

Norm-induced Cuts: Optimization with Lipschitzian Black-box Functions

ADRIAN GÖSS, ALEXANDER MARTIN, SEBASTIAN POKUTTA, KARTIKEY SHARMA

ABSTRACT. Optimal control problems usually involve constraints which model physical states and their possible transitions. These are represented by ordinary or partial differential equations (ODEs/PDEs) which add a component of infinite dimension to the problem. In recent literature, one method to simulate such ODEs/PDEs are physics-informed neural networks. Typically, neural networks are highly non-linear which makes their addition to optimization problems challenging. Hence, we leverage their often available Lipschitz property on a compact domain. The respective Lipschitz constants have to be computed only once and are accessible thereafter.

We present a method that, based on this property, iteratively adds cuts involving the violation of the constraints by the current incumbent and the Lipschitz constant. Hereby, the “shape” of a cut depends on the norm used. We prove the correctness of the method by showing that it either returns an optimal solution when terminating or creates a sequence with optimal accumulation points. This is complemented by a discussion about the termination in the infeasible case, as well as an analysis of the problem complexity. For the analysis, we show that the lower and upper iteration bound asymptotically coincide when the relative approximation error goes to zero. In the end, we visualize the method on a small example based on a two-dimensional non-convex optimization problem, as well as stress the necessity of having a globally optimal oracle for the sub-problems by another example.

1. INTRODUCTION

Mathematical optimization problems exist in a variety of levels of specificity and are investigated in each kind. Here, we address the following general form of an optimization problem.

$$\begin{aligned} \min_x & f(x) \\ \text{s.t. } & r(x) \leq \mathbf{0}, \\ & x \in \Omega. \end{aligned} \tag{P}$$

We assume Ω to be a non-empty, compact subset of the finite dimensional vector space X over a field F , i.e., $\Omega \subset X \cong F^n$, $n \in \mathbb{N}$. The space X is associated with a norm $\|\cdot\|_X$. That is, in particular, $(X, \|\cdot\|_X)$ is a normed space. Set Ω serves as a domain for the function $r : \Omega \rightarrow \mathbb{R}^m$, mapping it into the m -dimensional real space for some $m \in \mathbb{N}$. We mention that $\mathbf{0} \in \mathbb{R}^m$ denotes the vector of m zeros and the inequality in (P) is to be interpreted component-wise. In addition, function r is assumed to be Lipschitz continuous on Ω with a known Lipschitz constant $L > 0$, i.e., for any $x, y \in \Omega$, it holds that

$$\|r(x) - r(y)\|_{\mathbb{R}^m} \leq L\|x - y\|_X,$$

Date: May 11, 2023.

2010 Mathematics Subject Classification. 90C26, 90C30, 90C56, 90C60 .

Key words and phrases. Mixed-integer nonlinear optimization, Decomposition, Lipschitz optimization, black-box function evaluation

with $\|\cdot\|_{\mathbb{R}^m}$ denoting some norm in \mathbb{R}^m . Note that the value of the constant L depends on the specific choice of norms. From now on, we will omit the subscripts for each norm, where it is unambiguous. Further, we emphasize that the explicit form of r can be unknown, though its evaluation at any point $x \in \Omega$ must be possible. Lastly, the objective function $f : \Omega \rightarrow \mathbb{R}$ is assumed to be continuous on Ω .

We tackle the problem (P) for an arbitrary field F in order to stress the wide applicability of the approach presented in this article. Nevertheless, the main part of the mathematical literature in the area of optimization or programming, considers $F = \mathbb{R}$. Doing so, solving (P) is considered as *constrained global optimization* or *constrained global programming*. Depending on the properties of the involved objects f , r , and Ω , the field of mathematical programming and the developed approaches change. For example, linearity of f and r combined with box-constraints defining Ω leads to the class of *linear programming*. Convexity in Ω and r determines the problem for *convex programming*, whereas *non-linear programming* tackles general non-linear (possibly continuously differentiable) constraints r . A broad overview of this topic is provided in the book [9].

In our problem formulation, it is assumed that we can evaluate the constraint function r without having an explicit representation. Such problems are tackled in the field of *black-box optimization* or, more commonly, *derivative-free optimization*. Here, the applicability of any particular approach also depends on the structure of Ω . For example, if Ω is defined by linear constraints only, a directional direct-search approach can be leveraged which applies a Lagrangian reformulation; see [12]. For non-smooth constraints defining Ω , filtering techniques as proposed in [1] may be more suitable. Further, in terms of direct search, a well-known method called *mesh adaptive direct search (MADS)* (see [2]) is capable of handling general constraints in Ω , too. Alternatively, [3] proposes a trust-region interpolation-based approach to tackle such problems. For a general overview on constrained derivative-free optimization, we refer to [4, Chapter 13].

As mentioned in [18, Chapter 6], even when minimizing a continuous function over a simple box-shaped domain, in a worst-case situation a finite bound on the number of iterations to find an optimal solution cannot be expected. Hence, as described earlier, we make the assumption of a Lipschitz property to r . Since continuous differentiability of r on Ω already implies the Lipschitz continuity of r , we consider our assumption as a rather weak one. In particular, we touch the field of *Lipschitz optimization* which is treated in [9, Chapter XI]. Here, problems with univariate, Lipschitzian objectives are tackled by saw-tooth cover methods. More general problems, similar to (P), are solved by branch-and-bound schemes or outer approximation.

The authors of [6] leverage the Lipschitz continuity of constraint functions to tackle a specific kind of problem. They assume Ω to be a polytope, variable vector x to be partially integral, and r to be only involved in equality constraints on the continuous part of x . For this kind of problem, they propose a method which relaxes the equality constraints and iteratively adds cuts induced by the maximum norm to the set Ω . Using binary variables to make a model of these cuts, they obtain a mixed-integer program which is solved in each iteration. The principles of our method align with this approach. Nevertheless, we consider an arbitrary compact domain Ω with some norm and allow the constraint involving r to be an inequality instead. Hence, our method is a generalization of the one in [6], even though we do not discuss properties of the resulting problem class after adding the cuts due to the general character of the original problem.

The authors of [5] discussed the Lipschitz continuity of neural networks and introduced a method to compute the respective Lipschitz constants. In [13], the

authors use this Lipschitz continuity of neural networks, in order to incorporate them into the optimization problem. They leverage an approach introduced in [16], relaxing the Lipschitz constraints and iteratively refining the domain to smaller parallelepipedal shapes in order to solve bilevel problems with an unknown lower problem. Our motivation is also based on incorporating the Lipschitz continuity of neural networks. In contrast to [13], we aim to deal with ordinary or partially differential equations in the context of physics, in particular with gas network control (see [10]). These can be approximated by machine learning techniques, e.g., see [15], and hopefully re-integrated into the optimization problem with the underlying method. This justifies the inclusion of non-convex constraints by our method to trade off against infinite dimensional differential equation constraints.

Our contribution consists in presenting a method generalizing the one in [6] which is thus applicable to a wide and abstract class of problems. We prove its correctness by showing that it either returns an optimal point or creates a sequence with optimal accumulation points. In terms of problem complexity, we show asymptotic convergence of its lower and upper complexity bounds motivating the introduction of the method. At the end, we stress the necessity of a globally optimal oracle for the solution of the sub-problems by illustrating the convergence to the maximum in an one dimensional minimization problem with a locally optimal oracle.

The remainder of the paper is structured as follows. In Section 2, we give basic definitions in order to formulate the introduced method clearly and to keep the proof of correctness in Section 3 concise. Section 4 is dedicated to an analysis regarding a termination bound in the case of an infeasible problem w.r.t. the properties of r and Ω , as well as to a discussion about the problem complexity. Lastly, we give an illustration of the procedure and present an example which shows its weakness in terms of local solutions. We close the paper with Section 5 by summarizing the content and pointing to further research directions.

