ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Liuzzi, S. Lucidi

A DERIVATIVE-FREE ALGORITHM FOR INEQUALITY CONSTRAINED NONLINEAR PROGRAMMING

R. 659 Aprile 2007

Giampaolo Liuzzi – CNR - Consiglio Nazionale delle Ricerche, IASI - Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Viale Manzoni 30 - 00185 Roma - Italy (liuzzi@iasi.cnr.it)..

Stefano Lucidi – Università degli Studi di Roma "La Sapienza", Dipartimento di Informatica e Sistemistica "A. Ruberti", Via Ariosto 25 - 00185 Roma - Italy (lucidi@dis.uniroma1.it)..

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.rm.cnr.it
URL: http://www.iasi.rm.cnr.it

Abstract

In this paper we consider inequality constrained nonlinear optimization problems where the first order derivatives of the objective function and the constraints cannot be used. Our starting point is the possibility to transform the original constrained problem into an unconstrained or linearly constrained minimization of a non-smooth exact penalty function. This approach shows two main difficulties: the first one is the non-smoothness of this class of exact penalty functions which may cause derivative-free codes to converge to non-stationary points of the problem; the second one is the fact that the equivalence between stationary points of the constrained problem and those of the exact penalty function can be stated only when the penalty parameter is smaller than a threshold value which is not known *a priori*. In this paper we propose a derivative-free algorithm which overcomes the preceding difficulties and produces a sequence of points that admits a subsequence converging to a Karush-Kuhn-Tucker point of the constrained problem. In particular the proposed algorithm is based on a smoothing of the non-differentiable exact penalty function and includes an updating rule which, after at most a finite number of updates, is able to determine a "right value" of the penalty parameter. Furthermore we present the results obtained on a real world problem concerning the estimation of parameters in an insulin-glucose model of the human body.

Key words: Derivative-free optimization, constrained optimization, nonlinear programming, non-differentiable exact penalty functions

AMS subject classifications: 65K05, 90C30, 90C56

1. Introduction

We consider the following problem

$$\begin{array}{ll} \min & f(x), \\ s.t. & g(x) \leq 0 \\ & Ax \leq b, \end{array} \qquad (1)$$

where $x \in \Re^n$, $f : R^n \to R$, $g : R^n \to R^m$, $A \in R^{p\times n}$, $b \in R^p$ and we assume that f and g are twice continuously differentiable on R^n. We denote by $a_j^\top$, $j = 1, \dots, p$, the rows of matrix A and by

$$\mathcal{F} = \{x \in \Re^n : \; Ax \leq b, \; g(x) \leq 0\}$$

the feasible set of Problem (1). We assume that the derivatives of the objective and nonlinear constraint functions can be neither calculated nor approximated explicitly. Indeed, in many engineering problems the analytic expressions of the functions defining the objective and constraints of the problem are not available and their values are computed by means of complex simulation computer programs. For further motivations on the necessity of using derivative-free methods we refer the reader to the survey paper [21].

In the literature, some globally convergent derivative-free methods for the solution of Problem (1) have been proposed. In [26] a pattern search algorithm is used within a sequential augmented Lagrangian approach. Essentially, the method embeds the pattern search algorithm proposed in [25] within the augmented Lagrangian method [5] which is the basis for the subroutine AUGLG in the LANCELOT optimization package.

In [1] the filter method proposed in [9] is adapted to include a patter search minimization strategy. Basically, the method employs a "filter" for acceptance of the points produced by the patter search local optimizer.

In [2] a so-called *extreme barrier* approach, namely the constrained problem is converted to an unconstrained one by setting the objective to infinity for infeasible point, is employed. To minimize this extreme barrier function, the authors propose an extension of the generalized pattern search class of algorithms which allows local exploration in an asymptotically dense set of directions.

Similarly to [26] and [2], in this paper we propose an algorithm which is based on the idea of employing a derivative-free method to an unconstrained reformulation of Problem (1). Our approach differs from the precedeeing ones in that we start from the possibility to solve Problem (1) by an unconstrained minimization of a non-smooth exact penalty function. However non-smooth exact penalty functions cannot be straightforwardly combined to a globally convergent derivative-free algorithm. Indeed, the following theoretical and computational aspects should be carefully taken into account.

- Ill-conditioning of merit functions. This aspect makes the minimization of such functions a difficult task especially for derivative-free codes which use only evaluations of the objective function.

- Non-differentiability of the penalty function. The lack of non-differentiability may have negative effects on the convergence of derivative-free methods to stationary points of the penalty function. Indeed, most of the unconstrained derivative-free methods require the objective function to be at least continuously differentiable.

- Equivalence between stationary points of the penalty function and KKT pairs of Problem (1). An exact penalty function enjoys its exactness properties only if the penalty parameter is below a certain threshold value which is not known “a priori”. This aspect is crucial also in the case of available derivatives.

As regards the first point, we introduce a new exact penalty function which only penalizes the nonlinear constraints and does this penalization in such a way to reduce as much as possible the ill-conditioning.
The non-differentiability of the new penalty function has been tackled by employing the smoothing technique proposed in [4, 30]. In particular, in order to find a stationary point of the penalty function by minimizing the smooth approximation, we adapted the method proposed in [22] to solve linearly constrained finite minimax problems.
As for the last point, the properties of the smooth approximation allow us to define a suitable updating rule for the penalty parameter ϵ. This rule, after a finite number of reductions, is able to find a right value for ϵ so as to convey the desirable exactness properties to the penalty function.
To conclude we propose a globally convergent algorithm which is based on the derivative-free minimization of a smooth approximation of a non-differentiable exact penalty function which does an ℓ_∞ penalization of the constraints. Moreover, this new algorithms exploits the structure of the problem by allowing for an explicit handling of the linear constraints.
As regards a possible practical interest of the proposed approach, we recall the encouraging results described in [11]. In fact, [11] reports an extensive numerical testing and comparison which point out significant computational advantages in using a smooth approximation of an exact ℓ_∞ penalty function in the field of derivative-free methods.
Even though the main aim of the paper is the definition of a new algorithm and the study of its theoretical properties, we also show that a rough implementation of the method is able to successfully solve a real world problem concerning the estimation of parameters in an insulin-glucose model of the human body. This result seems to further confirm the conclusion of [11].
The paper is organized as follows. In Section 2, the exact penalty function approach is introduced and discussed. Section 3 is devoted to the description of a smooth approximation technique along with some preliminary properties. In Section 4, the derivative-free method is reported and its global convergence is studied. Section 5 is devoted to the solution of the constrained parameter estimation problem. Finally, in Section 6, we draw some conclusions.

2. An exact penalty function approach

As already said in the introduction, the first step of our approach is that of defining and using a penalty function which be more tractable from a computational point of view. Namely, a penalty function which

(i) has a structure that presents fewer nonlinearities than previous exact penalty functions [8];

(ii) allows direct handling of linear and bound constraints thus penalizing only the nonlinear ones.

