bound on the size of a useful set D. Such a bound can be used when designing a practical method for constructing a domain. The following result gives such an upper bound in the Lipschitz case in the sense that a larger set D will not affect the value of $\mathbf{y}^*$ when solving the minimax problem.

PROPOSITION 4.9 *Assume that* $\phi(r) = cr$*. Let* $\hat{\mathbf{x}} \in \arg\max_{\mathbf{x}\in\Omega} \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x})$*. Suppose that* $\mathbf{y} \in \Omega$ *is such that* $h(\mathbf{y}) < \mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_{k=1,\ldots,N} f_k$*. Then* $\mathbf{y} \in B(\hat{\mathbf{x}}, R)$*, where* $R = \left(\mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}}) - \max_k f_k\right)/c$*. Furthermore, the corresponding control points* $\mathbf{z}_k$*,* $k = 1, \ldots, M$*, fulfill* $\mathbf{z}_k \in B(\hat{\mathbf{x}}, 2R)$.

Proof In order for the vector $\mathbf{y}$ to affect the maximum value of the upper envelope, and hence the maximum loss that is to be minimized, it has to be sufficiently close to $\hat{\mathbf{x}}$. From the expression (20), it follows that $\mathbf{y} \in B(\hat{\mathbf{x}}, R)$. In Proposition 4.8 containing necessary and sufficient optimality conditions for the minimax problem, it is shown that $\mathbf{y}$ affects the upper envelope at a set of control points. The value of the upper envelope function at each of the control points is obviously smaller than $\mathcal{U}_{\mathcal{F}_{N,c}}(\hat{\mathbf{x}})$. Therefore, as above, their distance to $\mathbf{y}$ must be smaller than R, and hence $\mathbf{z}_k \in B(\hat{\mathbf{x}}, 2R)$. This concludes the arguments for the proposition. ∎

4.2 Constructions of suitable domains

There are a number of criteria which seem natural to pose on a suitable domain D for the minimax problem:

1. D should be a domain of uniqueness for the minimax problem.
2. D should have a simple, e.g., convex, shape (simplifying the computation of the minimax solution and the corresponding control points).
3. The computational effort for computing D should be reasonable.

Finding an algorithm which generates domains fulfilling all these criteria has turned out to be the hardest subproblem for the problem considered in this article. Up to now, all algorithms found, meet some, but not all, of the above criteria. An exception is the one with $D = \{\hat{\mathbf{x}}\}$, however, that is clearly not what we want. Below we define constructions which are proved to be domains of uniqueness for the two-functions minimax problem. Unfortunately, these sets are in general not convex and no fail-safe algorithm has been found to generate them. An attempt to define a cheaply computable approximation is presented in the end of this section.

DEFINITION 4.10 (The descent set) *The* descent set $D_g(S)$ *for a starting set* $S \subset \Omega$ *with respect to a function* $g \in C(\Omega)$ *is the union of all points* $\mathbf{y} \in \Omega$ *for which there exists a continuous function* $\gamma : [0, T] \to \Omega$ *such that* $\gamma(0) \in S$, $\mathbf{y} = \gamma(T)$, *and* $g \circ \gamma$ *is a decreasing function on* $[0, T]$.

From this definition follows that $D_g(S) \setminus S$ cannot contain any local maximum for g. For this reason the descent set for $S = \{\hat{\mathbf{x}}\} = \arg\max_{\mathbf{x}\in\Omega} g(\mathbf{x})$ is a candidate for the reference set for the minimax problem (24). This is actually how the reference set for the lower two subfigures of Figure 2 was chosen. However, the requirements for the descent set is slightly too weak, implying that such a domain can result in jumps in the radius function. As a remedy, we propose the *Lipschitz descent* set, for which we require strict and controllable decay of the function $g \circ \gamma$.

DEFINITION 4.11 (The Lipschitz descent set) *The* Lipschitz descent set $D_g(S, \alpha)$, $\alpha > 0$, *for a starting set* $S \subset \Omega$ *with respect to a function* $g \in C(\Omega)$ *is the union of all points* $\mathbf{y} \in \Omega$ *for which there exists a continuously differentiable function* $\gamma : [0, T] \to \Omega$ *with* $\|\dot{\gamma}(t)\| = 1$ *for all* $t \in [0, T]$ *such that* $\gamma(0) \in S$, $\mathbf{y} = \gamma(T)$, *and* $g \circ \gamma$ *is a uniformly decreasing function* $g(\gamma(s)) - g(\gamma(t)) \geq \alpha(t - s)$ *whenever* $T \geq t \geq s \geq 0$.

The next result is a consequence of the above definition.

PROPOSITION 4.12 *Let $D_g(S,\alpha)$ be a Lipschitz descent set for a starting set S with respect to g. Then the radius function* (30) *for $D = D_g(S,\alpha)$ is Lipschitz continuous for $d \leq \min_{\mathbf{x}\in S} g(\mathbf{x})$ with norm at most $1/\alpha$.*

Proof Let $q > 0$. We will prove that the distance from an arbitrarily chosen point in $\{\mathbf{x} \in D_g(S,\alpha) : g(\mathbf{x}) \geq d - q\}$ to a point in $\{\mathbf{x} \in D_g(S,\alpha) : g(\mathbf{x}) \geq d\}$ is at most q/α. This implies that if $B(\mathbf{x}_d, r_d)$ is the smallest ball enclosing $\{\mathbf{x} \in D_g(S,\alpha) : g(\mathbf{x}) \geq d\}$ then $B(\mathbf{x}_d, r_d + q/\alpha)$ is enclosing $\{\mathbf{x} \in D_g(S,\alpha) : g(\mathbf{x}) \geq d - q\}$. This proves that the radius function is Lipschitz continuous with Lipschitz constant $1/\alpha$.

Now, let $\mathbf{y} \in \{\mathbf{x} \in D_g(S,\alpha) : g(\mathbf{x}) \geq d - q\}$. If $g(\mathbf{y}) \geq d$ there is nothing to prove so assume that $g(\mathbf{y}) < d$. By assumption there exists a continuously differentiable function $\gamma : [0,T] \to \Omega$ with $\|\dot\gamma(t)\| = 1$ for all $t \in [0,T]$ such that $\mathbf{y} = \gamma(T)$, $\gamma(0) \in S$ and $g(\gamma(s)) - g(\gamma(t)) \geq \alpha(t-s)$ whenever $t \geq s \geq 0$. By continuity there exists an $s' < t$ such that $\mathbf{x} = \gamma(s')$ with $g(\mathbf{x}) = d$ (since $g(\gamma(0)) \geq d > g(\gamma(t)) = g(\mathbf{y})$). Now, $\|\mathbf{y} - \mathbf{x}\| = \|\gamma(t) - \gamma(s')\| \leq t - s'$ and hence

$$\|\mathbf{y} - \mathbf{x}\| \leq \frac{\alpha(t-s')}{\alpha} \leq \frac{g(\gamma(s')) - g(\gamma(t))}{\alpha} \leq \frac{d - (d-q)}{\alpha} = \frac{q}{\alpha},$$

which yields the result. ∎

To make this notion useful for implementation into a computer code we need discrete analogues of the descent set and the Lipschitz descent set. A *simplex* is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^n$. Let K be a covering of the domain Ω of non-overlapping simplices. In two dimensions such coverings are called triangulations. This induces a graph connecting the different vertices in K: two vertices are connected with an edge if there exists a simplex to which they both belong.