2. NORM-INDUCED CUT METHOD

In this section, we formulate a method that is able to tackle Problem (P). We start by giving precise definitions which clarify the occurring mathematical objects. Afterwards, the solution method is introduced, iteratively solving a relaxed problem and adding cuts in order to enforce feasibility. The mentioned cuts are implicitly determined by a point in Ω , the norm, and the function r with its Lipschitz constant L . We define them formally below.

Definition 2.1. *Let $y \in \Omega$ such that $r(y) > \mathbf{0}$ and let $L > 0$ define the Lipschitz constant of function r . Then, for $x \in \Omega$, we call the inequality*

$$\frac{\|r(y)\|}{L} \leq \|x - y\|,$$

a norm-induced cut in y .

Whenever referring to the *radius of a norm-induced cut*, we consider the constant term $\|r(y)\|/L$. In the case of $X \cong \mathbb{R}^n$ we note that a norm-induced cut is non-convex in variable x independent from the choice of norm.

Now, we denote the feasible set of (P) by

$$Q := \{x \in \Omega \mid r(x) \leq \mathbf{0}\} \subseteq \Omega.$$

In the following, we define a family of sets which involve norm-induced cuts and which are relaxations to Q as shown below in Lemma 3.3.

Definition 2.2. Let $k \in \mathbb{N}$ be a positive integer and $(x^i)_{i=0, \dots, k-1} \subseteq \Omega$ a sequence such that $r(x^i) > \mathbf{0}$ for all $i = 0, \dots, k-1$. Then, we define

$$Q_k := \left\{ x \in \Omega \mid \forall i \in \{0, \dots, k-1\} : \frac{\|r(x^i)\|}{L} \leq \|x - x^i\| \right\},$$

the k th relaxed set of Q . It results from Ω by relaxing the inequality constraint $r(x) \leq \mathbf{0}$ and adding k norm-induced cuts for each element x^i of the sequence. For $k = 0$, we naturally define $Q_0 := \Omega$.

The resulting family of relaxed problems is denoted as

$$\min_{x \in Q_k} f(x), \tag{P}_k$$

for $k \in \mathbb{N}_0$. We interpret k as the iteration index of our method. The following remark shows a connection between two consecutive relaxed problems (P_k) and (P_{k+1}) , in particular, between their feasible sets Q_k and Q_{k+1} , respectively.

Remark 2.3. Let $k \in \mathbb{N}_0$ and let $(x^i)_{i=0, \dots, k} \subseteq \Omega$ be the sequence to define the k th and the $(k+1)$ th relaxed set Q_k and Q_{k+1} , respectively. Then, Q_{k+1} is given by

$$Q_{k+1} = Q_k \cap \left\{ x \in \Omega \mid \frac{\|r(x^k)\|}{L} \leq \|x - x^k\| \right\}.$$

In other terms, the $(k+1)$ th relaxed set results from the k th relaxed set by adding the *norm-induced cut* in x^k , i.e.,

$$\frac{\|r(x^k)\|}{L} \leq \|x - x^k\|.$$

Now, we are able to state our method for solving problem (P) . It is called *Norm-induced Cutting Method (NIC)*, following its main functionality of iteratively adding norm-induced cuts.

In particular, it starts by relaxing the explicit inequality constraint in (P) , i.e., $r(x) \leq \mathbf{0}$, and solves the resulting relaxed problem (P_k) for $k = 0$. It is assumed that we have a suitable method to solve (P_k) for all $k \in \mathbb{N}_0$. If the relaxed problem is infeasible, so is the original one and a message is returned accordingly. Otherwise, the method checks if the solution of the relaxed problem x^0 is also feasible for the original problem by evaluating function r at x^0 and checking for the inequality. If it is satisfied, the current incumbent x^0 is returned as a solution to the original problem. Otherwise, a norm-induced cut in incumbent x^0 is added. This defines the set Q_1 . The resulting problem (P_1) is then solved in the next iteration.

In conclusion, the method adds norm-induced cuts as long as the incumbents are infeasible for the original problem (P) . Below we give a formal description of the **NIC** method in pseudo code. We point out that any specific stopping criterion is omitted, as we examine the method's behavior for an infinite number of iterations in Section 3. The practical case of satisfaction with an approximate solution is tackled in Section 4.

Note that the norm-induced cuts are added with a lower bound to the norm of the difference between the current incumbent and variable x . This bound is the radius of the norm-induced cut and includes the violation of $r(x) \leq \mathbf{0}$ at the incumbent and the Lipschitz constant L . Informally, the greater the violation or the smaller L , the greater the radius of the norm-induced cut and, thus, the greater the “excluded area” in the next iteration.

In order to discuss properties of the **NIC** method in the following sections, we introduce the notation of a ball and the terms ε -packing/-covering. We start with the former.

Norm-induced Cutting Method (NIC)

Input: Set Ω , function r , Lipschitz constant L

Output: A message of (P)'s infeasibility, a solution to (P), or a sequence $(x^i)_{i \in \mathbb{N}_0}$

```

1: set  $k \leftarrow 0$ 
2: solve (Pk)
3: if (Pk) is infeasible then
4:   return '(P) is infeasible'
5: else
6:   denote  $x^k$  as the solution of (Pk)
7: end if
8: if  $r(x^k) \leq \mathbf{0}$  then
9:   return solution  $x^k$  to (P)
10: end if
11: set  $k \leftarrow k + 1$ 
12: goto (2)
    
```

Definition 2.4. Let $\varepsilon > 0$, $\bar{x} \in X$, and $\|\cdot\|$ a norm on X . Then, we define the ε -ball around \bar{x} as

$$B_\varepsilon(\bar{x}) := \{x \in X \mid \|x - \bar{x}\| < \varepsilon\}.$$

For the case $\bar{x} = \mathbf{0}$, we write the short form

$$B_\varepsilon := B_\varepsilon(\mathbf{0}) = \{x \in X \mid \|x\| < \varepsilon\}.$$

For the fundamental mathematics regarding ε -packing/-covering, we refer to [17]. We use the notation from the lecture notes [8] and adapt the definition to fit our question.

Definition 2.5. Let $k \in \mathbb{N}_0$, $\varepsilon > 0$, and $\|\cdot\|$ a norm on X . Further, consider $A \subseteq X$ and $P := \{x^0, \dots, x^k\} \subseteq A$ a set of points in A .

- a) If $A \subseteq \bigcup_{i=0}^k B_\varepsilon(x^i)$, we call P an ε -covering of A . Furthermore, we define the minimal cardinality of such a P as

$$N(A, \|\cdot\|, \varepsilon) := \min \{k \mid \exists \varepsilon\text{-covering of } A \text{ with size } k\},$$

the covering number of A w.r.t. ε .

- b) If $\{B_{\varepsilon/2}(x^i)\}_{i=0, \dots, k}$ are pairwise disjoint, i.e., $\|x^i - x^j\| \geq \varepsilon$ for all $i, j = 0, \dots, k$, $i \neq j$, we call P an ε -packing of A . We also define the maximal cardinality of such a P , i.e.,

$$M(A, \|\cdot\|, \varepsilon) := \max \{k \mid \exists \varepsilon\text{-packing of } A \text{ with size } k\},$$

the packing number of A w.r.t. ε .