As regards point (i), non-differentiable exact penalty functions were introduced for the first time in [31] for solving nonlinear programming problems of the form

$$\begin{array}{ll} \min & f(x) \\ s.t. & g(x) \leq 0. \end{array} \tag{2}$$

In [31] the penalty function had the expression

$$P(x;\epsilon) = f(x) + \frac{1}{\epsilon}\max\{0, g_1(x), \ldots, g_m(x)\}.$$

However, the function $P(x;\epsilon)$ is not "globally exact" (as defined in [8]), namely, the solution of problem (2) is not completely equivalent to an unconstrained minimization of $P(x;\epsilon)$.
In order to come up with a globally exact penalty function, it is necessary to introduce a compact relaxation of the feasible set $\mathcal{F}$.
Following references [8], given a vector $\alpha \in \Re^m$ such that $\alpha_i > 0$, $i = 1, \ldots, m$, the set

$$\mathcal{D}_\alpha = \{x \in R^n : g(x) \leq \alpha\},$$

is considered and assumed to be compact. Then, on the interior of set $\mathcal{D}_\alpha$, the penalty function

$$Q(x;\epsilon) = f(x) + \frac{1}{\epsilon}\max\left\{0, \frac{g_1(x)}{\alpha_1 - g_1(x)}, \ldots, \frac{g_m(x)}{\alpha_m - g_m(x)}\right\}$$

can be defined, where the terms $\alpha_i - g_i(x)$ makes $Q(x;\epsilon)$ go to infinity when x approaches the boundary of $\mathcal{D}_\alpha$, thus guaranteeing the compactness of its level sets.
In [8], it is shown that $Q(x;\epsilon)$ is a globally exact penalty function for problem (2), that is, a value $\epsilon^\star > 0$ for the penalty parameter exists such that for every $\epsilon \in (0, \epsilon^\star]$ the solution of problem (2) is equivalent to the solution of

$$\min Q(x;\epsilon) \quad s.t. \quad x \in \mathring{\mathcal{D}}_\alpha, \tag{3}$$

which, essentialy, amounts to an unconstrained minimization of $Q(x;\epsilon)$, due to the fact that $\mathring{\mathcal{D}}_\alpha$ is an open set. More in particular, a value $\epsilon^\star > 0$ for the penalty parameter exists such that for every $\epsilon \in (0, \epsilon^\star]$, every local (global, stationary) point of Problem (3) is a local (global, KKT) point of Problem (2) and conversely.
However, we note that the structure of $Q(x;\epsilon)$ is such that two contrasting effects, tied with the choice of parameters α and ϵ, may arise. Indeed, rewriting

$$Q(x;\epsilon) = f(x) + \max\left\{0, \frac{g_1(x)}{\epsilon(\alpha_1 - g_1(x))}, \ldots, \frac{g_m(x)}{\epsilon(\alpha_m - g_m(x))}\right\},$$

two conflicting requirements become apparent. On the one hand, in order to limit the ill-conditioning of the penalty function near the boundary of the compact set $\mathcal{D}_\alpha$, sufficiently large α_i's should be chosen. On the other hand, the exactness properties only follow when the constraints are sufficiently penalized, that is when terms $\epsilon(\alpha_i - g_i(x))$ are sufficiently small on all $\mathcal{D}_\alpha$. Hence, choosing large α_i's requires very small values of ϵ thus increasing the ill-conditioning of the penalty function wherever at least one term $(\alpha_i - g_i(x))$ exists which is

not excessively large. Besides, reasonable values for ϵ preclude the possibility of choosing large values for the α_i's that is the possibility of choosing sets $\mathcal{D}_\alpha$ having the boundary sufficiently away from the feasible region.
In order to overcome the preceding difficulties, we introduce the following new penalty function

$$Z(x;\epsilon) = f(x) + \max\left\{0, \left(\frac{1}{\epsilon} + \frac{1}{\alpha_1 - g_1(x)}\right) g_1(x), \dots, \left(\frac{1}{\epsilon} + \frac{1}{\alpha_m - g_m(x)}\right) g_m(x)\right\}$$

where the terms $g_i(x)/\epsilon$ (needed to achieve the exactness properties) and $g_i(x)/(\alpha_i - g_i(x))$ (required to guarantee the compactness of the level sets), are split in such a way that they no longer interfere one another. Introducing the functions

$$\hat{g}_i(x;\epsilon) = \left(1 + \frac{\epsilon}{\alpha_i - g_i(x)}\right) g_i(x), \qquad i = 1, \dots, m$$

and $\hat{g}_0(x;\epsilon) = 0$, $Z(x;\epsilon)$ can be rewritten in compact form as

$$Z(x;\epsilon) = f(x) + \frac{1}{\epsilon} \max_{i=0,1,\dots,m} \{\hat{g}_i(x;\epsilon)\}.$$

As regards point (ii), the need for an "ad-hoc" handling of the constraints arises every time they can be partitioned into two subsets of "difficult" and "easy" constraints. A more traditional approach to Problem (1) would be that of penalizing all the constraints by adding to the objective function a penalty term for every constraint. However, every penalty term increases the nonlinearities and the ill-conditioning of the penalty function. This, in turn, would surely result in a difficult problem to solve especially for a derivative-free method. On the other hand, many efficient derivative-free methods exists which are able to solve linearly and bound constrained optimization problems by explicitly handling the linear and bound constraints [25, 24, 29]. For this reason, instead of Problem (2), we consider problem (1) and penalize just the nonlinear constraints.
To this aim, we prove that a threshold value $\epsilon^\star > 0$ exists such that, for all $\epsilon \in (0, \epsilon^\star]$, Problem (1) is equivalent to

$$\begin{array}{ll} \min & Z(x;\epsilon) \\ s.t. & Ax \le b \\ & x \in \overset{\circ}{\mathcal{D}}_\alpha \; . \end{array} \tag{4}$$

Note that, the new structure of the penalty function along with the presence of the linear constraints and, hence, the fact that $Z(x;\epsilon)$ is to be minimized on the set $\{x \in \Re^n : \; Ax \le b, \; x \in \overset{\circ}{\mathcal{D}}_\alpha\}$ makes the analysis of [7, 8] inapplicable. Therefore, in the following subsections, we analyze the theoretical properties and connections between problems (4) and (1), by adapting the analysis carried out in [7, 8].
In the sequel we denote

$$\mathcal{S}_\alpha = \overset{\circ}{\mathcal{D}}_\alpha \cap \{x \in \Re^n : \; Ax \le b\}.$$

2.1. Definitions and assumptions

In order to state the equivalence between Problems (1) and (4), we introduce the following assumptions which we require to hold true throughout the paper. They are standard assumptions in a constrained context.

Assumption 1. *The set $\bar{\mathcal{S}}_\alpha$ is compact*

Assumption 2. *At any point $x \in \bar{\mathcal{S}}_\alpha$, a vector $d \in T(x)$ exists such that*

$$\nabla g_i(x)^\top d < 0, \qquad \textit{for all } i \textit{ s.t. } g_i(x) \geq 0.$$

These two assumptions are needed basically to guarantee, the first one, that the penalty function has compact level sets and, the second one, that the feasible region of Problem (1) is not empty with a non-empty interior.
As concerns Problem (1), the presence of the linear inequality constraints allows us to define necessary optimality conditions under somewhat weaker assumptions than usual. In particular, under Assumption 2 it is possible to state the following (KKT) necessary conditions for local optimality of Problem (1).

Proposition 2.1. *Let $\bar{x} \in \mathcal{F}$ be a local solution of Problem (1). Then,*

$$\begin{aligned} &\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + A^\top \bar{\mu} = 0 \\ &\bar{\lambda}^\top g(\bar{x}) = 0, \quad \bar{\lambda} \geq 0, \\ &\bar{\mu}^\top (Ax - b) = 0, \quad \bar{\mu} \geq 0, \end{aligned} \tag{5}$$

for some vectors $\bar{\lambda} \in R^m$ and $\bar{\mu} \in \Re^p$.