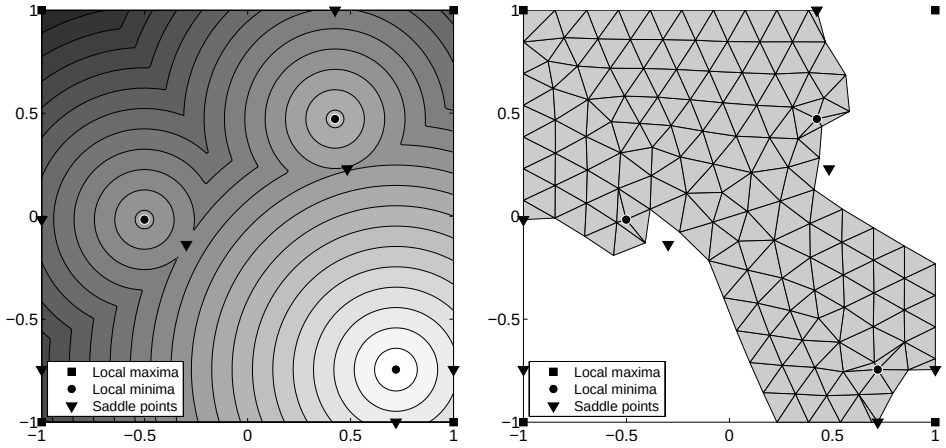
DEFINITION 4.13 (The discrete descent set) *Let the starting set S be a subset of all vertices in K. A vertex $\nu \in K$ belongs to the discrete descent set of vertices if there exists a chain of vertices $\{\nu_k\}_{k=0}^{N_\nu}$ in K such that (i) $\nu_0 \in S$, (ii) $\nu = \nu_{N_\nu}$, (iii) for $k = 1, \ldots, N_\nu$ there exists a simplex in K which contains both ν_{k-1} and ν_k, and (iv)*

$$g(\nu_{k-1}) - g(\nu_k) \geq 0. \tag{32}$$

The discrete descent set $D_g^d(S)$ of simplices is the union of all simplices whose all vertices belong to the discrete descent set of vertices. For the discrete Lipschitz descent set $D_g^d(S,\alpha)$, (32) *should be replaced by*

$$g(\nu_{k-1}) - g(\nu_k) \geq \alpha\|\nu_{k-1} - \nu_k\|. \tag{33}$$

Figure 3 shows an example of a discrete descent set $D_g^d(S)$ for a simple example. Each extreme point of the domain $\Omega = [-1,1]^2$ is a local maximum to the upper envelope function $\mathcal{U}_{\mathcal{F}_{N,c}}$ and the extreme point $\hat{\mathbf{x}} = (-1,1)^{\mathrm{T}}$ is the global maximum. In this example the starting set $S = \{\hat{\mathbf{x}}\}$. In the figure the saddle points of $\mathcal{U}_{\mathcal{F}_{N,c}}$ have been marked. At these points $D_g^d(S)$ is partly blocked from continuing to grow which



(a) Level curves for the upper envelope function. (b) The discrete descent set for a given triangulation K.

Figure 3: Illustration of a discrete descent set.

prevents it to reach the other local maxima. As can be seen in the figure, $D_g^d(S)$ depends critically on the triangulation of the domain, which also makes it smaller than the true descent set. This dependence seems unavoidable unless the triangulation is adapted during the computation of the descent set itself.

5 Solution methods

Up until now we have explored the theoretical properties of the minimax problem with an emphasize on the case when the function class consists of Lipschitz continuous functions. For this case we have been able to derive explicit expressions and optimality conditions that can be exploited in order to solve the minimax problems.

In this section we provide solution methods for the minimax problem in the Lipschitz case with the norm bounded by the constant c. The optimization problem under consideration is thus to solve (15), i.e., (utilizing Proposition 3.1 and the relation (21)) to find $\mathbf{y}^*$ that solves

$$\min_{\mathbf{y}\in D}\max_{\mathbf{x}\in D}\min\left\{\min_{k=1,\dots,N}\{f_k + c\|\mathbf{x}-\mathbf{x}_k\|\} - \max_{k=1,\dots,N} f_k,\, c\|\mathbf{y}-\mathbf{x}\|\right\}. \tag{34}$$

In order to construct a suitable domain $D \subseteq \Omega$, first the maximum gain solution $\hat{\mathbf{x}}$, i.e., the maximizer to the upper envelope function (20) has to be found. A procedure for this is presented in Section 5.2. We begin, however, with a discussion about the possible choices of the Lipschitz constant c, and the convergence properties when using the minimax method repeatedly for determining the sample points $\mathbf{x}_k$ when solving the original maximization problem (1).

5.1 The Lipschitz constant and convergence

As discussed in Section 3, the Lipschitz constant of a function can be estimated from below using a set of already sampled points and the inequality (18). For any value of c larger than this underestimation, the minimax optimization problem (34) will have a solution. If it is also possible to estimate the Lipschitz constant from above (e.g. if the function is an explicit composition of elementary functions, or if there are physical limitations on the function derivatives), then a repeated use of the minimax strategy guarantees convergence towards a global optimum of (1), this without ending up with the sample set being dense in the whole domain. The convergence is clear since if a global optimum $\mathbf{x}^*$ is not already among the sampled points $\mathbf{x}_k$, $k = 1, \ldots, N$, then it holds that the maximum loss $\mathcal{ML}(\mathcal{F}_{N,c}, D, E_N) = \max_{\mathbf{x}\in D} \mathcal{U}_{\mathcal{F}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k > 0$, since $\mathcal{U}_{\mathcal{F}}(\mathbf{x}^*) \geq f(\mathbf{x}^*) > f_k$, $k = 1, \ldots, N$. Furthermore, unless the maximum loss is equal to zero, it decreases strictly after a finite number of iterations of the main algorithm described in Section 2.2 (when, at a certain iteration, all, possibly non-unique but finitely many, global maxima of $\mathcal{U}_{\mathcal{F}_{N,c}}$, have been used to define a domain D). This concludes the convergence argument.

Which value of c larger than the underestimation given by (18) that is best to use cannot be said with certainty. A too large value of c will lead to a too large emphasis on filling out the space, yielding unnecessary samples in regions where the function values are low. A too small value of c might, however, lead to that subregions are erroneously discarded, and that the global optimum is missed.