3. CONVERGENCE AND OPTIMALITY RESULTS

This section is dedicated to show the correctness of NIC. Depending on the feasibility of the original problem (P), we can show feasibility and even optimality of a returned solution or of the accumulation points of a sequence of incumbents produced by the method. In the end, we give a statement regarding the availability of local solutions in step 2. This is complemented by a negative example in Section 4. We start by remarking on the boundedness of (P) in case of feasibility.

Remark 3.1. As r is Lipschitz continuous, it is also continuous. By re-writing the feasible set of (P) as

$$Q = \Omega \cap r^{-1}((-\infty, 0]^m),$$

where $r^{-1}(\cdot)$ notes the pre-image of r , we can use the closure of $(-\infty, 0]^m$ to derive the closure of $r^{-1}(-\infty, 0]^m$. Together with the compactness of Ω , this implies that Q is compact. Therefore, Problem (P) minimizes a continuous function over a compact set and, thus, if $Q \neq \emptyset$, the problem has a finite solution value.

With this in mind, we state two lemmata. The first one ensures that NIC does not find any incumbent twice, the second one implies that (P_k) is a relaxation of (P) for each $k \in \mathbb{N}$, as mentioned in Definition 2.2.

Lemma 3.2. *Consider some $k \in \mathbb{N}_0$ and a solution x^k to (P_k) such that $r(x^k) > \mathbf{0}$. Then, $x^k \notin Q_{k+1}$, i.e., the incumbent x^k is infeasible for problem (P_{k+1}) .*

Proof. Let the assumptions as in the lemma hold true. As x^k is a solution to (P_k) , we have $x^k \in Q_k$. Due to strict positivity of $r(x^k)$ and L , it follows

$$\|x^k - x^k\| = 0 < \frac{\|r(x^k)\|}{L}.$$

Then, Remark 2.3 directly implies that $x^k \notin Q_{k+1}$ and the claim follows. \square

Lemma 3.3. *For any $k \in \mathbb{N}_0$, it holds that $Q \subseteq Q_k$.*

Proof. If $k = 0$, $Q_k = \Omega$ by construction and the claim follows from Definition 2.2.

Now, let $k \in \mathbb{N}$ arbitrary but fixed and let $(x^i)_{i=0, \dots, k-1}$ be a sequence to determine Q_k . Then, it holds true that

$$x^i \in \Omega \quad \wedge \quad r(x^i) > \mathbf{0},$$

for $i = 0, \dots, k-1$. Further, consider some $\bar{x} \in Q$, thus, $\bar{x} \in \Omega$ and $r(\bar{x}) \leq \mathbf{0}$. We can conclude that

$$0 < \frac{\|r(x^i)\|}{L} = \frac{1}{L} \|r(x^i) + r(\bar{x}) - r(\bar{x})\| \leq \frac{1}{L} \|r(\bar{x}) - r(x^i)\| \leq \|\bar{x} - x^i\|,$$

for all $i = 1, \dots, k-1$. Hence, for all $i = 0, \dots, k-1$, \bar{x} satisfies the norm-induced cut in x^i . With Definition 2.2 it follows $\bar{x} \in Q_k$ and the claim is proven. \square

In other terms, Lemma 3.3 shows that infeasibility of Problem (P_k) automatically implies the infeasibility of Problem (P), in particular, $Q_k = \emptyset \Rightarrow Q = \emptyset$. Therefore, step 4 associated with the corresponding if-clause is reasonable and NIC gives a correct answer if it stops in step 4. Further, it results from Lemma 3.3 that (P_k) is a relaxation of (P) for $k \in \mathbb{N}_0$. Therefore, the solution value of (P_k) , which is obtained during NIC, gives a lower bound to the value of (P). This can be used in meta-level solution frameworks, e.g., in Branch-and-Bound which is common in discrete optimization scenarios; see [11] for the original source.

In the following, we prove that NIC either outputs a feasible solution or produces a sequence with feasible accumulation points for (P) if running for infinite time.

Theorem 3.4. *Let (P) be feasible, i.e., $Q \neq \emptyset$. Then, either*

- a) *NIC stops with a feasible solution x^* for (P) in step 9, or*
- b) *NIC creates a sequence of incumbents $(x^i)_{i \in \mathbb{N}_0}$, there exists at least one convergent subsequence of this sequence, and each convergent subsequence has a limit feasible for (P). That is, $(x^i)_{i \in \mathbb{N}_0}$ has at least one accumulation point and all accumulation points of $(x^i)_{i \in \mathbb{N}_0}$ are feasible for (P).*

Proof. With the assumption and Lemma 3.3, it is clear that

$$\emptyset \neq Q \subseteq \bigcap_k Q_k. \tag{1}$$

Here, k indicates the iteration index and is specified in the case distinction below. From (1) we have that $Q_k \neq \emptyset$ and, thus, NIC does not stop in step 4. Hence, it either stops in step 9 or runs for infinite time. We start with the former case.

a) Let $k \in \mathbb{N}_0$ be the index of the stopping iteration. According to Definition 2.2, $Q_k \subseteq \Omega$ and, thus, the incumbent satisfies $x^k \in \Omega$. Since stopping in step 9 requires $r(x^k) \leq \mathbf{0}$, it follows that **NIC** returns $x^* := x^k \in Q$, i.e., a feasible point for (P).

b) **NIC** runs for infinite time, i.e., stopping in step 4 is not possible and the missing of a stop in step 9 is assumed. Again, as $Q_k \subseteq \Omega$ for all $k \in \mathbb{N}_0$, we receive a sequence $(x^i)_{i \in \mathbb{N}_0} \subseteq \Omega$. With Ω being compact, $(x^i)_{i \in \mathbb{N}_0}$ has at least one convergent subsequence $(x^{i_j})_{j \in \mathbb{N}_0}$ with

$$\lim_{j \rightarrow \infty} x^{i_j} = x^* \in \Omega.$$

We want to prove the feasibility of x^* for (P) by contradiction and, thus, assume $x^* \notin Q$, i.e., $r(x^*) > \mathbf{0}$. As **NIC** is assumed to not terminate, we have

$$\forall j \in \mathbb{N}_0 : r(x^{i_j}) > \mathbf{0}.$$

Furthermore, having compactness of Ω and the (Lipschitz) continuity of r on Ω , we can deduce the existence of an $\varepsilon > 0$, such that

$$0 < 2\varepsilon < \|r(x^*)\|.$$

Due to the convergence of $(x^{i_j})_{j \in \mathbb{N}_0}$ to x^* , there also exists a $j(\varepsilon) \in \mathbb{N}_0$ such that

$$\|x^* - x^{i_j}\| < \frac{\varepsilon}{L}, \quad \text{for all } j \geq j(\varepsilon). \quad (2)$$

That is, for all $j \geq j(\varepsilon)$, the element x^{i_j} of the subsequence is inside an ε/L -ball around x^* . Using the triangle inequality and the results from above, we can conclude the relation

$$\begin{aligned} 2\varepsilon < \|r(x^*)\| &= \|r(x^*) - r(x^{i_j}) + r(x^{i_j})\| \\ &\leq \|r(x^*) - r(x^{i_j})\| + \|r(x^{i_j})\| \\ &\leq L\|x^* - x^{i_j}\| + \|r(x^{i_j})\| \leq \varepsilon + \|r(x^{i_j})\|, \end{aligned} \quad (3)$$

for all $j \geq j(\varepsilon)$. In conclusion, we see that

$$\|r(x^{i_j})\| > \varepsilon, \quad \text{for all } j \geq j(\varepsilon).$$

That is, for all $j \geq j(\varepsilon)$, a norm-induced cut in $x^{i_j} \in Q_{i_j}$ of the form

$$\|x - x^{i_j}\| \geq \frac{\|r(x^{i_j})\|}{L} > \frac{\varepsilon}{L} \quad (4)$$

is added, resulting in Q_{i_j+1} . In order to show termination, Lemma 3.2 and (4) imply that it suffices to prove that only a finite number t of x^{i_j} can be contained in the original feasible set $Q_0 = \Omega$, having a pairwise distance (w.r.t. the norm) of at least ε/L . After adding all the respective cuts, **NIC** would end up with an empty Q_k for some k which is a contradiction and, thus, implies termination. We can write the problem of finding such a number as

$$\begin{aligned} &\max_{t \in \mathbb{N}} t \\ &\text{s.t. } x_\alpha \in \Omega, \quad \text{for all } \alpha = 1, \dots, t, \\ &\|x_\beta - x_\gamma\| \geq \frac{\varepsilon}{L}, \quad \text{for all } \beta, \gamma = 1, \dots, t, \beta \neq \gamma. \end{aligned} \quad (5)$$