Proof. The proof follows by considering Propositions 3.3.11 and 3.3.12 in [3] along with the Motzkin theorem of the alternative [27]. □

As regards Problem (4), we recall that the directional derivative $DZ(x, d; \epsilon)$ of $Z(x; \epsilon)$ at $x \in \mathcal{S}_\alpha$ along direction $d \in R^n$ exists and is given by (see, for instance, [4])

$$DZ(x, d; \epsilon) = \nabla f(x)^\top d + \frac{1}{\epsilon} \max_{i \in B(x;\epsilon)} \left\{ \nabla \hat{g}_i(x; \epsilon)^\top d \right\},$$

where

$$B(x; \epsilon) = \left\{ i \in \{0, 1, \ldots, m\} : \hat{g}_i(x; \epsilon) = \max_{j=0,1,\ldots,m} \{\hat{g}_j(x; \epsilon)\} \right\},$$

and

$$\nabla \hat{g}_i(x; \epsilon) = \left(1 + \frac{\epsilon \alpha_i}{(\alpha_i - g_i(x))^2}\right) \nabla g_i(x), \qquad i = 1, \ldots, m.$$

Therefore, the usual definition of stationarity for Problem (4) can be given as follows.

Definition 2.2. *A point $\bar{x} \in \mathcal{S}_\alpha$ is a stationary point of Problem (4) if*

$$DZ(\bar{x}, d; \epsilon) \geq 0, \quad \textit{for all } d \in T(\bar{x}),$$

where

$$T(x) = \{d \in \Re^n : a_j^\top d \leq 0, \textit{ for all } j \textit{ s.t. } a_j^\top x = b_j\} \tag{6}$$

is the cone of feasible directions *with respect to the linear inequality constraints.*

By exploiting the particular structure of Problem (4), that is, the expression of the penalty function $Z(x; \epsilon)$, it is possible to state a different characterization of its stationary points.

Proposition 2.3. *For any given $\epsilon > 0$, a point $\bar{x} \in \mathcal{S}_\alpha$ is a stationary point of Problem (4) if and only if λ_i, $i \in B(\bar{x}; \epsilon)$, exist such that*

$$\lambda_i \geq 0, \quad i \in B(\bar{x}; \epsilon), \qquad \sum_{i \in B(\bar{x};\epsilon)} \lambda_i = 1, \tag{7}$$

$$\left(\nabla f(\bar{x}) + \frac{1}{\epsilon} \sum_{i \in B(\bar{x};\epsilon)} \lambda_i \nabla \hat{g}_i(\bar{x}; \epsilon) \right)^\top d \geq 0, \quad \textit{for all } d \in T(\bar{x}). \tag{8}$$

Proof. If $\bar{x} \in \mathcal{S}_\alpha$ is a stationary point of Problem (4), then at least one index $\bar{\imath} \in B(\bar{x}; \epsilon)$ exists such that $\nabla \hat{g}_{\bar{\imath}}(\bar{x}; \epsilon)^\top d \geq -\nabla f(\bar{x})^\top d$, for all $d \in T(\bar{x})$. Then conditions (7) and (8) hold with $\lambda_{\bar{\imath}} = 1$ and $\lambda_i = 0$, for all $i \in B(\bar{x}; \epsilon) \setminus \{\bar{\imath}\}$.
On the contrary, let $\bar{x} \in \mathcal{S}_\alpha$ be a point which satisfies conditions (7) and (8), then we can write:

$$\begin{aligned} 0 &\leq \left(\nabla f(\bar{x}) + \frac{1}{\epsilon} \sum_{i \in B(\bar{x};\epsilon)} \lambda_i \nabla \hat{g}_i(\bar{x}; \epsilon) \right)^\top d \\ &\leq \left(\nabla f(\bar{x})^\top d + \frac{1}{\epsilon} \max_{i \in B(\bar{x};\epsilon)} \{\nabla \hat{g}_i(\bar{x}; \epsilon)^\top d\} \sum_{i \in B(\bar{x};\epsilon)} \lambda_i \right) \\ &= \left(\nabla f(\bar{x})^\top d + \frac{1}{\epsilon} \max_{i \in B(\bar{x};\epsilon)} \{\nabla \hat{g}_i(\bar{x}; \epsilon)^\top d\} \right). \end{aligned}$$

for all $d \in T(\bar{x})$, which shows that $\bar{x}$ is a stationary point of Problem (4). □

We end this subsection by introducing the following index sets which will be used in the sequel of the paper.

$$I_0(x) := \{i : g_i(x) = 0\}, \quad I_\pi(x) := \{i : g_i(x) \geq 0\}, \quad I_\nu(x) := \{i : g_i(x) < 0\},$$
$$J(x) := \{j : a_j^\top x = b_j\}.$$

2.2. Exactness properties

The Exactness properties of the penalty function $Z(x;\epsilon)$ heavily hinges on the following lemma which is a slight modification of Theorem 2.2 of reference [12].

Lemma 2.4. *Let $\hat{x} \in \{x \in \Re^n : \ Ax \leq b\}$. Then, an open neighborhood $\mathcal{B}(\hat{x};\rho)$ of $\hat{x}$ and a direction $d \in T(\hat{x})$ exist such that, for all $i \in I_\pi(\hat{x})$, we have*

$$\nabla g_i(x)^\top d \leq -1, \qquad \forall\, x \in \mathcal{B}(\hat{x};\rho) \cap \{x \in \Re^n : \ Ax \leq b\}, \tag{9}$$

$$\nabla \hat{g}_i(x;\epsilon)^\top d \leq -1, \qquad \forall\, x \in \mathcal{B}(\hat{x};\rho) \cap \mathcal{S}_\alpha, \ \forall\, \epsilon > 0. \tag{10}$$

Proof. By Assumption 2, a vector $\hat{z} \in T(\hat{x})$ exists such that $\nabla g_i(\hat{x})^\top \hat{z} < 0$, for all $i \in I_\pi(\hat{x})$. Hence, by continuity, a $\rho > 0$ exists such that

$$\nabla g_i(x)^\top \hat{z} \leq -\gamma/2,$$

for all $x \in \mathcal{B}(\hat{x};\rho) \cap \{x \in \Re^n : \ Ax \leq b\}$ and $i \in I_\pi(\hat{x})$ where

$$-\gamma = \max_{i \in I_\pi(\hat{x})} \{\nabla g_i(\hat{x})^\top \hat{z}\} < 0.$$

So that (9) follows by choosing $d = 2\hat{z}/\gamma$.

For $x \in \mathcal{B}(\hat{x};\rho) \cap \mathcal{S}_\alpha$ we can write

$$\nabla g_i(x)^\top d = \frac{(\alpha_i - g_i(x))^2}{(\alpha_i - g_i(x))^2 + \epsilon\alpha_i} \nabla \hat{g}_i(x;\epsilon)^\top d \leq -1, \qquad \forall\, i \in I_\pi(\hat{x}),$$

so that, considering

$$\frac{(\alpha_i - g_i(x))^2 + \epsilon\alpha_i}{(\alpha_i - g_i(x))^2} > 1, \qquad \forall\, i \in I_\pi(\hat{x}), \ \forall\, \epsilon > 0$$

we have

$$\nabla \hat{g}_i(x;\epsilon)^\top d \leq -1, \qquad \forall\, i \in I_\pi(\hat{x}), \ \forall\, \epsilon > 0 \tag{11}$$

which proves (10). □

The analysis of the exactness properties of $Z(x;\epsilon)$ follows the same reasonings used in references [7] and [8]. For the sake of clarity, here we only report the statement of the main results and refer the interested reader to the appendix for a thorough development and analysis of the exactness properties.
The following propositions establish a connection between stationary points of the exact penalty function $Z(x;\epsilon)$ which are feasible and KKT pairs of Problem (1).