It is possible that using local estimates of the Lipschitz constant—rather than the global estimate that we consider in this article—would increase the efficiency of our method. One idea is to let the Lipschitz constant in the definition (20) of the upper envelope be dependent on k, i.e., on the evaluation points. Then $c = c_k$ could be estimated locally by using only close surrounding sample points in the lower estimate according to (18).

5.2 A spatial Branch-and-Bound scheme for finding the maximizer of the upper envelope function

Given a box-constrained region $\Omega = [\boldsymbol{\ell}, \mathbf{u}] \subset \mathbb{R}^n$, an unknown function $f : \Omega \to \mathbb{R}$, and a set of sampled points $\{(\mathbf{x}_1, f_1), \ldots, (\mathbf{x}_N, f_N)\}$ this section presents a model of, and a spatial Branch-and-Bound scheme for, finding $\hat{\mathbf{x}}$, i.e., the maximizer to the upper envelope function $\mathcal{U}_{\mathcal{F}_{N,c}}$ as defined by (20).

From Proposition 3.1 we have that $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) = \min_{k=1,\ldots,N} \{f_k + c\|\mathbf{x} - \mathbf{x}_k\|\}$ for the Lipschitz space of functions. Therefore, we introduce an auxiliary variable θ and find $\hat{\mathbf{x}}$ by solving the problem to

$$\begin{aligned} \underset{\mathbf{x},\theta}{\text{minimize}} \quad & -\theta, \\ \text{subject to} \quad & \theta \leq f_k + c\|\mathbf{x} - \mathbf{x}_k\|, \quad k = 1, \ldots, N, \\ & \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}, \\ & \theta \in \mathbb{R}. \end{aligned} \tag{35}$$

This is a nonlinear programming problem for which Algorithm 1 below presents a local optimization algorithm based on sequential linearization; it will be used within the Branch-and-Bound framework.

Algorithm 1 Find a local maximizer $\mathbf{x}^*$ of the Lipschitz upper envelope function

Require: A set of sampled points with corresponding function values $\{(\mathbf{x}_k, f_k)\}_{k=1}^N$, a consistent Lipschitz constant c (fulfilling the inequality (19)), and a domain $\Omega = [\boldsymbol{\ell}, \mathbf{u}]$. An initial guess $\mathbf{x}^0$ of $\hat{\mathbf{x}}$ and a solution tolerance $\varepsilon_{\text{gap}} > 0$.

Ensure: A local maximizer $\mathbf{x}^*$ of $\mathcal{U}_{\mathcal{F}_{N,c}}$ within the tolorance ε_{gap} in terms of objective value.

Start

Define $\mathbf{p} = (\mathbf{x}^{\mathrm{T}}, \theta)^{\mathrm{T}}$.

$\theta^0 \leftarrow f(\mathbf{x}^0)$; $i \leftarrow 0$.

repeat

$i \leftarrow i + 1$.

Linearize the constraints $\theta \leq f_k + c\|\mathbf{x} - \mathbf{x}_k\|,\ k = 1, \ldots, N$, around $\mathbf{p} = \mathbf{p}^{i-1}$ $\Rightarrow\ \mathbf{A}^i\mathbf{p} \leq \mathbf{b}^i$.

Collect the lower and upper bounds on the decision variables in the vectors $\boldsymbol{\ell}_{\mathbf{p}}$ and $\mathbf{u}_{\mathbf{p}}$.

Solve the linear program

$$\begin{aligned} \mathbf{p}^i \in \arg\min_{\mathbf{p}} \quad & -\theta, \\ \text{s.t.} \quad & \mathbf{A}^i\mathbf{p} \leq \mathbf{b}^i, \\ & \boldsymbol{\ell}_{\mathbf{p}} \leq \mathbf{p} \leq \mathbf{u}_{\mathbf{p}}, \end{aligned} \tag{36}$$

until $\|\mathbf{p}^i - \mathbf{p}^{i-1}\| < \varepsilon_{\text{gap}}$.

Return: $\mathbf{x}^* = \mathbf{x}^i$.

It can be shown (see e.g. [2], Chapter 4) that when (and only when) the optimal solution to (36) corresponds to a zero step in $\mathbf{p}$ (i.e., $\mathbf{p}^i = \mathbf{p}^{i-1}$), then a Karush-Kuhn-Tucker (KKT) point to the nonlinear program (35) has been found, thus motivating the termination criterion of Algorithm 1. In practice, such a KKT point normally corresponds to a local maximum of $\mathcal{U}_{\mathcal{F}_{N,c}}$ in the concave region in which the initial guess $\mathbf{x}^0$ is located. That the nonlinear constraints in (35) are not differentiable everywhere (they are actually non-differentiable at each point in the evaluation set $\cup_{k=1}^N \mathbf{x}_k$) is not a severe practical issue, since they will be differentiable around each local maximum of $\mathcal{U}_{\mathcal{F}_{N,c}}$. Rapid convergence of the suggested method towards a maximum of $\mathcal{U}_{\mathcal{F}_{N,c}}$ has been obtained in our numerical experiments. This is also expected from the discussion in Chapter 10 in [2], since such a method behaves essentially like Newton's algorithm when entering a close neighborhood of the final solution.

Each solution found by Algorithm 1 is a lower bound on the global maximum of $\mathcal{U}_{\mathcal{F}_{N,c}}$. Furthermore, upper bounds are easily obtained by employing the Lipschitz property of $\mathcal{U}_{\mathcal{F}_{N,c}}$. With $\mathbf{x}$ restricted to some set $\bar{\Omega} \subseteq \Omega$ and with $B\left(\mathbf{x}_{\bar{\Omega}}, r(\bar{\Omega})\right)$ denoting the smallest ball enclosing $\bar{\Omega}$, we have that $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}^*) \leq \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}_{\bar{\Omega}}) + cr(\bar{\Omega})$. The bounding procedures are utilized in a breadth-first spatial Branch-and-Bound

scheme for the global maximizer $\hat{\mathbf{x}}$ of $\mathcal{U}_{\mathcal{F}_{N,c}}$. We begin with a family T of unprocessed boxes, starting with $T = \{\Omega\}$. At each iteration—or level in the Branch-and-Bound tree—the upper envelope function $\mathcal{U}_{\mathcal{F}_{N,c}}$ is evaluated at the mid point of each box in the family. The upper bound of $\mathcal{U}_{\mathcal{F}_{N,c}}$ restricted to each box is computed as above, and if it is concluded that the global maximum can be contained in the box, then it is split into 2^n new boxes that are included in T, and a new lower bound to the global maximum is computed by running Algorithm 1 restricted to the box using its mid point as a starting guess $\mathbf{x}^0$. The procedure is terminated when a local maximum that is guaranteed to be within $\varepsilon_{\texttt{gap}} > 0$ of the global maximum has been found. In practice, the algorithm can also be terminated with a suboptimal solution after a limited number of local optimizations or at a limited depth in the search tree.