In other terms, (5) aims to find a maximal ε/L -packing of Ω , compare Definition 2.5. This motivates us to set $M(\delta) = M(\Omega, \|\cdot\|, \delta)$ and $N(\delta) = N(\Omega, \|\cdot\|, \delta)$ the packing and covering number of Ω w.r.t. some $\delta > 0$, respectively. Therefore, determining the optimal value of (5) is equivalent to finding $M(\varepsilon/L)$. For normed spaces, it is well known that $M(2\delta) \leq N(\delta)$ (see, e.g., [17]) and, thus, $M(\varepsilon/L) \leq N(\varepsilon/(2L))$. That is, our claim of a finite termination is proven if there exists a finite $\varepsilon/(2L)$ -covering.

Now, as X is a finite dimensional vector space, it is isomorphic to \mathbb{R}^n with an isomorphism $\phi : X \rightarrow \mathbb{R}^n$. As Ω is compact, $\phi(\Omega)$ is also compact and analogously $\phi(B_{\varepsilon/(2L)})$ is open. By the theorem of Heine-Borel there exists a finite cover of $\phi(\Omega)$ with translations of $\phi(B_{\varepsilon/(2L)})$. Therefore, the equivalent statement holds true for Ω and $B_{\varepsilon/(2L)}$, and it follows $N(\varepsilon/(2L)) < \infty$. This shows $M(\varepsilon/L) < \infty$. Hence, **NIC** terminates in a finite amount of iterations, a contradiction. Therefore, $x^* \in \Omega$ and claim b) is proven. \square

In other terms, this shows that **NIC** solves the feasibility problem of **(P)** in case it really is feasible. The following remark takes the availability of possibly smaller Lipschitz constants into account and comments on their effects on Theorem 3.4.

Remark 3.5. *Let's assume we have access to the local Lipschitz constants L_x of r for each $x \in \Omega$. That is,*

$$\forall x \in \Omega \exists L_x > 0 : \|r(x) - r(y)\| \leq L_x \|x - y\|, \quad \text{for all } y \in \Omega. \quad (6)$$

First, note that the statements of Lemma 3.2 and Lemma 3.3 still hold when using the local Lipschitz constants L_{x^k} and L_{x^i} in the proofs, respectively. Second, as the global Lipschitz constant L is an upper bound to every local one, i.e., $L \geq L_x$ for all $x \in \Omega$, inequality (2) can be re-written to $\|x^ - x^{ij}\| < \varepsilon/L_{x^{ij}}$, leading to the same result as in inequality (3) and, thus, a possible greater right-hand side in inequality (4). Therefore, the availability of local Lipschitz constants can lead to larger cuts and may result in faster convergence.*

*Further, during the execution of **NIC** in iteration $k \in \mathbb{N}_0$, we have to consider the Lipschitz continuity of r only on Q_k . We note that, while this has no influence on the behavior of the algorithm per se, it might be possible to leverage a bounded domain such as Q_k to compute the Lipschitz constant efficiently, as well as obtain a lower constant. This leads to the same effect as mentioned above.*

*It is also possible to combine locality with the current feasible region. In particular, if we are in iteration $k \in \mathbb{N}_0$ and have solved **(P_k)** to receive x^k , we could compute L_{x^k} for $y \in Q_k$ instead of doing so for all $y \in \Omega$ as denoted in (6). Then, we could proceed and add the cut with respect to L_{x^k} .*

Besides the question of feasibility of the output, its quality/optimality also needs to be investigated. Note that Theorem 3.4 does not require the solution in step 2 to be optimal per se. Therefore, **NIC** produces a feasible output even if **(P_k)** is only solved for a feasible solution. However, when investigating the optimality of the returned solution or the accumulation points, respectively, we assume the global optimality of x^k for **(P_k)** in step 2 and apply a case distinction as in Theorem 3.4.

Theorem 3.6. *Let **(P)** be feasible, i.e., $Q \neq \emptyset$, and let step 2 produce a globally optimal point x^k for **(P_k)**, $k \in \mathbb{N}_0$. Then, either*

- a) **NIC** stops with a globally optimal solution x^* of **(P)** in step 9, or
- b) **NIC** creates a sequence of incumbents $(x^i)_{i \in \mathbb{N}_0}$, there exists at least one convergent subsequence of this sequence, and each convergent subsequence has a limit that is globally optimal for **(P)**. That is, $(x^i)_{i \in \mathbb{N}_0}$ has at least one accumulation point and all accumulation points of $(x^i)_{i \in \mathbb{N}_0}$ are globally optimal for **(P)**.

Proof. With **(P)** being feasible, Theorem 3.4 shows that **NIC** either stops in step 9 or iterates infinitely often. We start with the former case.

a) Let $k \in \mathbb{N}_0$ be the index of the stopping iteration, in which x^k is returned. According to Theorem 3.4, x^k is feasible for **(P)**, i.e., $x^k \in Q$, and, thus, it is

$$\min_{x \in Q} f(x) \leq f(x^k).$$

Further, by assumption x^k is optimal for (P_k) and Lemma 3.3 states $Q \subseteq Q_k$ which implies

$$f(x^k) = \min_{x \in Q_k} f(x) \leq \min_{x \in Q} f(x).$$

Together, we can conclude

$$\min_{x \in Q} f(x) = f(x^k),$$

which proves claim a).

b) Assuming **NIC** iterates infinitely, we denote $(x^i)_{i \in \mathbb{N}_0}$ as the resulting sequence of incumbents. Theorem 3.4 states that it has at least one accumulation point and that all its accumulation points are feasible for (P) . Hence, let $(x^{i_j})_{j \in \mathbb{N}_0}$ denote a convergent subsequence of $(x^i)_{i \in \mathbb{N}_0}$ with limit $x^* \in Q$. The feasibility of x^* for (P) immediately gives

$$\min_{x \in Q} f(x) \leq f(x^*).$$

Applying Lemma 3.3 on the index of the subsequence, leads to $Q \subseteq Q_{i_j}$ and, thus,

$$f(x^{i_j}) = \min_{x \in Q_{i_j}} f(x) \leq \min_{x \in Q} f(x),$$

for $j \in \mathbb{N}_0$. Leveraging the continuity of f when taking the limit, we can derive

$$\lim_{j \rightarrow \infty} f(x^{i_j}) = f(\lim_{j \rightarrow \infty} x^{i_j}) = f(x^*) \leq \min_{x \in Q} f(x).$$

Therefore, $f(x^*) = \min_{x \in Q} f(x)$ which finishes the proof. \square

So far, we have proven that **NIC** is suited to tackle feasible problems which have a structure like (P) . We can further extend the structure by either including several functions instead of one or by defining the function in question on lower dimensional parts of the feasible set. The combination of these two extension is also treatable. A formalization of this matter is given below.