Proposition 2.5. *Let $\bar{x} \in \mathcal{F}$. Then, for any $\epsilon > 0$, if $\bar{x}$ is a critical point of $Z(x;\epsilon)$, there exist multipliers $\bar{\lambda} \in R^m$ and $\bar{\mu} \in R^p$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple for Problem (1).*

For sufficiently small values of the penalty parameter ϵ, a one-to-one correspondence between KKT pairs of Problem (1) and critical points of the penalty function $Z(x;\epsilon)$ exists.

Proposition 2.6. *There exists an $\epsilon^\star > 0$ such that, for all $\epsilon \in (0, \epsilon^\star]$, if $\bar{x} \in \mathcal{S}_\alpha$ is a critical point of $Z(x;\epsilon)$, there exist multipliers $\bar{\lambda} \in R^m$ such that $(\bar{x}, \bar{\lambda})$ is a KKT pair of Problem (1) and conversely.*

Finally, the last proposition describes a connection between local and global minimum points of the penalty function and Problem (1).

Proposition 2.7. *There exists an $\epsilon^\star > 0$ such that, for all $\epsilon \in (0, \epsilon^\star]$:*

(a) *if $x_\epsilon \in \mathcal{S}_\alpha$ is a (strict) local unconstrained minimum point of $Z(x;\epsilon)$, then x_ϵ is a (strict) local constrained minimum point of Problem (1);*

(b) *if $x^\star \in \mathcal{S}_\alpha$ is a global unconstrained minimum point of $Z(x;\epsilon)$ on $\mathcal{S}_\alpha$, then $x^\star$ is a global solution to Problem (1) and conversely.*

3. Smooth approximation and preliminary results

In this section we concentrate on the derivative-free approach to the solution of Problem (1). As already argued, in order to solve this problem we resort to a reformulation of Problem (1) by means of the exact penalty function $Z(x;\epsilon)$, so that, on the basis of the exactness properties studied in the preceding section, we can concentrate on the solution of the linearly constrained Problem (4).

To tackle with the non-differentiability of function $Z(x;\epsilon)$, we adopt a smoothing technique ([4, 30]) which consists of solving a sequence of smooth problems approximating the non-smooth one in the limit. Let $\mu > 0$ be a smoothing parameter and define

$$Z(x;\mu,\epsilon) = f(x) + \mu \ln\left(1 + \sum_{i=1}^{m} \exp\left(\frac{\hat{g}_i(x;\epsilon)}{\mu\epsilon}\right)\right).$$

We report some properties of $Z(x;\mu,\epsilon)$ [30].

Proposition 3.1.

(i) *For any given $x \in R^n$ and $\epsilon > 0$, $Z(x;\mu,\epsilon)$ is increasing with respect to μ and*

$$Z(x;\epsilon) \le Z(x;\mu,\epsilon) \le Z(x;\epsilon) + \mu \ln m; \tag{12}$$

(ii) *$Z(x;\mu,\epsilon)$ is twice continuously differentiable for all $\mu > 0$, $\epsilon > 0$ and*

$$\nabla_x Z(x;\mu,\epsilon) = \nabla f(x) + \frac{1}{\epsilon}\sum_{i=1}^{m} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon), \tag{13}$$

$$\begin{aligned}\nabla_x^2 Z(x;\mu,\epsilon) = {} & \nabla^2 f(x) + \frac{1}{\epsilon}\sum_{i=1}^{m} \left(\lambda_i(x;\mu,\epsilon)\nabla^2 \hat{g}_i(x;\epsilon)\right) \\ & + \frac{1}{\mu\epsilon^2}\sum_{i=1}^{m}\left(\lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top\right) \\ & - \frac{1}{\mu\epsilon^2}\left(\sum_{i=1}^{m} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)\right)\left(\sum_{i=1}^{m} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)\right)^\top,\end{aligned} \tag{14}$$

where

$$\lambda_i(x;\mu,\epsilon) = \frac{\exp\left(\dfrac{\hat{g}_i(x;\epsilon)}{\mu\epsilon}\right)}{1+\displaystyle\sum_{j=1}^{m}\exp\left(\dfrac{\hat{g}_j(x;\epsilon)}{\mu\epsilon}\right)} \in (0,1), \quad i = 0,1,\ldots,m, \tag{15}$$

and $\sum_{i=0}^{m} \lambda_i(x;\mu,\epsilon) = 1.$

Thus, having introduced the smoothing function $Z(x;\mu,\epsilon)$, we consider the following smooth approximating problem,

$$\begin{array}{l} \min_x Z(x;\mu,\epsilon) \\ Ax \le b \\ x \in \mathcal{S}_\alpha, \end{array} \tag{16}$$

where the approximating parameter μ and the penalty parameter ϵ will be adaptively reduced during the optimization process.

Considered the above Problem (16), it is important to study the connections that $Z(x;\mu,\epsilon)$ has with the original constrained problem. In particular, $Z(x;\mu,\epsilon)$ should be able, when minimized, to drive the algorithm away from infeasible points with respect to the nonlinear inequality constraints. Indeed, the following proposition states an important result needed to prove convergence of the algorithm. Namely, on every $\hat{x} \in \mathcal{S}_\alpha$ a sufficiently small neighborhood of $\hat{x}$ exists such that in every infeasible point belonging to this neighborhood a direction $d \in T(\hat{x})$ exists such that the directional derivative of the smooth approximating function along direction d is negative and uniformly bounded away from zero, for ϵ sufficiently small.

Proposition 3.2. *Let $\hat{x} \in \mathcal{S}_\alpha$ and $\mu_{MAX} > 0$ any given scalar. Then, $\epsilon(\hat{x}) > 0$ and $\sigma(\hat{x}) > 0$ exist such that for all $x \in \mathcal{B}(\hat{x};\sigma(\hat{x})) \cap \mathcal{S}_\alpha$ and $g(x) \not\leq 0$, $\epsilon \in (0,\epsilon(\hat{x})]$ a direction $d \in T(\hat{x})$ exists such that:*

$$\nabla Z(x;\mu,\epsilon)^\top d \le -\frac{1}{2\epsilon(m+1)},$$

for all $\mu \in (0,\mu_{MAX}]$.

Proof. Let $\mathcal{B}(\hat{x};\rho)$ and d be the neighborhood and the direction considered in Lemma 2.4. By continuity, we can find a neighborhood $\mathcal{B}(\hat{x};\sigma(\hat{x})) \subseteq \mathcal{B}(\hat{x};\rho)$ such that for $i \notin I_\pi(\hat{x})$ and $x \in \mathcal{B}(\hat{x};\sigma(\hat{x}))$, we have $g_i(x) < 0$; it follows that $I_\pi(x) \subseteq I_\pi(\hat{x})$ and $I_\nu(\hat{x}) \subseteq I_\nu(x)$, for $x \in \mathcal{B}(\hat{x};\sigma(\hat{x}))$.