5.3 The algorithm for finding the minimax solution

This section presents an algorithm for finding the minimizer $\mathbf{y}^*$ to the two-functions minimax problem (34) in the Lipschitz case, given the maximizer $\hat{\mathbf{x}}$ of the upper envelope function $\mathcal{U}_{\mathcal{F}_{N,c}}$ and a domain D of uniqueness, i.e., $D \in \mathcal{D} := \{D \subseteq \Omega \mid D$ such that the assumptions of Proposition 4.5 are fulfilled$\}$, is known, and, furthermore, that it has a polyhedral representation. The procedure is based on the characterizations made in Proposition 4.8 and the solution idea is similar to that of Algorithm 1.

From Proposition 4.8 it can be concluded that by utilizing an auxiliary variable $\theta \in \mathbb{R}$, the two-functions minimax problem (34) can be rewritten as

$$
\begin{aligned}
\underset{\mathbf{y},\,\theta,\,\mathbf{z}_j,\,\lambda_j,\,j=1,\ldots,n+1}{\text{minimize}} \quad & -\theta, \\
\text{subject to} \quad & \theta \le f_k + c\|\mathbf{z}_j - \mathbf{x}_k\|, && k = 1,\ldots,N, && \text{(37a)}\\
& && j = 1,\ldots,n+1, \\
& \theta \le \max_{k=1,\ldots,N} f_k + c\|\mathbf{y} - \mathbf{z}_j\|, && j = 1,\ldots,n+1, && \text{(37b)}\\
& \mathbf{z}_j \in D, && j = 1,\ldots,n+1, \\
& \mathbf{y} \in D, \\
& \theta \in \mathbb{R}, \\
& \mathbf{y} = \sum_{j=1}^{n+1} \lambda_j \mathbf{z}_j, && && \text{(37c)}\\
& \sum_{j=1}^{n+1} \lambda_j = 1, && && \text{(37d)}\\
& \lambda_j \in [0,1], && j = 1,\ldots,n+1. && \text{(37e)}
\end{aligned}
$$

The construction is similar to that of (35), but in (37) the maximization is done at the control points $\mathbf{z}_1,\ldots,\mathbf{z}_{n+1}$ instead of at $\mathbf{x}$. The term $\max_{k=1,\ldots,N} f_k$ in (37b) represents, due to Lemma 3.3, the worst-case objective value at the searched sample point

$\mathbf{y}$. By the construction of (37a)–(37b) and from the minimization of θ, the value of the inner function of (34), given by

$$\min\left\{\min_{k=1,\ldots,N}\{f_k + c\|\mathbf{x}-\mathbf{x}_k\|\} - \max_{k=1,\ldots,N} f_k,\ c\|\mathbf{y}-\mathbf{x}\|\right\},$$

will, for $\mathbf{y}=\mathbf{y}^*$, take its largest value at $\mathbf{x}=\mathbf{z}_j,\ j=1,\ldots,n+1$.

The constraints (37c)–(37e) are actually redundant and will be automatically fulfilled in an optimal solution. However, in the iterative algorithm presented below, we have noted that their utilization reduce the risk of ending up in poor local minima.

Algorithm 2 Find a global optimizer $\mathbf{y}^*$ to the minimax problem (34)

Require: A set of sampled points with corresponding function values $\{(\mathbf{x}_k, f_k)\}_{k=1}^N$, a consistent Lipschitz constant c (fulfilling the inequality (19)), a polyhedral domain $D=\{\mathbf{x}\in\Omega : \mathbf{A}_D\mathbf{x}\le\mathbf{b}_D\}$, the maximizer $\hat{\mathbf{x}}$ of the Lipschitz upper envelope $\mathcal{U}_{\mathcal{F}_{N,c}}$ and a solution tolerance $\varepsilon_{\text{gap}}>0$.

Ensure: A global minimizer $\mathbf{y}^*$ to the minimax problem (34) within the tolerance ε_{gap} in terms of objective value.

Start

$\mathbf{y}^0 \leftarrow \hat{\mathbf{x}}$; $\mathbf{z}_j^0 \leftarrow \mathbf{y}^0 + \varepsilon\mathbf{v}_j,\ j=1,\ldots,n+1$, where $V=\{\mathbf{v}_1,\ldots,\mathbf{v}_{n+1}\}$ is a positive basis with uniform angles (see [5]) and $\varepsilon>0$ is a small constant.

Let $\boldsymbol{\lambda}^0$ be such that (37c)–(37e) hold for $\boldsymbol{\lambda}=\boldsymbol{\lambda}^0$.

Define $\mathbf{p} := (\mathbf{y}^{\mathrm{T}}, \mathbf{z}_1{}^{\mathrm{T}}, \ldots, \mathbf{z}_{n+1}{}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}}, \theta)^{\mathrm{T}}$.

$i \leftarrow 0$.

repeat

$i \leftarrow i+1$.

Linearize the inequality constraints (37a) and (37b) around $\mathbf{p}=\mathbf{p}^{i-1}$ and compose these with the domain constraints for $\mathbf{y}^i$ and $\mathbf{z}_j^i,\ j=1,\ldots,n+1$ $\Rightarrow\ \mathbf{A}^i\mathbf{p}\le\mathbf{b}^i$.

Linearize the equality constraints (37c) around $\mathbf{p}=\mathbf{p}^{i-1}$ and compose these with (37d) $\Rightarrow\ \mathbf{A}_{eq}^i\mathbf{p}=\mathbf{b}_{eq}^i$.

Collect the lower and upper bounds on the decision variables in the vectors $\boldsymbol{\ell}_{\mathbf{p}}$ and $\mathbf{u}_{\mathbf{p}}$.

Solve the linear program

$$\begin{aligned}\mathbf{p}^i \in \arg\min_{\mathbf{p}} \quad & -\theta,\\ \text{s.t.} \quad & \mathbf{A}^i\mathbf{p}\le\mathbf{b}^i,\\ & \mathbf{A}_{eq}^i\mathbf{p}=\mathbf{b}_{eq}^i,\\ & \boldsymbol{\ell}_{\mathbf{p}}\le\mathbf{p}\le\mathbf{u}_{\mathbf{p}},\end{aligned}$$

until $\|\mathbf{p}^i-\mathbf{p}^{i-1}\|<\varepsilon_{\text{gap}}$.

Return: $\mathbf{y}^*=\mathbf{y}^i$.

As for Algorithm 1, the solution that is returned from Algorithm 2 after a zero step in $\mathbf{p}$ corresponds to a KKT point of the nonlinear program (37). In Section 4.1 it

is shown that there are domains in $\mathcal{D}$ such that the minimax problem has a unique optimum. However, even with $D \in \mathcal{D}$, it is possible that the output of Algorithm 2 is a local, but not global, minimax solution. This is due to the optimization being performed over a larger space containing also the variables $\mathbf{z}_j$ and not only the variable $\mathbf{y}$ considered in Proposition 4.5. Heuristic techniques can be employed to improve on the possible termination at a local, but not global, optimum. For example, the Branch-and-Bound algorithm presented in Section 5.2 can be applied to the Lipschitz upper envelope $\mathcal{U}_{\mathcal{F}_{N,c}}$ given the evaluation set including the found minimax solution $\mathbf{y}^*$. If a value that is larger than the function values at the found control points $\mathbf{z}_1, \ldots, \mathbf{z}_{n+1}$ is the result from the Branch-and-Bound scheme, then $\mathbf{y}^*$ is not the global optimal solution. One can now create a scheme where $\mathbf{y}^*$ is moved towards the maximizer of the updated Lipschitz envelope, possibly in combination with a repeated application of Algorithm 2, until there is a set of control points $\mathbf{z}_1, \ldots, \mathbf{z}_M$ fulfilling Proposition 4.8, thus guaranteeing that the global optimal solution to (34) has been found.