Remark 3.7. As r can be multi-dimensional, i.e., $m > 1$, and $r(x) \leq \mathbf{0}$ is interpreted component-wise, we can re-write (P) as

$$\begin{aligned} & \min_x f(x) \\ & \text{s.t. } r_p(x) \leq \mathbf{0}, \quad \text{for } p = 1, \dots, R, \\ & \quad x \in \Omega, \end{aligned} \tag{P_R}$$

with $R \in \mathbb{N}$ and Lipschitz continuous functions $r_p : \Omega \rightarrow \mathbb{R}^{m_p}$, $m_p \in \mathbb{N}$ with their Lipschitz constants $L_p > 0$. Then, we can perform **NIC** analogously by checking for each $r_p(x^k) \leq \mathbf{0}$ before step 9. Moreover, if the condition is not satisfied for all $p = 1, \dots, R$, we can add one norm-induced cut for each p with $r_p(x^k) > \mathbf{0}$. In particular, such a cut is given by

$$\|x - x^k\| \geq \|r_p(x^k)\|/L_p.$$

This may result in more than one cut per iteration. One can see that versions of Lemma 3.2 and Lemma 3.3 adapted to the r_p 's are equally valid with such a procedure. Even further, the proofs of Theorem 3.4 and Theorem 3.6 can be derived analogously by replacing each occurrence of r and L with all r_p 's or the violated ones, respectively, and the associated L_p .

Apart from that, we can assume r in (P) to map from a lower dimensional subset of Ω into the \mathbb{R}^m . In particular, let $x = (x', x'') \in \Omega = \Omega' \times \Omega''$ with $\dim(\Omega')$, $\dim(\Omega'') \geq 1$, $\dim(\Omega') + \dim(\Omega'') = n$, and $r : \Omega' \rightarrow \mathbb{R}^m$. Denoting the first $\dim(\Omega')$ elements of a vector with the prime, we can investigate the case of an incumbent x^k with $r((x^k)') > \mathbf{0}$. As a vector $((x^k)', x'')$, for all $x'' \in \Omega''$, can

never be a valid solution for (\mathbf{P}) , the norm-induced cut in x^k can be added in $(x^k)'$ instead. That is, we can add

$$\|x' - (x^k)'\| \geq \|r((x^k)')\|/L,$$

if a suitable notion of norm is available for Ω' .

In conclusion, one can even combine several inequalities depending on different subvectors of x . The only necessary condition is the availability of respective Lipschitz constants and norms.

We showed that **NIC** works correctly when we have a global solution method for (\mathbf{P}_k) . But the question arises whether the algorithm still fulfills the desired behavior when only a local solution method is available. In the following, we verify that a returned point of step 9 is indeed a locally optimal point, if only such can be found in step 2.

Theorem 3.8. *Let (\mathbf{P}) be feasible, i.e., $Q \neq \emptyset$, and let step 2 produce a locally optimal point x^k for (\mathbf{P}_k) , $k \in \mathbb{N}_0$. Then, if **NIC** stops in step 9, it returns a locally optimal solution x^* .*

Proof. Assume that **NIC** stops in step 9 in iteration $k \in \mathbb{N}_0$. As x^k is a locally optimal solution for (\mathbf{P}_k) , there exists an $\varepsilon > 0$ such that

$$f(x^k) = \min_{x \in B_\varepsilon(x^k) \cap Q_k} f(x).$$

By Theorem 3.4 $x^k \in Q$ and, thus, $x^k \in Q \cap B_\varepsilon(x^k)$. We infer

$$\min_{x \in Q \cap B_\varepsilon(x^k)} f(x) \leq f(x^k).$$

Further, the result from Lemma 3.3 implies $Q \cap B_\varepsilon(x^k) \subseteq Q_k \cap B_\varepsilon(x^k)$. We can derive

$$f(x^k) = \min_{x \in Q_k \cap B_\varepsilon(x^k)} f(x) \leq \min_{x \in Q \cap B_\varepsilon(x^k)} f(x),$$

and, thus, conclude that

$$\min_{x \in Q \cap B_\varepsilon(x^k)} f(x) = f(x^k).$$

Therefore, x^k is a local optimum for (\mathbf{P}) and the proof is complete. \square

If **NIC** does not terminate with returning a particular point, we can not guarantee the local optimality of the accumulation points. On the contrary, the sequence might even converge to the worst possible point which we show in Example 4.7 in the following section.

4. TERMINATION, PROBLEM COMPLEXITY AND ILLUSTRATIONS

This section discusses a termination bound of **NIC** based on the properties of r in case of infeasibility in (\mathbf{P}) , the problem complexity, and illustrative examples giving rise to particular assumptions made. We start with the termination bound and operate on \mathbb{R} throughout the entire section.

4.1. Termination for Infeasible Problems. Let (\mathbf{P}) be infeasible, i.e., $Q = \emptyset$. Hence, $r(x) > \mathbf{0}$ for all $x \in \Omega$. Due to the compactness of Ω and the (Lipschitz) continuity of r , there exists $\delta > 0$ with $\delta < \min_{x \in \Omega} \|r(x)\|$. In analogy to the proof of Theorem 3.4, we can deduce inequality (4) regarding the norm-induced cuts for a sequence of incumbents $(x^i)_{i \in \mathbb{N}_0}$, i.e.,

$$\|x - x^i\| \geq \frac{\|r(x^i)\|}{L} > \frac{\delta}{L}, \quad \text{for all } i \in \mathbb{N}_0.$$

Ω	$\ \cdot\ $	T
$\Omega_1 := \prod_{j=1}^n [a_j, b_j]$	$\ \cdot\ _\infty$	$\prod_{j=1}^n [(L/\delta)(b_j - a_j) + 1]$
$\Omega_2 := \prod_{j=1}^n [a_j, b_j] \cap \mathbb{Z}^n$	$\ \cdot\ _\infty$	$ \Omega_2 $
$\Omega_3 := \{x \mid \ x\ _2 < D\}$	$\ \cdot\ _2$	$(2LD/\delta + 1)^n$

TABLE 1. Examples for T in (7) for $\delta/L < 1$ depending on the structure of Ω and the choice of norm.

That is, each incumbent x^i satisfies the upper inequality with respect to all previous incumbents x^0, \dots, x^{i-1} ; compare Remark 2.3. In particular, for all $l \in \mathbb{N}$, it is

$$\|x^l - x^i\| > \frac{\delta}{L}, \quad \text{for all } i = 0, \dots, l-1.$$

Thus, the number of iterations until termination can be bounded by the maximal number of incumbents in Ω with a norm of the pairwise difference of more than δ/L . This was formalized in Definition 2.5 and refers to a maximal δ/L -packing of Ω .

In particular, our aim comprises of computing $M(\delta/L) := M(\Omega, \|\cdot\|, \delta/L)$ or an upper bound for it. To the best of our knowledge there is no such bound for a general normed finite-dimensional vector space. Though, when restricting X to \mathbb{R}^n , the situation differs, see [8, 17]:

Remark 4.1. Let $X = \mathbb{R}^n$ and $\text{vol}(\cdot)$ denote the Lebesgue measure (“volume”). Further, we define $B := B_1$ as the unit ball under the considered norm $\|\cdot\|$ and set $B/2 := B_{1/2}$. Then, it holds

$$M(\delta/L) \leq T, \tag{7}$$

where

$$T := \frac{\text{vol}(\Omega + (\delta/L)B/2)}{\text{vol}((\delta/L)B/2)},$$

and the sum of two sets is considered the Minkowski sum.

Depending on the structure of Ω and the specific choice of norm, this leads to particular upper bounds dependent on δ and L . We give some examples in Table 1.