Let now $x \in \mathcal{B}(\hat{x};\sigma(\hat{x})) \cap \mathcal{S}_\alpha$ be an infeasible point with respect to the nonlinear inequality constraints; then, there must exists at least an index $i \in I_\pi(x)$ such that $g_i(x) > 0$ which implies $I_\pi(x) \neq \phi$. By recalling expression (13), we can write

$$\begin{array}{rcl} \nabla Z(x;\mu,\epsilon)^\top d &=& \nabla f(x)^\top d \\ &+& \dfrac{1}{\epsilon}\left(\displaystyle\sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top d + \sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top d\right). \end{array}$$

By Lemma 2.4 we have that $\nabla \hat{g}_i(x;\epsilon)^\top d \le -1$, $i \in I_\pi(\hat{x})$, so that we can write

$$\begin{aligned}\nabla Z(x;\mu,\epsilon)^\top d &\le \nabla f(x)^\top d \\ &+ \frac{1}{\epsilon}\left(\sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top d - \sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu,\epsilon)\right).\end{aligned} \tag{17}$$

Let $\bar{\imath} \in I_\pi(\hat{x})$ be an index such that $\hat{g}_{\bar{\imath}}(x;\epsilon) = \max_{i\in I_\pi(\hat{x})}\{\hat{g}_i(x;\epsilon)\}$. It is easily seen that $\sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu,\epsilon) \ge \lambda_{\bar{\imath}}(x;\mu,\epsilon)$ and, since

$$\frac{\exp(\hat{g}_i(x;\epsilon)/\mu\epsilon)}{\exp(\hat{g}_{\bar{\imath}}(x;\epsilon)/\mu\epsilon)} \le 1, \qquad \text{for all } i \in \{1,\dots,m\},$$

that

$$\lambda_{\bar{\imath}}(x;\mu,\epsilon) \ge \frac{1}{1+1+\displaystyle\sum_{i=1,i\ne\bar{\imath}}^{m} \frac{\exp(\hat{g}_i(x;\epsilon)/\mu\epsilon)}{\exp(\hat{g}_{\bar{\imath}}(x;\epsilon)/\mu\epsilon)}} \ge \frac{1}{1+m}.$$

Hence we get

$$\sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu,\epsilon) \ge 1/(1+m). \tag{18}$$

By considering (17) and (18), we get

$$\begin{aligned}\nabla Z(x;\mu,\epsilon)^\top d &\le \nabla f(x)^\top d \\ &+ \frac{1}{\epsilon}\left(\sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top d - \frac{1}{1+m}\right).\end{aligned} \tag{19}$$

Now, since $I_\nu(\hat{x}) \subseteq I_\nu(x)$, for $x \in \mathcal{B}(\hat{x};\sigma(\hat{x}))$, by expression (15), it follows that, for any given $\mu > 0$ and $x \in \mathcal{B}(\hat{x};\sigma(\hat{x})) \cap \mathcal{S}_\alpha$ not feasible,

$$\lim_{\epsilon\to 0^+} \lambda_i(x;\mu,\epsilon) = 0, \qquad\qquad i \in I_\nu(\hat{x}).$$

Hence, by the boundedness of $\nabla \hat{g}_i(x;\epsilon)^\top d$ and $\nabla f(x)^\top d$, an $\epsilon(\hat{x}) > 0$ exists such that for all $\epsilon \in (0,\epsilon(\hat{x})]$ we have

$$\sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu,\epsilon)\nabla \hat{g}_i(x;\epsilon)^\top d < \frac{1}{4(m+1)}, \qquad \forall\ \mu \in (0,\mu_{MAX}] \tag{20}$$

$$\nabla f(x)^\top d < \frac{1}{4\epsilon(m+1)}. \tag{21}$$

The result follows from (20), (21) and (19). □

In order to guarantee the global convergence of the algorithm in case when first derivatives are unavailable, it is necessary to get alternative information by sampling the objective function along a suitable set of search directions. Specifically, we follow the approach proposed in [24],

which uses a set of search directions that positively span a "ν-approximation" of the cone of feasible directions; or in other words, the cone of feasible directions with respect to the ν-active linear constraints.
Formally, for any $\nu > 0$ and $x \in \mathcal{S}_\alpha$, we define the set of indices of ν-active linear constraints by

$$J(x;\nu) = \{j : a_j^\top x \geq b_j - \nu\},$$

and the ν-approximation of the cone of feasible directions by

$$T(x;\nu) = \{d \in \Re^n : a_j^\top d \leq 0, \forall j \in J(x;\nu)\}.$$

The following proposition (see [24]) describes some properties of sets $J(x;\nu)$ and $T(x;\nu)$.

Proposition 3.3. *Let $\{x_k\}$ be a sequence of iterates converging towards a point $\bar{x} \in \mathcal{S}_\alpha$. Then, there exists a value $\nu^* > 0$ (depending on $\bar{x}$ only) such that for every $\nu \in (0, \nu^*]$ there exists $\bar{k}_\nu$ such that*

$$J(x_k;\nu) = J(\bar{x}), \tag{22}$$

$$T(x_k;\nu) = T(\bar{x}), \tag{23}$$

for all $k \geq \bar{k}_\nu$.

Proof. See the proof of Proposition 1 in [24]. □

The first step toward defining a derivative-free method for the solution of Problem (16) is to associate a suitable set of search directions with each point x_k produced by the algorithm. This set should have the property that the local behavior of the objective function in each direction in the set provides sufficient information to overcome the lack of the gradient. Formally, we introduce the following assumption.

Assumption 3. *Let $\{x_k\}$ be a sequence of points belonging to $\mathcal{S}_\alpha$ and $\{D_k\}$ a sequence of sets of search directions. Then, for all k,*

$$D_k = \{d_k^i : \|d_k^i\| = 1, \quad i = 1, \ldots, r_k\},$$

and, for some constant $\bar{\nu} > 0$,

$$cone\{D_k \cap T(x_k;\nu)\} = T(x_k;\nu) \quad \forall \nu \in [0, \bar{\nu}].$$

Moreover, $\bigcup_{k=0}^{\infty} D_k$ is a finite set and r_k is bounded.

Assumption 3 is quite a standard assumption in a derivative-free context and is needed to guarantee that the search directions are well-defined and able to capture sufficiently well the local geometry of the feasible set. An example on how to compute a set of directions satisfying the above assumption can be found in the paper [25].

The proposition which follows is essential to prove convergence of the proposed algorithm to a KKT point. In particular, it points out the minimal requirements on the sampling of the smoothed penalty function $Z(x;\mu,\epsilon)$ along the directions p_k^i, $i = 1, \ldots, r_k$ and on the updating of both the smoothing parameter μ and penalty parameter ϵ which are able:

- to enforce global convergence toward stationary points of Problem (4) (point (a));
- to guarantee that feasibility of Problem (1) is attained in a finite number of steps (point (b)).

Proposition 3.4. *Let $\{x_k\}$ be a sequence of points and $\bar{x}$ be a limit point of a subsequence $\{x_k\}_K$, for some infinite set $K \subseteq \{0,1,\ldots\}$, such that $\bar{x} \in \mathcal{S}_\alpha$. Let $\{D_k\}$, with $D_k = \{d_k^1,\ldots,d_k^{r_k}\}$, be a sequence of sets of directions which satisfy Assumption 3 and $J_k = \{i \in \{1,\ldots,r_k\} : d_k^i \in T(x_k;\nu)\}$ with $\nu \in (0, \min\{\bar{\nu},\nu^\star\}]$, where $\nu^\star$ and $\bar{\nu}$ are defined in Proposition 3.3 and Assumption 3, respectively. Suppose that the following conditions hold:*

(i) for each $k \in K$ and $i \in J_k$, there exist y_k^i and scalars $\xi_k^i > 0$ such that:

$$y_k^i + \xi_k^i p_k^i \in \mathcal{S}_\alpha, \qquad Z(y_k^i + \xi_k^i p_k^i;\mu_k,\epsilon_k) \geq Z(y_k^i;\mu_k,\epsilon_k) - o(\xi_k^i); \tag{24}$$

(ii) and, furthermore, $\{\mu_k\}_K$ is a bounded sequence and

$$\lim_{k\to\infty, k\in K} \epsilon_k = 0, \qquad \lim_{k\to\infty, k\in K} \frac{\max_{i\in J_k}\{\xi_k^i, \|x_k - y_k^i\|\}}{\mu_k\epsilon_k} = 0. \tag{25}$$

It results that a $\bar{k} \geq 0$ exists such that, for all $k \in K$ and $k \geq \bar{k}$, x_k is feasible for Problem (1).