6 Numerical experiments

In this section, the algorithms proposed in Section 5 are illustrated and the minimax and the maximum gain strategies are compared. The aim of the minimax strategy is to minimize the (worst-case) optimality gap, i.e., the maximum loss, after the sampling of a new point in order to find a good point to evaluate in the next iteration for the outer optimization problem (1). In a first example, the purpose is to show that the strategy leads to desirable results when applying it repeatedly on a simple function. In a second experiment, the algorithms are then applied to a standard benchmark problem in global and simulation-based optimization.

The focus, and purpose, of this article is to a perform a theoretical investigation of whether a maximum loss strategy can be applied to global optimization, and to analyze the principles that such an application would rely on. Hence, the focus is not to construct and validate a competitive algorithm. Therefore—although generally valid—the dimension of the decision space is limited to $\mathbb{R}^2$ in the numerical studies. In higher dimensions, however, the problem of implementing an efficient method for finding a domain D of uniqueness for the minimax problem remains. The algorithms are coded in MATLAB® ([12]) and the numerical tests have been run on a standard desktop computer.

6.1 Problem I - A concave quadratic objective function

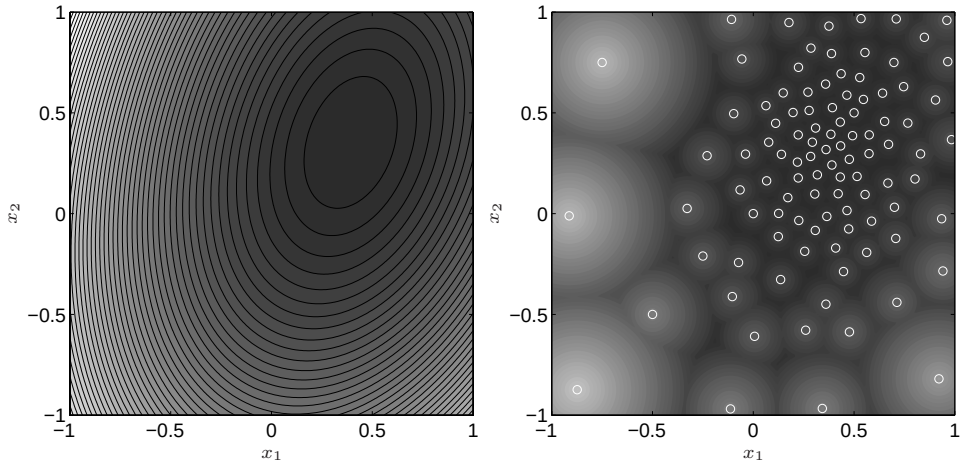
The following simple problem in which a concave quadratic function is to be maximized over a box-constrained domain is used to illustrate the behaviour of the min-

imax strategy:

$$\begin{aligned} \underset{\mathbf{x}}{\mathrm{m}axim\mathit{ize}} \quad & f(\mathbf{x}) := (\mathbf{x}-\mathbf{q})^{\mathrm{T}}\mathbf{Q}(\mathbf{x}-\mathbf{q}), \\ \text{subject to} \quad & -1 \le x_1 \le 1, \\ & -1 \le x_2 \le 1, \end{aligned} \tag{38}$$

$$\text{where } \mathbf{Q} = \begin{pmatrix} -1 & 0.2 \\ 0.2 & -0.5 \end{pmatrix} \text{ and } \mathbf{q} = \begin{pmatrix} \pi/8 & \sqrt{2}/4 \end{pmatrix}^{\mathrm{T}}.$$

In Figure 4(a) the level curves of f is shown and in Figure 4(b) the upper envelope function is shown together with the 100 first sample points that are found by applying the minimax strategy repeatedly with the evaluation set at each iteration of the main algorithm, described in Section 2.2, updated with the latest minimax solution $\mathbf{y}^*$. Here, the mid point $(0, 0)$ has been used as the sole starting member of the evaluation set and the Lipschitz constant has been estimated adaptively based on the points in the evaluation set at each main iteration. As can be seen, the sampling becomes more dense in regions with higher function values, a behavior which obviously is a desirable property since the larger the function value, the more interesting the region.



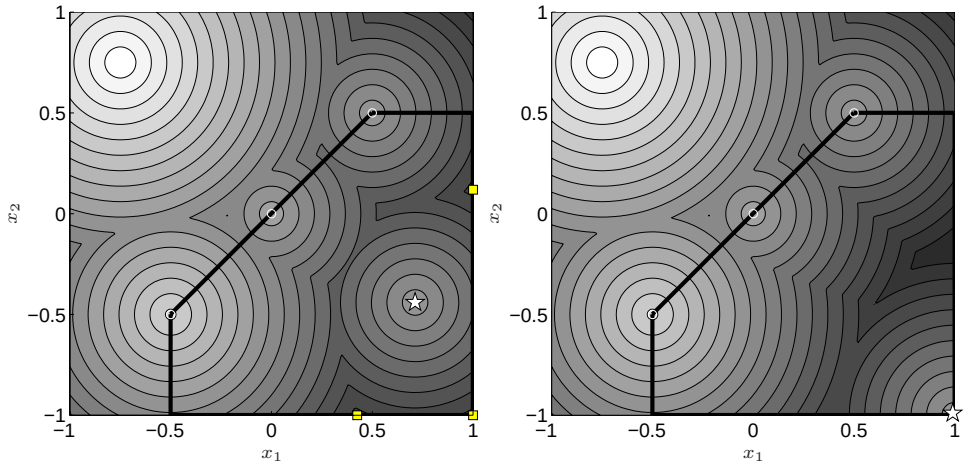
(a) Level curves for f in (38).

(b) The upper envelope $\mathcal{U}_{\mathcal{F}_{N,c}}$ and the found sample points (rings).

Figure 4: A concave quadratic test function.

In Figure 5, the evaluated points are shown in a particular iteration of the main algorithm, in Figure 5(a) using the minimax strategy and in Figure 5(b) using the maximum gain strategy. To find the region D of uniqueness, an ad-hoc method based on the saddle points of the envelope function (whose details are left out here) has

been used. Also for the maximum gain strategy it has been assumed that the function value at the next evaluation point, i.e., $\hat{\mathbf{x}}$, is equal to that of the evaluated point with the largest value found so far; this implying the worst-case optimality gap as shown in Lemma 3.3. The figures show that the worst-case optimality gap becomes lower when using the minimax strategy compared to when employing the maximum gain strategy. This behavior is expected since the worst-case optimality gap is essentially what is optimized within the minimax strategy.