4.2. Problem Complexity. As we tackle an abstract problem class, the expectation of finding an exact solution in finite time is inadequate; see [14, 18]. Hence, for the investigation of problem complexity, we only consider approximate solutions.

Definition 4.2. Let $\varepsilon > 0$ and let r_i denote the i -th component of r , $i = 1, \dots, m$. Then, we call $x^* \in \Omega$ an ε -approximate solution for (P) if and only if

$$\forall i = 1, \dots, m : r_i(x^*) \leq \varepsilon.$$

As shown in [18], continuity of f is not sufficient to find an approximate solution in a finite amount of steps. Thus, we assume the objective f to be also Lipschitz. Additionally, in order to enable a discussion about the problem complexity, the interior of Ω is assumed to be non-empty, as well an optimal solution x^* to (P) is assumed to be contained in an ε -ball inside Ω . This ensures that if NIC adds norm-induced cuts with a radius of ε , it is not possible to cut off every neighboring point in Ω .

For the lower bound on the problem complexity, we use a result from [14]. In order to state it clearly, we define the following constants regarding Ω .

Definition 4.3 ([14]). Let $\varepsilon > 0$ and let Ω have a non-empty interior. Then, we set the radius of Ω as

$$\rho(\Omega) := \min\{s \mid \exists x \in \Omega : \|x - y\| \leq s, \text{ for all } y \in \Omega\}.$$

Further, the asphericity of Ω is defined as

$$\alpha(\Omega) = \inf\{\beta \mid \exists x, y \in \Omega, s \geq 0 : B_s(x) \subseteq \Omega \subseteq B_{\beta s}(y)\}.$$

In other terms, the asphericity defines the ratio of radii of the minimal ball containing Ω and the maximal ball contained in Ω .

Since the magnitude of each r_i has a great influence on how to choose a proper approximation guarantee ε , we use the radius $\rho(\Omega)$ and assume r to be divided by $\rho(\Omega)L$, “normalizing” the inequality.

Now, problem complexity considers the number of oracle calls (step 2) until achieving an approximate solution. We denote this number with $N(\varepsilon)$ for $\varepsilon > 0$. As **NIC** is presented to aim for an exact solution, we adjust it in the following way.

- a) We only check for component-wise inequality with respect to ε in the if-clause before step 9,
- b) and we add the norm-induced cut from Definition 2.1 with respect to a slightly changed radius of $\max\{\|r(x^k)\|/L, \varepsilon\}$.

Then, we can give the following lower bound on the problem complexity.

Theorem 4.4 ([14]). *Let $\varepsilon > 0$ and let Ω have a non-empty interior. Then,*

$$N(\varepsilon) \geq \left(\frac{c}{\alpha(\Omega)}\right)^n \frac{1}{\varepsilon^n},$$

with some constant $c > 0$.

For the upper bound, we can use Remark 4.1 with ε instead of δ/L . That is, instead of giving a termination bound dependent on values of r , we do it in terms of the approximation guarantee ε . We formalize this in the following.

Theorem 4.5. *Let $\varepsilon > 0$ and let Ω have a non-empty interior. Then,*

$$N(\varepsilon) \leq (2\rho(\Omega) + \varepsilon)^n \frac{1}{\varepsilon^n}.$$

Proof. With the definition of the radius in Definition 4.3, there exists $\bar{x} \in \Omega$ such that $\Omega \subseteq B_{\rho(\Omega)}(\bar{x})$. Together with Remark 4.1, we can conclude

$$\begin{aligned} N(\varepsilon) &\leq \frac{\text{vol}(\Omega + \varepsilon B/2)}{\text{vol}(\varepsilon B/2)} \leq \frac{\text{vol}(B_{\rho(\Omega)} + \varepsilon B/2)}{\text{vol}(\varepsilon B/2)} \\ &= \frac{\text{vol}(B_{\rho(\Omega)+\varepsilon/2})}{\text{vol}(\varepsilon B/2)} = \left(\frac{\rho(\Omega) + \varepsilon/2}{\varepsilon/2}\right)^n \\ &= \left(\frac{2\rho(\Omega)}{\varepsilon} + 1\right)^n = (2\rho(\Omega) + \varepsilon)^n \frac{1}{\varepsilon^n}. \end{aligned}$$

□

In conclusion, considering the dimension n fixed, we observe that the lower and upper complexity bound show identical order of magnitude in ε .

4.3. Examples. Having investigated the problem complexity, we lastly aim to illustrate the behavior of **NIC**. We give an example of a two dimensional problem, involving non-linear constraints in form of a sine function. We note that such a problem could not be treated by standard solvers such as [7] without reformulation.

Example 4.6. *Consider the non-linear problem*

$$\begin{aligned} \min_{x_1, x_2} \quad & |x_1 - x_2| + x_1 \\ \text{s.t.} \quad & -\sin(x_1) - x_2 \leq 0, \\ & x_1, x_2 \in [-1, 1]. \end{aligned} \tag{P_{sin}}$$

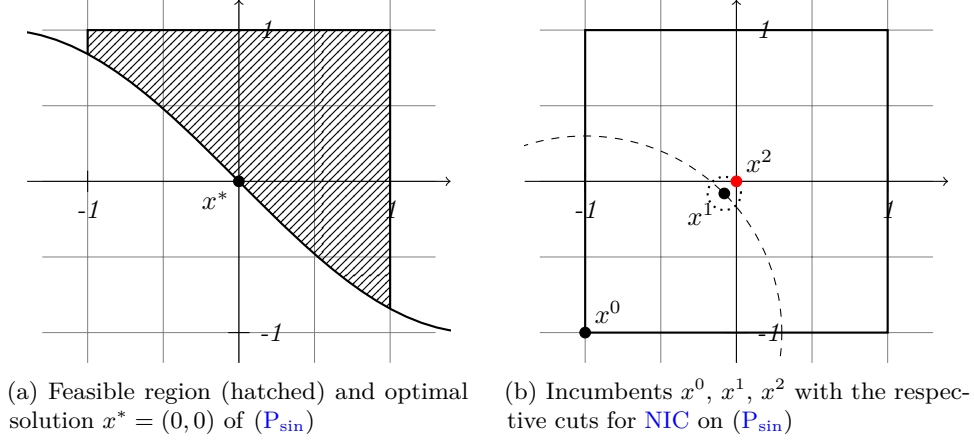


FIGURE 1. Illustration of Example 4.6

Following our notation, we have $f(x) = f(x_1, x_2) = |x_1 - x_2| + x_1$, $r(x) = -\sin(x_1) - x_2$, and $\Omega = [-1, 1]^2$. The optimal solution of Problem (P_{\sin}) is $x^* = (0, 0)$. See Figure 1a for a graphical representation of x^* and the feasible region $Q = \Omega \cap \{x \in \mathbb{R}^2 \mid r(x) \leq 0\}$.

We specify an accuracy of $\varepsilon = 10^{-4}$. Further, we consider the \mathbb{R}^2 with the Euclidean norm $\|\cdot\|_2 = \sqrt{x_1^2 + x_2^2}$ and the \mathbb{R} with the absolute value $|\cdot|$.