Proof. As a first step, we show that for every bounded sequence $\{p_k\} \subset T(x_k;\nu)$ of directions, an infinite set $\bar{K} \subseteq K$ exists such that

$$\lim_{k\to\infty, k\in\bar{K}} \epsilon_k \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k \geq 0. \tag{26}$$

By assumption, the limit point $\bar{x} \in \mathcal{S}_\alpha$, an open neighborhood $\mathcal{B}(\bar{x})$ exists which is strictly contained within $\mathcal{S}_\alpha$. Therefore, by points (a) and (b), we have that, for $k \in K$ and sufficiently large and for all $i \in J_k$, it results that x_k, y_k^i and $y_k^i + \xi_k^i p_k^i$ belong to $\mathcal{B}(\bar{x})$.
By applying the Mean-Value Theorem to (24), we can write

$$-o(\xi_k^i) \leq Z(y_k^i + \xi_k^i p_k^i;\mu_k,\epsilon_k) - Z(y_k^i;\mu_k\epsilon_k) = \xi_k^i \nabla Z(u_k^i;\mu_k,\epsilon_k)^\top p_k^i, \qquad i \in J_k, \tag{27}$$

where $u_k^i = y_k^i + t_k^i \xi_k^i p_k^i$, with $t_k^i \in (0,1)$. By using the Mean-Value Theorem again and the Cauchy-Schwarz inequality, we can write

$$\begin{aligned} \xi_k^i \nabla Z(u_k^i;\mu_k,\epsilon_k)^\top p_k^i &= \xi_k^i \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k^i + \xi_k^i (u_k^i - x_k)^\top \nabla^2 Z(\tilde{u}_k^i;\mu_k,\epsilon_k) p_k^i \\ &\leq \xi_k^i \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k^i + \xi_k^i \|u_k^i - x_k\| \|\nabla^2 Z(\tilde{u}_k^i;\mu_k,\epsilon_k) p_k^i\|, \end{aligned}$$

where $\tilde{u}_k^i = x_k + \tilde{t}_k^i (u_k^i - x_k)$, with $\tilde{t}_k^i \in (0,1)$. By considering expression (14) of $\nabla^2 Z(\tilde{u}_k^i;\mu_k,\epsilon_k)$ and the triangle inequality, we get that

$$
\begin{aligned}
&\xi_k^i \nabla Z(u_k^i;\mu_k,\epsilon_k)^\top p_k^i \le \xi_k^i \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k^i \\
&\quad + \xi_k^i \|u_k^i - x_k\| \left\{ \left\|\nabla^2 f(\tilde u_k^i) p_k^i\right\| + \frac{1}{\epsilon_k}\left\| \sum_{j=1}^m \lambda_j(\tilde u_k^i;\mu_k,\epsilon_k)\nabla^2 \hat g_j(\tilde u_k^i;\epsilon_k) p_k^i \right\| \right. \\
&\quad + \frac{1}{\mu_k\epsilon_k^2}\left\| \sum_{j=1}^m \lambda_i(\tilde u_k^i;\mu_k,\epsilon_k)\nabla \hat g_j(\tilde u_k^i;\epsilon_k)\nabla \hat g_j(\tilde u_k^i;\epsilon_k)^\top p_k^i - \left(\sum_{j=1}^m \lambda_i(\tilde u_k^i;\mu_k,\epsilon_k)\nabla \hat g_j(\tilde u_k^i;\epsilon_k) \right) \right. \\
&\quad \left.\left. \cdot \left(\sum_{j=1}^m \lambda_i(\tilde u_k^i;\mu_k,\epsilon_k)\nabla \hat g_j(\tilde u_k^i;\epsilon_k) \right)^\top p_k^i \right\| \right\}.
\end{aligned}
$$

Since $\{x_k\}_K$ converges, it follows from Assumption 3 and (15) that, for all i and j, $\{x_k\}_K$, $\{\tilde u_k^i\}$, $\{\lambda_j(\tilde u_k^i;\mu_k,\epsilon_k)\}$, $\{p_k^i\}$ are bounded sequences. Therefore, by the continuity assumption on $f(x)$ and $g(x)$, we can find constants c_1, c_2 and c_3 such that

$$
\xi_k^i \nabla Z(u_k^i;\mu_k,\epsilon_k)^\top p_k^i \le \xi_k^i \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k^i + \xi_k^i \left(c_1 + \frac{1}{\epsilon_k} c_2 + \frac{1}{\mu_k\epsilon_k^2} c_3 \right) \|u_k^i - x_k\|. \tag{28}
$$

By (24), (27) and (28), we obtain

$$
\nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k^i + \left(c_1 + \frac{1}{\epsilon_k} c_2 + \frac{1}{\mu_k\epsilon_k^2} c_3 \right) \|u_k^i - x_k\| \ge -\frac{o(\xi_k^i)}{\xi_k^i}
$$

from which, taking into account (13) we can write

$$
\begin{aligned}
\left(\nabla f(x_k) + \frac{1}{\epsilon_k}\sum_{j=1}^m \lambda_j(x_k;\mu_k,\epsilon_k)\nabla \hat g_j(x_k;\epsilon_k) \right)^\top p_k^i \quad + \\
\left(c_1 + \frac{c_2}{\epsilon_k} + \frac{c_3}{\mu_k\epsilon_k^2} \right) \|u_k^i - x_k\| \quad \ge \quad -\frac{o(\xi_k^i)}{\xi_k^i}.
\end{aligned} \tag{29}
$$

Since, by Assumption 3, $\bigcup_{k\in K} D_k$ is a finite set and recalling the boundedness of each sequences $\{\lambda_j(x_k;\mu_k,\epsilon_k)\}$, $j = 1,\ldots,m$, an infinite set $\bar K \subseteq K$ exists such that

$$
\lim_{\substack{k \to \infty \\ k \in \bar K}} \quad x_k = \bar x, \tag{30}
$$

$$
\lim_{\substack{k \to \infty \\ k \in \bar K}} \quad \lambda_j(x_k;\mu_k,\epsilon_k) = \bar\lambda_j \qquad j = 1,\ldots,m. \tag{31}
$$

Furthermore, given the fact that r_k is bounded, a finite set $J \subseteq \{1,2,\ldots\}$ and $\bar p^i \in \Re^n$, $i \in J$, exist such that

$$
\lim_{\substack{k \to \infty \\ k \in \bar K}} \quad p_k^i = \bar p^i, \qquad \forall\ i \in J. \tag{32}
$$

By Proposition 3.3, for all $k \in \bar{K}$ and sufficiently large, $T(x_k;\nu) = T(\bar{x})$, so that, considering the boundedness of sequence $\{p_k\}$, a direction $\bar{p} \in T(\bar{x})$ exists such that

$$\lim_{\substack{k \to \infty \\ k \in \bar{K}}} p_k = \bar{p}, \tag{33}$$

Moreover, by Assumption 3 and by Corollary 10.2 of [25], we know that, for every index $k \in \bar{K}$, $\beta_k^i \geq 0$, $i \in J$, and $\bar{c} > 0$ exist such that

$$p_k = \sum_{i \in J} \beta_k^i \bar{p}_k^i \quad \text{and} \quad \|\beta_k\| \leq \bar{c}\|p_k\|. \tag{34}$$