(a) Minimax solution.

(b) Maximum gain solution.

Figure 5: Results at a particular iteration using the minimax (a) and the maximum gain (b) strategies. The star denotes in (a) the optimal minimax solution (with the corresponding control points marked with squares) and in (b) the maximum gain solution. The black edges define the region D.

If one can afford to evaluate a larger number of samples, it is not absolutely clear that minimax is superior to maximum gain in terms of the resulting optimality gap. The minimax strategy is a greedy strategy for minimizing the long-term optimality gap with no guarantee of being the best one. The maximum gain strategy, however, does not focus explicitly on the optimality gap at all. In the next example, we give some numerical results from a comparison of the minimax and the maximum gain strategies, using a larger number of samples.

6.2 Problem II - A standard test problem

The following is a modification of the standard Branin test problem for global optimization [3], having three global optima in a box-constrained subset of $\mathbb{R}^2$. We have transformed the original minimization problem to an equivalent maximization one. The modified problem (39), given by maximizing (the negation of) the logarithm of

the original objective function, possesses a lower Lipschitz constant than does the original one. Aside from this, the theoretical properties of the problem are, however, unchanged.

$$\begin{aligned} \underset{\mathbf{x}}{\text{m}axim\!ize} \quad & f(\mathbf{x}) := -\log\big(f_{bra}(\mathbf{x})\big), \\ \text{subject to} \quad & f_{bra}(\mathbf{x}) = \Big(x_2 - \frac{51}{40\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\Big)^2 + 10\Big(1 - \frac{1}{8\pi}\Big)\cos x_1 + 10, \\ & -5 \le x_1 \le 10, \\ & 0 \le x_2 \le 15. \end{aligned} \tag{39}$$

The Lipschitz constant of f in the given domain is $c \approx 5.34$ and is attained at $\mathbf{x} \approx (-3.33, 12.25)$ (compared to $c \approx 114$ for the original function f_{bra}).

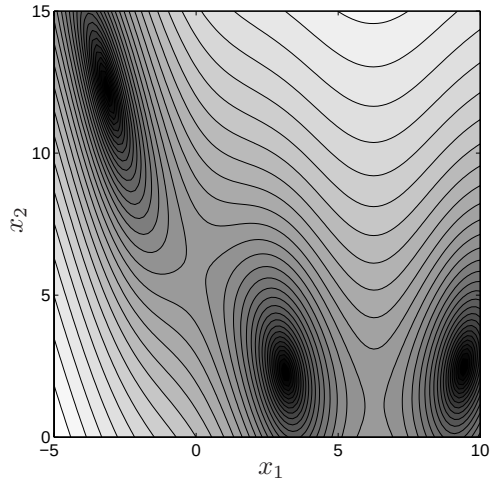
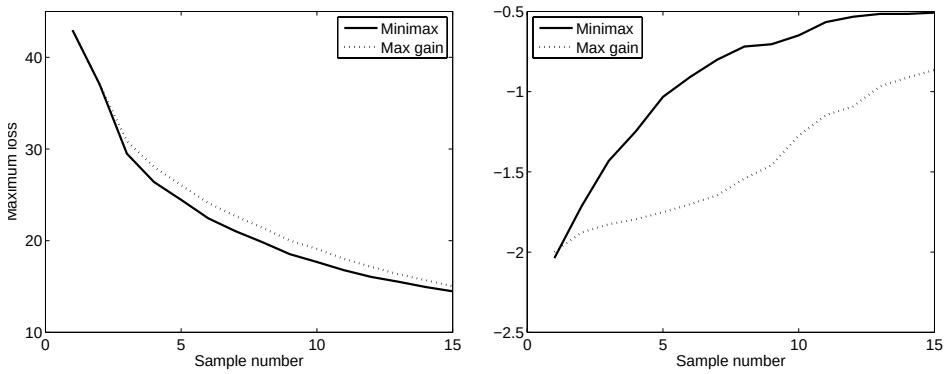


Figure 6: Level curves of the function f in (39).

In a numerical study to compare the average performance of the minimax strategy compared to that of the maximum gain strategy, we have applied both strategies 50 times to problem (39). In each run, we start with a sample set of 5 points positioned in a randomized latin hypercube. Thereafter, 15 more samples are sequentially generated, using the minimax and the maximum gain strategy, respectively. Here we employ the true Lipschitz constant, $c = 5.34$.

In Figure 7(a) the averaged maximum loss, i.e., (20) inserted in(10), is compared for the minimax and the maximum gain strategies. From the figure, we conclude that, by using the minimax strategy when generating new evaluation points, the optimality gap will in the worst case be smaller than when employing the more traditional maximum gain strategy. This seems to be particularly true for a small number of samples, which can also be expected. This since the minimax problem is in each main iteration solved locally over a region D and the number of local maxima of the upper envelope is increased with a larger number of sample points. Therefore,



(a) Maximal loss. (b) Best solution found.

Figure 7: Results when applying the minimax and the maximum gain strategy, respectivly, on (39). The results are averaged over 50 runs.

by decreasing the sizes of the suitable domains, the amount of extra information used by the minimax strategy compared to that of the maximum gain strategy decreases with the size of the evaluation set. Therefore, the advantage of the minimax strategy is also reduced.

Figure 7(b) shows that the minimax strategy for this test problem yields a better lower bound on the global optimum value compared to that generated by the maximum gain strategy. This result is somewhat surprising since the method is focusing on the upper bound, and not explicitly on the lower bound.

6.3 Discussion

Concluding the numerical experiments, the minimax strategy behaves as desired with many (few) sample points in regions with high (low) function values. Different strategies for selecting the value of the Lipschitz constant c affect the balance between space filling, i.e., global search, and sampling within promising regions, i.e., local search. In some situations (see Figure 5) the minimax strategy clearly chooses better sample points. In a comparison using a classical test problem we have shown that the minimax strategy does lead to a reduced worst-case optimality gap compared to the minimax strategy. However, the computational effort required for solving the minimax problem compared to that required for the maximum gain problem is significantly larger with yet another solution step according to the following: First, both methods solves for the maximum gain solution $\hat{\mathbf{x}}$. Then, in the minimax strategy, a suitable domain D is constructed, which is in general not an easy task, especially not in high dimensions. Thereafter a similar problem as the one for $\hat{\mathbf{x}}$ has to be solved to find $\mathbf{y}^*$.

If the minimax strategy is to be preferred over the maximum gain strategy for a global maximization problem, the following two main aspects have to be considered:

1. What is the computational effort for evaluating the objective function f compared to that of solving the minimax problem (34)? This effort should be comparatively high for the minimax strategy to be preferred.

2. What is the main concern for the optimization – a guarantee on a small gap in terms of objective value for the resulting point compared to the true optimum value, or the hope of finding a point whose objective value is as high as possible? For the minimax strategy to be preferred, the former concern should be the most important.