As r is differentiable, we can apply the mean value theorem and calculate its global Lipschitz constant L on Ω as the maximal norm of the gradient of r over Ω . In a formal manner, this reads

$$L = \max_{x \in \Omega} \|\nabla r(x)\|_2 = \max_{x \in \Omega} \left\| \begin{pmatrix} -\cos(x_1) \\ -1 \end{pmatrix} \right\|_2 = \sqrt{2}.$$

In the first iteration, the algorithm initializes $k \leftarrow 0$ and afterwards solves $\min \{f(x) \mid x \in Q_0 = \Omega\}$. As the problem is feasible, we receive a solution $x^0 = (x_1^0, x_2^0) = (-1, -1)$. Evaluating r at x^0 gives $r(x^0) \approx 1.84$, which does not satisfy the required accuracy. Therefore, NIC adds a cut of the form

$$\|x - x^0\|_2 \geq \frac{|r(x^0)|}{L}, \quad (8)$$

with the right-hand side being approximately 1.30. We illustrate this cut with a dashed circle and its center, the current incumbent x^0 , with a black dot in the lower left corner in Figure 1b.

Turning to the second iteration, NIC minimizes f over Ω associated with the cut (8). This results in the incumbent $x^1 = (x_1^1, x_2^1)$ with $x_1^1 = x_2^1 \approx -0.08$. Again, computing the value for r , we receive $r(x^1) \approx 0.16$ which does not satisfy the accuracy. Hence, the second cut

$$\|x - x^1\|_2 \geq \frac{|r(x^1)|}{L}, \quad (9)$$

with a right-hand side of approximately 0.11, is added. This is illustrated with the dotted circle and the respective incumbent in Figure 1b.

Lastly, minimizing f over Ω with the cuts (8) and (9), we receive $x^2 = (x_1^2, x_2^2)$ with $x_1^2 = x_2^2 \approx 4 \cdot 10^{-5}$. This gives $r(x^2) \approx 8 \cdot 10^{-5}$ which satisfies the accuracy and leads to termination of the algorithm. Therefore, we stop with an approximate solution x^2 which is close to the original solution x^* . We visualized the found point in a different color in Figure 1b.

Note that we globally solved the subproblems tackled in the upper example. In Theorem 3.8 we have shown that a point returned in step 9 under a local method for step 2 is a locally optimal point for (P). In the following, we will give an example that does not terminate and, thus, creates a sequence which even converges to the global maximum instead of the desired minimum.

Example 4.7. *Here, we consider a one dimensional problem in \mathbb{R} with the absolute value $|\cdot|$ which reads*

$$\begin{aligned} \min \quad & -|x| \\ \text{s.t.} \quad & \frac{1}{3}x^3 \geq 0, \\ & x \in [-1, 1]. \end{aligned} \tag{P_{\text{bad}}}$$

Again, following our notation, we have $f(x) = -|x|$, $r(x) = -\frac{1}{3}x^3$, and $\Omega = [-1, 1]$. We note that the fraction in the inequality regarding r could be omitted. But, when we calculate the Lipschitz constant of r on Ω , we can use the differentiability of r and the fraction to get $L = \max_{x \in \Omega} |r'(x)| = \max_{x \in [-1, 1]} |x^2| = 1$.

In order to understand Problem (P_{bad}), we can re-write the problem as $\max \{x \mid x \in [0, 1]\}$ and see that its optimal point is at $x^* = 1$. In contrast, the “worst” possible point in terms of the optimization problem would be $\hat{x} = 0$, as this is the maximum of Problem (P_{bad}) or the minimum of the re-written problem. We will show that NIC converges to \hat{x} if the solution method used in step 2 is a local one only.

We assume step 2 to always return a locally optimal solution in $[-1, 0)$ if it exists. Therefore, relaxing the inequality $r(x) \leq 0$ in (P_{bad}), we receive the incumbent $x^0 = -1$ with a violation of $r(x^0) = 1/3$. Adding the cut $|x - x^0| \geq r(x^0)/L = 1/3$, we end up with $Q_1 = [-2/3, 1]$ as the feasible set of the next problem to investigate. We can show by induction that our assumption about the solution method leads to the relation

$$x^{k+1} = x^k + r(x^k) = x^k - \frac{1}{3}(x^k)^3, \tag{10}$$

for $k \in \mathbb{N}_0$ and our incumbents x^k . Further, using induction for one more time, we can prove that $x^k \in [-1, 0)$ for all $k \in \mathbb{N}_0$. So, our x^k 's are actually the points returned in step 2.

Now, we define a function $\nu(x) := x - \frac{1}{3}x^3$ as the right-hand side of (10). Note that this is a contraction on $[-1, 0]$. Therefore, we can use Banach's fixed-point theorem to see that ν has a unique fixed-point \hat{x} on $[-1, 0]$. The use of the notation from above is intended as the fixed-point is indeed $\hat{x} = 0$, which can be seen by $\nu(0) = 0$. Further, Banach's theorem gives that by starting with any $x^0 \in [-1, 0]$, the sequence defined by $x^{k+1} = \nu(x^k)$, which is equivalent to (10), converges to \hat{x} . Therefore, $\lim_{k \rightarrow \infty} x^k = \hat{x} = 0$.

In summary, this shows that the sequence produced by NIC converges to the “worst” point. If we have a solution method returning a global optimal point, then the algorithm would return $x^* = 1$ in its second iteration at the latest.

5. CONCLUSION

We have introduced the NIC method in order to tackle the abstract problem class (P). The method only requires a continuous objective, a compact domain, and Lipschitz continuous constraint functions. Leveraging an oracle for the global solution of (P_k), we have proven the correctness of NIC in terms of returning an optimal solution when terminating or, if not, creating a sequence whose accumulation points are all optimal. This has been complemented by discussions about the termination bound in the infeasible case and the problem complexity, showing asymptotic

convergence of the bounds in the latter. The theoretical content has been rounded off with two examples, illustrating the method and visualizing the possible effects of a local oracle.

As mentioned in a side note above, there exist optimization software like Gurobi [7] which successfully tackle quadratic constraint problems, but struggle when facing general non-linear constraints. Our approach enables such software to solve even more general problems by executing NIC using them as the oracle in step 2. Contrasting the theoretical character of the present article, an application to wide problem classes accompanied by an extensive computational study can build one further research direction.

Since the magnitude of the radius of the norm-induced cuts depend on the Lipschitz constant L in the denominator, an application of NIC to constraints with small L seems reasonable. For example, the field of optimal control of power networks include sine or cosine constraints with small L and, thus, could be tackled by NIC. This gives rise to another interesting topic of research.

ACKNOWLEDGEMENTS

We thank the Deutsche Forschungsgemeinschaft for their support within project A05 in the *Sonderforschungsbereich/Transregio 154 Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks*.

REFERENCES

- [1] C. Audet and J. E. Dennis. “A Pattern Search Filter Method for Nonlinear Programming without Derivatives.” In: *SIAM Journal on Optimization* 14.4 (2004), pp. 980–1010. DOI: [10.1137/S105262340138983X](https://doi.org/10.1137/S105262340138983X).
- [2] C. Audet and J. E. Dennis. “Mesh Adaptive Direct Search Algorithms for Constrained Optimization.” In: *SIAM Journal on Optimization* 17.1 (2006), pp. 188–217. DOI: [10.1137/040603371](https://doi.org/10.1137/040603371).
- [3] F. V. Berghen. “CONDOR, a parallel, direct, constrained optimizer for high-computing-load, black box objective functions.” In: *Proceedings of the third MIT conference on Computational Fluid and Solid Mechanics, Elsevier*. 2005.
- [4] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. Philadelphia: SIAM, 2009. DOI: [10.1137/1.9780898718768](https://doi.org/10.1137/1.9780898718768). eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9780898718768>.
- [5] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. Pappas. “Efficient and Accurate Estimation of Lipschitz Constants for Deep Neural Networks.” In: *Advances in Neural Information Processing Systems*. Vol. 32. Vancouver: Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/95e1533eb1b20a9777749fb94fdb944-Paper.pdf.
- [6] J. Grübel, R. Krug, M. Schmidt, and W. Wollner. “A Successive Linear Relaxation Method for MINLPs with Multivariate Lipschitz Continuous Non-linearities.” In: (2023). arXiv: [2208.06444](https://arxiv.org/abs/2208.06444) [math.OA].
- [7] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2023. URL: <https://www.gurobi.com>.
- [8] Y. Han. *Lecture notes in Informatic-theoretic Lower Bounds in Data Science (Lecture 12)*. May 2021. URL: https://web.stanford.edu/class/ee378c/scribe_lec12.pdf.
- [9] R. Horst and H. Tuy. *Global optimization: Deterministic approaches*. Third. Heidelberg: Springer Berlin, 1996. DOI: [10.1007/978-3-662-03199-5](https://doi.org/10.1007/978-3-662-03199-5).
- [10] T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. *Evaluating gas network capacities*. Philadelphia: SIAM, 2015. DOI: [10.1137/1.9781611973693](https://doi.org/10.1137/1.9781611973693).