Since $u_k^i = y_k^i + t_k^i \xi_k^i p_k^i$, with $t_k^i \in (0,1)$, and, by Assumption 3, p_k^i, $i \in J$ are bounded, we have that

$$\left(c_1 + \frac{1}{\epsilon_k}c_2 + \frac{1}{\mu_k \epsilon_k^2}c_3\right)\|u_k^j - x_k\| \leq \left(c_1 + \frac{1}{\epsilon_k}c_2 + \frac{1}{\mu_k \epsilon_k^2}c_3\right)(\|y_k^j - x_k\| + \xi_k^j), \qquad \forall j \in J,$$

which, by using (25), implies

$$\lim_{\substack{k \to \infty \\ k \in \bar{K}}} \left(\epsilon_k c_1 + c_2 + \frac{1}{\mu_k \epsilon_k}c_3\right)\|u_k^j - x_k\| = 0, \qquad \forall j \in J. \tag{35}$$

By multiplying (29) by $\epsilon_k \beta_k^i$, $i \in J$, and summing up, we get, for every index $k \in \bar{K}$

$$\begin{aligned}\left(\epsilon_k \nabla f(x_k) + \sum_{j=1}^m \lambda_j(x_k;\mu_k,\epsilon_k)\nabla \hat{g}_j(x_k;\epsilon_k)\right)^\top \sum_{i \in J} \beta_k^i p_k^i \geq \\ -\left(\left(c_1 + \frac{1}{\epsilon_k}c_2 + \frac{1}{\mu_k \epsilon_k^2}c_3\right)\|u_k^i - x_k\| + \frac{o(\xi_k^i)}{\xi_k^i}\right)\epsilon_k \sum_{i \in J}\beta_k^i.\end{aligned} \tag{36}$$

Now, taking the limit for $k \to \infty$, $k \in \bar{K}$, and recalling (35), (34) and the boundedness of β_k^i, we obtain

$$\lim_{k \to \infty, k \in \bar{K}} \epsilon_k \nabla Z(x_k;\mu_k,\epsilon_k)^\top p_k = \left(\sum_{j=1}^m \bar{\lambda}_j \nabla \hat{g}_j(\bar{x};\epsilon_k)\right)^\top \bar{p} \geq 0. \tag{37}$$

Let now suppose by contradiction that an infinite set $\hat{K} \subseteq K$ exists such that $\lim_{k \to \infty, k \in \hat{K}} x_k = \bar{x}$ and $g(x_k) \not\leq 0$, for all $k \in \hat{K}$. By virtue of Proposition 3.2, given the fact that (25) holds and recalling that, by assumption, $\nu \in (0, \min\{\bar{\nu}, \nu^\star\}]$ so that, by Proposition 3.3, $T(x_k;\nu) = T(\bar{x})$, we have that a $\hat{k} \in \hat{K}$ and a direction $\hat{d} \in T(\bar{x})$ exist such that

$$\epsilon_k \nabla Z(x_k;\mu_k,\epsilon_k)^\top \hat{d} \leq -\frac{1}{2(m+1)}.$$

By setting $p_k = \hat{d}$, for all $k \in \hat{K}$, the above relation constitutes a contradiction with (37) thus completing the proof. □

4. A derivative-free method and global convergence result

In this section we define an algorithm for the solution of Problem (1). The starting points are the non-differentiable exact penalty function $Z(x;\epsilon)$ defined in Section 2 along with its exactness properties and the smooth approximating function introduced in Section 3. Hence, it would be plausible to employ the algorithm proposed in [22]. Roughly speaking, the latter algorithm inexactly solves a sequence of problems (16) when the smoothing parameter μ is driven to zero at a suitable rate. The convergence analysis carried out in [22] guarantees that a subsequence exists which converges toward a stationary point of Problem (4). However, it should be noted that the mentioned result is unsatisfactory in that a proper value for the penalty parameter ϵ is not known "a priori" (namely, ϵ should be smaller than the threshold $\epsilon^\star$ introduced in Propositions 2.6 and 2.7). This implies that the stationary point of Problem (4) might have no connections with the solution of Problem (1).

In this section, by exploiting Proposition 3.4, we define a derivative-free algorithm for Problem (1) which hinges on a suitable automatic updating rule of the penalty parameter that in a finite number of steps is able to find a value below the mentioned threshold $\epsilon^\star$.

The method that we propose uses as a building block the algorithm proposed in [22]. For the sake of clarity, we report a single iteration of the method therein proposed and refer to it as iteration map $\mathcal{M}$.

Iteration map $\mathcal{M}(x,\mu,\tilde{\alpha}^0,\epsilon,q_1) \mapsto (\check{x},\check{\mu},\check{\alpha}^0,\check{\alpha}^{max})$

Data. $\gamma > 0,\ \theta \in (0,1)$.

Step 1. *(Computation of search directions)*

Choose a set of directions $D = \{p^1,\dots,p^r\}$ satisfying Assumption 3.

Step 2. *(Minimization on the cone$\{D\}$)*

Step 2.1. *(Initialization)*

Set $i = 1$, $y^i = x$, $\tilde{\alpha}^i = \tilde{\alpha}^0$.

Step 2.2. *(Computation of the initial stepsize)*

Compute the maximum steplength $\bar{\alpha}^i$ such that $A(y^i + \bar{\alpha}^i p^i) \le b$ and set $\hat{\alpha}^i = \min\{\bar{\alpha}^i, \tilde{\alpha}^i\}$.

Step 2.3. *(Test on the search direction)*

If $\big(\hat{\alpha}^i > 0$ and $Z(y^i + \hat{\alpha}^i p^i;\mu,\epsilon) < Z(y^i;\mu,\epsilon) - \gamma(\hat{\alpha}^i)^2$ and $y^i + \hat{\alpha}^i p^i \in \mathcal{S}_\alpha\big)$ then

compute $\alpha^i =$ *Expansion Step*$(\bar{\alpha}^i, \hat{\alpha}^i, y^i, p^i)$

and set $\tilde{\alpha}^{i+1} = \alpha^i$;

otherwise set $\alpha^i = 0$ and $\tilde{\alpha}^{i+1} = \theta\tilde{\alpha}^i$.

Step 2.4. *(New point)*

Set $y^{i+1} = y^i + \alpha^i p^i$.

Step 2.5. *(Test on the minimization on the cone$\{D\}$)*

If $i = r$, go to Step 3;

otherwise set $i = i + 1$ and go to Step 2.2.

Step 3. *(Iterate outputs)*

Set $\breve{x} = y^{i+1}$.

Set $\breve{\alpha}^0 = \tilde{\alpha}^{i+1}$ and $\breve{\alpha}^{max} = \max_{i=1,\dots,r+1} \{\tilde{\alpha}^i\}$; choose $\breve{\mu} = \min \left\{ \mu, (\breve{\alpha}^{max})^{q_1} \right\}$.

return $(\breve{x}, \breve{\mu}, \breve{\alpha}^0, \breve{\alpha}^{max})$.

where the Expansion Step in step 2.3 is defined as follows

Expansion Step $(\bar{\alpha}^i, \hat{\alpha}^i, y^i, p^i; \alpha^i)$

Data. $\gamma > 0$, $\delta \in (0, 1)$.

Step 1. Set $\alpha = \hat{\alpha}^i$.