7 Conclusions and outlook

In this article we have presented and explored a minimax principle for global optimization, a principle which apparently has not been used before. Although the principle is conceptually appealing it is far from straightforward to implement it into a working computer code. Here we have focused on applying it to the Lipschitz space of functions. Finding a new evaluation point according to this strategy involves three steps. The first step is to find the point where the current maximum loss is attained (which is also the maximum gain solution). The second step is to find an appropriate reference set for the minimax problem, i.e., the domain in which the solution is to be sought. The third step is to compute the minimax solution. For the local optimizations in steps one and three, we have developed algorithms based on sequential linearization. For step one, global optimality is ensured by a Branch-and-Bound technique. Practically it has turned out that step two, finding the reference set, is the most difficult and the most critical step for the whole strategy. On one hand, it is desirable that the reference set is as large as possible so that the minimization of the local maximum loss also decreases the global maximum loss effectively in the long run. On the other hand, the reference set should define a unique minimax solution and be easy to handle within the linear programming problems emerging from the sequential linearization (e.g. by being a convex polytope). Some theoretical aspects on the reference set are described and ideas for procedures for choosing the reference set are discussed. However, to implement the minimax principle for the Lipschitz space in higher dimensions these procedures needs to be improved further. It is also in higher dimensions that this strategy possesses better odds to be favorable to the maximum gain strategy since the boundary of a set in higher dimensions is more dominant (e.g., the number of extreme points of a hyperrectangle grows exponentially with the dimension).

A natural question is whether it is possible to implement the minimax strategy for other spaces of functions with higher regularity than the Lipschitz space, such as Sobolev spaces [1] or the native spaces for radial basis functions (RBF) [4, 6, 20]. We then partition the problem of finding a new evaluation point into the same steps as for the Lipschitz case. The first step, to find the maximum gain solution, leads to

a global optimization problem of the upper envelope function (which hopefully is relatively cheap to evaluate); it is in that respect very similar to the Lipschitz case. The following steps are more unclear. In general it will not be possible to formulate the minimax problem in terms of the upper envelope function. Since the minimax solution should be balanced in the sense given by Proposition 4.8, it may be possible to generalize the concept of control points (at which the minimal maximum loss is attained supposing the worst-case function value $f(\mathbf{y}^*)$ at the optimal minimax solution $\mathbf{y}^*$) for some auxiliary function. Preliminary studies in the RBF case have shown that this function involves the minimal norm interpolant and the power function measuring the uncertainty of the interpolant at different points. The analysis of these subproblems depends, however, much on the underlying function space and may be very complicated to carry out.

The high computational effort associated with our approach may seem discouraging. However, for a successful application, exact solutions to the minimax problems are not necessary; it is enough if the algorithm terminates at an approximately optimal solution with a small maximum loss. Therefore, one should develop estimates for approximate solutions and simplify the various problems involved according to this. A successful tool for this might be heuristic methods for solving the smallest enclosing ball problem (28).

The minimax and the maximum gain solutions differ the most when the maximum gain solution is located at the boundary of the domain. If the maximum gain solution is in the interior the solutions might be very close (indeed identical for the Lipschitz case in one dimension). Thus, simplifying assumptions, such as $\mathbf{y}^* = \hat{\mathbf{x}}$ whenever $\hat{\mathbf{x}}$ is not close to the boundary, might be reasonable from a computational point of view. However, in many engineering applications, very time-consuming simulations are needed for the computation of objective function values. Then it may be worthwhile to use a costly algorithm for the selection of sample points, provided that it can be proved efficient.

As a final comment, the definition of maximum loss is in this article generally stated, and may therefore be used also in other contexts. One can, for example, think of product portfolio optimization where the function set is interpreted as the set of utility functions for the potential customers. A product family strategy can then be formulated as that to decide on a limited number of products to make available such that the maximum loss in the end is minimized, meaning that the distance in utility for the customer with the worst distance is minimized. Of course, putting such a strategy into practice will involve many challenges besides the problems of the kind studied in this article, however, the concept presented can be used for a formalization of a strategy for product portfolio optimization.

A Appendix

Proof of Lemma 3.3 Pick a vector $\mathbf{y} \in \Omega$ arbitrarily. Suppose first that $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}) \leq \max_{k=1,\ldots,N} f_k$. Then, for any $\mathbf{x} \in \Omega$, it holds that $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k \leq \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}) \leq c\|\mathbf{x} - \mathbf{y}\|$, where the latter inequality is due to the Lipschitz

property, and therefore the right hand side of (23) equals $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k$. Since $s(\mathbf{y}) \leq \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}) \leq \max_{k=1,\ldots,N} f_k$, $s(\mathbf{x}_k) = f_k$, $k = 1, \ldots, N$, and

$$\max_{s \in \mathcal{F}_{N,c}} s(\mathbf{x}) = \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}),$$

it follows that also the left hand side of (23) equals $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k$.

Now, suppose conversely that, $\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y}) \geq \max_{k=1,\ldots,N} f_k$ holds and let $\alpha \in [\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}), \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]$. By Kirszbraun's Theorem ([11]), $\exists s \in \mathcal{F}_{N,c}$ with $s(\mathbf{y}) = \alpha$. By the Lipschitz assumption for the function set $\mathcal{F}_{N,c}$, it then holds that

$$s(\mathbf{x}) \leq s(\mathbf{y}) + c\|\mathbf{x} - y\| = \alpha + c\|\mathbf{x} - y\|, \qquad \mathbf{x} \in \Omega.$$

Define the function $\xi : \mathbb{R} \times \Omega \times \Omega \to \mathbb{R}$ by $\xi(\alpha, \mathbf{x}, \mathbf{y}) := \max_{s \in \mathcal{F}_{N,c} : s(\mathbf{y}) = \alpha} \{s(\mathbf{x})\}$. By the definition (11) and Proposition 3.1, for any $\alpha \in [\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}), \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]$ and $\mathbf{x}, \mathbf{y} \in \Omega$ it then holds that

$$\begin{aligned}
\xi(\alpha, \mathbf{x}, \mathbf{y}) &= \sup_{s \in \mathcal{B}(\Omega) : s(\mathbf{y}) = \alpha} \{s(\mathbf{x}) : s(\mathbf{x}) \leq f_k + c\|\mathbf{x} - \mathbf{x}_k\|, k = 1, \ldots, N\} \\
&= \sup_{s \in \mathcal{B}(\Omega) : s(\mathbf{y}) = \alpha} \left\{s(\mathbf{x}) : s(\mathbf{x}) \leq \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x})\right\} \\
&= \min \left\{\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}),\, \alpha + c\|\mathbf{x} - \mathbf{y}\|\right\}.
\end{aligned}$$