- [11] A. H. Land and A. G. Doig. “An Automatic Method of Solving Discrete Programming Problems.” In: *Econometrica* 28.3 (1960), pp. 497–520. URL: <http://www.jstor.org/stable/1910129>.
- [12] R. M. Lewis and V. Torczon. “Pattern Search Algorithms for Bound Constrained Minimization.” In: *SIAM Journal on Optimization* 9.4 (1999), pp. 1082–1099. DOI: [10.1137/S1052623496300507](https://doi.org/10.1137/S1052623496300507).
- [13] I. Molan and M. Schmidt. “Using neural networks to solve linear bilevel problems with unknown lower level.” In: *Optimization letters* (2023). DOI: [10.1007/s11590-022-01958-7](https://doi.org/10.1007/s11590-022-01958-7).
- [14] A. S. Nemirovskij and D. B. Yudin. *Problem complexity and method efficiency in optimization*. Norwich: Wiley-Interscience, 1983.
- [15] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations.” In: *Journal of Computational Physics* 378 (2019), pp. 686–707. DOI: [10.1016/j.jcp.2018.10.045](https://doi.org/10.1016/j.jcp.2018.10.045).
- [16] M. Schmidt, M. Sirvent, and W. Wollner. “A decomposition method for MINLPs with Lipschitz continuous nonlinearities.” In: *Mathematical Programming* 178.1 (2019), pp. 449–483. DOI: [10.1007/s10107-018-1309-x](https://doi.org/10.1007/s10107-018-1309-x).
- [17] V. M. Tikhomirov. “ ε -Entropy and ε -Capacity of Sets in Functional Spaces.” In: *Selected Works of A. N. Kolmogorov: Volume III: Information Theory and the Theory of Algorithms*. Ed. by A. N. Shirayev. Dordrecht: Springer, 1993, pp. 86–170. DOI: [10.1007/978-94-017-2973-4_7](https://doi.org/10.1007/978-94-017-2973-4_7).
- [18] S. A. Vavasis. *Nonlinear Optimization: Complexity Issues*. New York: Oxford University Press, 1991.

APPENDIX A. EXAMPLES FOR FORMULA (7)

Let $L > 0$ be the (upper bound to the) Lipschitz constant and $\delta > 0$, such that $\delta/L < 1$. We calculate the examples as appearing in table Table 1. For the sake of clarity, we define the unit ball $B = B_1$ with respect to the considered norm.

- a) Let $\Omega_1 = \prod_{j=1}^n [a_j, b_j]$ be an n -dimensional box with $a_j, b_j \in \mathbb{R}$ and $a_j < b_j$ for $j = 1, \dots, n$. We consider the maximum-norm, i.e., $\|x\|_\infty = \max_{i \in [n]} |x_i|$ for $x \in \mathbb{R}^n$ and $[n] := \{1, \dots, n\}$.

In this case, we have $B = [-1, 1]^n$ and, thus, $\text{vol}(B) = 2^n$. It follows, $\text{vol}((\delta/L)B/2) = (\delta/L)^n$ and

$$\Omega_1 + (\delta/L)B/2 = \prod_{j=1}^n [a_j - \frac{\delta}{2L}, b_j + \frac{\delta}{2L}].$$

Therefore, $\text{vol}(\Omega_1 + (\delta/L)B/2) = \prod_{j=1}^n (b_j - a_j + \delta/L)$ and we conclude

$$T = \frac{\text{vol}(\Omega_1 + (\delta/L)B/2)}{\text{vol}((\delta/L)B/2)} = \prod_{j=1}^n \left[\frac{L}{\delta} (b_j - a_j) + 1 \right].$$

- b) Let us consider the maximum-norm again and $\Omega_2 = \Omega_1 \cap \mathbb{Z}^n$ to be the integral (lattice) points in Ω_1 . Then, we can calculate the amount of lattice points as

$$|\Omega_2| = \prod_{j=1}^n ([b_j] - [a_j] + 1).$$

Since we assumed $\delta/L < 1$ in the beginning, it holds true that

$$(\{z_1\} + (\delta/L)B/2) \cap (\{z_2\} + (\delta/L)B/2) = \emptyset,$$

for $z_1, z_2 \in \Omega_2$, $z_1 \neq z_2$. Therefore, we can derive for the volume $\text{vol}(\Omega_2 + (\delta/L)B/2) = |\Omega_2| \text{vol}((\delta/L)B/2)$ and, thus,

$$T = \frac{\text{vol}(\Omega_2 + (\delta/L)B/2)}{\text{vol}((\delta/L)B/2)} = \frac{|\Omega_2| \text{vol}((\delta/L)B/2)}{\text{vol}((\delta/L)B/2)} = |\Omega_2|.$$

- c) Here, we consider the Euclidean or 2-norm, i.e., $\|x\|_2 = \sqrt{\left(\sum_{j=1}^n x_j^2\right)}$. For some radius $r > 0$, we have B_r and can derive from standard literature that

$$\text{vol}(B_r) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n = \text{vol}(B) r^n,$$

where $\Gamma(\cdot)$ denotes the gamma function. Now, for $D > 0$, let $\Omega_3 = B_D$ and it follows that

$$\text{vol}(\Omega_3 + (\delta/L)B/2) = \text{vol}(B_{D+(\delta/(2L))}) = \text{vol}(B)(D + (\delta/(2L)))^n.$$

Therefore, we can conclude

$$T = \frac{\text{vol}(B_D + (\delta/L)B/2)}{\text{vol}((\delta/L)B/2)} = \frac{(D + (\delta/(2L)))^n}{(\delta/(2L))^n} = \left(\frac{2L}{\delta} D + 1\right)^n.$$

(A. Göß) FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG, ANALYTICS & MIXED-INTEGERS OPTIMIZATION, CAUERSTR. 11, 91058 ERLANGEN, GERMANY

Email address: adrian.goess@fau.de

(A. Martin) UNIVERSITY OF TECHNOLOGY NUREMBERG, RESEARCH, INNOVATION AND ENTREPRENEURSHIP, ULMENSTR. 52I, 90443 NÜRNBERG, GERMANY

Email address: alexander.martin@utn.de

(S. Pokutta) ZUSE INSTITUTE BERLIN, MATHEMATICAL ALGORITHMIC INTELLIGENCE, TAKUSTR. 7, 14195 BERLIN, GERMANY

Email address: pokutta@zib.de

(K. Sharma) ZUSE INSTITUTE BERLIN, MATHEMATICAL ALGORITHMIC INTELLIGENCE, TAKUSTR. 7,
14195 BERLIN, GERMANY
Email address: `kartikey.sharma@zib.de`