Now, denoting the left hand side of the equation (23) by $\psi : \Omega \times \Omega \to \mathbb{R}$, where $\psi(\mathbf{x}, \mathbf{y}) := \max_{s \in \mathcal{F}_{N,c}} \left\{s(\mathbf{x}) - \max_{\mathbf{x}' \in E_N \cup \{\mathbf{y}\}} \{s(\mathbf{x}')\}\right\}$ it holds that

$$\begin{aligned}
\psi(\mathbf{x}, \mathbf{y}) &= \max_{s \in \mathcal{F}_{N,c}} \left\{s(\mathbf{x}) - \max\left\{\max_{k=1,\ldots,N} f_k,\, s(\mathbf{y})\right\}\right\} \\
&= \max_{\alpha \in [\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}), \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]} \{\eta(\alpha, \mathbf{x}, \mathbf{y})\},
\end{aligned}$$

where, for any $\alpha \in [\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}), \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]$ and $\mathbf{x}, \mathbf{y} \in \Omega$,

$$\begin{aligned}
\eta(\alpha, \mathbf{x}, \mathbf{y}) &:= \max_{s \in \mathcal{F}_{N,c} : s(\mathbf{y}) = \alpha} \left\{s(\mathbf{x}) - \max\left\{\max_{k=1,\ldots,N} f_k,\, \alpha\right\}\right\} \\
&= \max_{s \in \mathcal{F}_{N,c} : s(\mathbf{y}) = \alpha} \{s(\mathbf{x})\} - \max\left\{\max_{k=1,\ldots,N} f_k,\, \alpha\right\} \\
&= \xi(\alpha, \mathbf{x}, \mathbf{y}) - \max\left\{\max_{k=1,\ldots,N} f_k,\, \alpha\right\} \\
&= \min\left\{\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}),\, \alpha + c\|\mathbf{x} - \mathbf{y}\|\right\} - \max\left\{\max_{k=1,\ldots,N} f_k,\, \alpha\right\} \\
&= \min\Big\{\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \alpha, \\
&\qquad \alpha + c\|\mathbf{x} - \mathbf{y}\| - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x} - \mathbf{y}\|\Big\}.
\end{aligned}$$

For $\bar{\alpha} := \max_{k=1,\ldots,N} f_k$ and any $\varepsilon \neq 0$ it holds that

$$\eta(\bar{\alpha}, \mathbf{x}, \mathbf{y}) = \min\left\{\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \bar{\alpha},\, c\|\mathbf{x} - \mathbf{y}\|\right\} \geq \eta(\bar{\alpha} + \varepsilon, \mathbf{x}, \mathbf{y}).$$

It then follows that the maximum value $\max_{\alpha\in[\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}),\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]}\eta(\alpha,\mathbf{x},\mathbf{y})$ is attained at $\alpha = \bar{\alpha} = \max_{k=1,\ldots,N} f_k$. By assumption, $\bar{\alpha} \leq \mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})$, and from the definitions (7) and (11) it then follows that $\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{x}) \leq \bar{\alpha}$, $\mathbf{x} \in \Omega$. Hence, $\max_{k=1,\ldots,N} f_k \in [\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}),\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]$, thus implying that

$$\begin{aligned}\psi(\mathbf{x},\mathbf{y}) &= \max_{\alpha\in[\mathcal{L}_{\mathcal{F}_{N,c}}(\mathbf{y}),\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{y})]}\{\eta(\alpha,\mathbf{x},\mathbf{y})\}\\ &= \eta(\max_{k=1,\ldots,N} f_k,\mathbf{x},\mathbf{y})\\ &= \min\left\{\mathcal{U}_{\mathcal{F}_{N,c}}(\mathbf{x}) - \max_{k=1,\ldots,N} f_k,\, c\|\mathbf{x}-\mathbf{y}\|\right\}\end{aligned}$$

and the lemma follows. ∎

References

[1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev spaces*, vol. 140 of Pure and Applied Mathematics, Elsevier / Academic Press, New York, NY, USA, second ed., 2003.

[2] M. S. BAZARAA, H. D. SHERALI, AND C. M. SHETTY, *Nonlinear programming Theory and Algorithms*, John Wiley & Sons, Hoboken, NJ, USA, third ed., 2006.

[3] F. H. BRANIN, *Widely convergent method for finding multiple solutions of simultaneous nonlinear equations*, IBM Journal of Research and Development, 16 (1972), pp. 504–522.

[4] M. D. BUHMANN, *Radial basis functions*, vol. 12 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2003.

[5] A. R. CONN, K. SCHEINBERG, AND L. N. VICENTE, *Introduction to Derivative-Free Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, PA, USA, 2009.

[6] G. E. FASSHAUER, *Meshfree approximation methods with MATLAB*, vol. 6 of Interdisciplinary Mathematical Sciences, World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007.

[7] H. M. GUTMANN, *A radial basis function method for global optimization*, Journal of Global Optimization, 19 (2001), pp. 201–227.

[8] P. HANSEN, B. JAUMARD, AND S.-H. LU, *Global optimization of univariate Lipschitz functions I: Survey and properties*, Mathematical Programming, 55 (1992), pp. 251–272.

[9] D. R. JONES, *A taxonomy of global optimization methods based on response surfaces*, Journal of Global Optimization, 21 (2001), pp. 345–383.

[10] D. R. JONES, M. SCHONLAU, AND W. J. WELCH, *Efficient global optimization of expensive black-box functions*, Journal of Global Optimization, 13 (1998), pp. 455–492.

[11] M. D. KIRSZBRAUN, *Über die zusammenziehende und Lipschitzsche transformationen*, Fundamenta Mathematicae, 22 (1934), pp. 77–108.

[12] MATLAB, version 7.9.0 (R2009b). The Mathworks Inc., Natick, MA, USA, `www.mathworks.se/products`.

[13] D. Q. MAYNE AND E. POLAK, *Outer approximation algorithm for nondifferentiable optimization problems*, Journal of Optimization Theory and Applications, 42 (1984), pp. 19–30.

[14] C. C. MEEWELLA AND D. Q. MAYNE, *An algorithm for global optimization of Lipschitz continuous functions*, Journal of Optimization Theory and Applications, 57 (1988), pp. 307–322.

[15] R. H. MLADINEO, *An algorithm for finding the global maximum of a multimodal, multivariate function*, Mathematical Programming, 34 (1986), pp. 188–200.

[16] F. NIELSEN AND R. NOCK, *Approximating smallest enclosing balls*, in Computational science and its applications—ICCSA 2004. Part III, vol. 3045 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 2004, pp. 147–157.

[17] S. A. PIYAVSKII, *An algorithm for finding the absolute extremum of a function*, USSR Computational Mathematics and Mathematical Physics, 12 (1972), pp. 57–67.

[18] B. O. SHUBERT, *A sequential method seeking the global maximum of a function*, SIAM Journal on Numerical Analysis, 9 (1972), pp. 379–388.

[19] J. SYLVESTER, *A question in the geometry of situation*, Quarterly Journal of Mathematics, 1 (1857), p. 79.

[20] H. WENDLAND, *Scattered data approximation*, vol. 17 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2005.