Step 2. Let $\tilde{\alpha} = \min\{\bar{\alpha}^i, (\alpha/\delta)\}$.

Step 3. If $y^i + \tilde{\alpha} p^i \notin \mathcal{S}_\alpha$ or $\alpha = \bar{\alpha}^i$ or

$$Z\left(y^i + \tilde{\alpha} p^i; \mu, \epsilon\right) \geq Z(y^i; \mu, \epsilon) - \gamma\, (\tilde{\alpha})^2 .$$

return α.

Step 4. Set $\alpha = \tilde{\alpha}$ and go to Step 2.

We note that the iteration map $\mathcal{M}$ takes, as input arguments, current values for the iterate x, the smoothing parameter μ, the initial step size $\tilde{\alpha}^0$, the penalty parameter ϵ and exponent q_1 and returns, as output arguments, the newly computed iterate $\breve{x}$, smoothing parameter $\breve{\mu}$, initial stepsize for subsequent calls $\breve{\alpha}^0$ and maximum stepsize $\breve{\alpha}^{max}$. For a thorough description of the iteration map $\mathcal{M}$ we refer the interested reader to [22].

On the basis of the iteration map $\mathcal{M}$ so far described, we can define our derivative-free method for the solution of Problem (1).

DeFCon Algorithm

Data. $\tilde{x} \in \mathcal{S}_\alpha$, $\tilde{\alpha}_0^0 > 0$, $\epsilon_0 > 0$, $\mu_{MAX} > 0$, $\gamma > 0$, $\theta \in (0,1)$, $0 < q_1 < q_2 < 1$ and $\tau \in (0,1)$.

Step 0. Set $\mu_0 = \mu_{MAX}$, $z_0 = \tilde{x}$, $x_0 = \tilde{x}$, $j = 0$ and $\epsilon = \epsilon_j$.

Step 1. Set $k = 0$.

Step 2. *(Main iteration)*

Compute $\mathcal{M}(x_k, \mu_k, \tilde{\alpha}_k^0, \epsilon, q_1) \mapsto (x_{k+1}, \mu_{k+1}, \tilde{\alpha}_{k+1}^0, \tilde{\alpha}_{k+1}^{max})$.

Set $k = k + 1$.

Step 3. *(Penalty parameter testing)*

If $\dfrac{(\tilde{\alpha}_k^{max})^{q_2}}{\mu_k} < \min\left\{\epsilon, \max\{0, g_1(x_k), \ldots, g_m(x_k)\}\right\}$ then set $\epsilon = \tau\dfrac{(\tilde{\alpha}_k^{max})^{q_2}}{\mu_k}$,

If $Z(\tilde{x}; \mu_k, \epsilon) \leq Z(x_k; \mu_k, \epsilon)$ then set $x_0 = \tilde{x}$ else $x_0 = x_k$.

Set $\epsilon_{j+1} = \epsilon$, $j = j + 1$, $\mu_0 = \mu_k$ and go to Step 1.

Else go to Step 2.

As mentioned earlier in this section, the crucial aspect of Algorithm DeFCon resides in Step 3 that is in the penalty parameter testing and updating formula.
We remark that the requirement that $0 < q_1 < q_2 < 1$ is essential to prove convergence. In particular, $q_1 \in (0,1)$ is required to prove convergence of the method proposed in [22]; $q_2 \in (0,1)$ is needed to prove that the penalty parameter is updated only a finite number of times; finally, $q_1 < q_2$ is essential to prove that a feasible point is obtained in the limit by algorithm DeFCon.
We recall that the quantity $(\tilde{\alpha}_k^{max})^{q_2}$ can be viewed as a stationarity measure of the current iterate with respect to the smoothing function (see [14]). Then, on the basis of the analysis so far carried out, roughly speaking $(\tilde{\alpha}_k^{max})^{q_2}/\mu_k$ measures how good a solution the current iterate is for Problem (4). So that, the rationale behind the penalty parameter updating is to decrease ϵ whenever an improvement of the quality of the solution of Problem (4) does not correspond to a reduction of the infeasibility of the current iterate with respect to Problem (1). Note, in particular, that if x_k is feasible with respect to the nonlinear inequality constraints then the penalty parameter is left unchanged.

The following proposition is an important result in that it guarantees that the penalty parameter ϵ is reduced finitely many times.

Proposition 4.1. *Let $J = \{0, 1, \ldots\}$ be the index set generated by Algorithm DeFCon at Step 3. Then, J is finite.*

Proof. Suppose that, every time the test at step 3 on the penalty parameter is satisfied and before incrementing the counter j, the following quantities are stored: $\sigma_j^{max} = \tilde{\alpha}_k^{max}$, $\tilde{\sigma}_j^i = \tilde{\alpha}_j^i$,

$\sigma_j^i = \alpha_j^i$ for all $i = 1, \dots, r_{k-1} + 1$, $w_j^i = y_{k-1}^i$ and $d_j^i = p_{k-1}^i$, for all $i = 1, \dots, r_{k-1}$, and $t_j = r_{k-1}$, $z_j = x_k$, $\rho_j = \mu_k$. For the sake of completeness, we set $\rho_{-1} = \mu_{MAX}$.
Reasoning by contradiction, we suppose that J is infinite. By the test of Step 3 we have that

$$\frac{(\sigma_j^{max})^{q_2}}{\rho_j} \le \epsilon_j$$

hence

$$\epsilon_{j+1} = \tau \frac{(\sigma_j^{max})^{q_2}}{\rho_j} \le \tau \epsilon_j$$

so that we get

$$\lim_{j \to \infty} \epsilon_j = 0. \tag{38}$$

Let $\{z_j\}$ be the sequence produced by Algorithm DeFCon. Then, by Assumption 1, an accumulation point $\bar{z} \in \bar{\mathcal{S}}_\alpha$ exists. Let us relabel $\{z_j\}$ the subsequence which converges to $\bar{z}$. Suppose furthermore that $\bar{z} \in \partial \mathcal{D}_\alpha$. By the instructions at Step 3 of Algorithm DeFCon, iteration map $\mathcal{M}$ and by point (i) of Proposition 3.1, we get that

$$Z(z_j; \epsilon_j) \le Z(z_j; \rho_j, \epsilon_j) \le Z(x_0^{(j)}; \mu_0^{(j)}, \epsilon_j) \le Z(x_0^{(j)}; \rho_{j-1}, \epsilon_j), \tag{39}$$

where we denote by $\{x_k^{(j)}\}$ and $\{\mu_k^{(j)}\}$ the sequences produced by DeFCon when $\epsilon = \epsilon_j$.
Furthermore, by the second test at Step 3 of Algorithm DeFCon and by relation (12) we get

$$Z(x_0^{(j)}; \rho_{j-1}, \epsilon_j) \le Z(\tilde{x}; \rho_{j-1}, \epsilon_j) \le Z(\tilde{x}; \epsilon_j) + \rho_{j-1} \ln m. \tag{40}$$

Hence, by (39) and (40) and multiplying by ϵ_j, we obtain

$$\epsilon_j Z(z_j; \epsilon_j) \le \epsilon_j Z(\tilde{x}; \epsilon_j) + \epsilon_j \rho_{j-1} \ln m. \tag{41}$$

Since, when $j \to \infty$, $z_j \to \bar{z} \in \partial \mathcal{D}_\alpha$, an index $i \in \{1, \dots, m\}$ must exist such that $g_i(z_j) \to \alpha_i$ so that $\hat{g}_i(z_j; \epsilon_j) \to +\infty$. Therefore, we get that

$$\lim_{j \to \infty} \epsilon_j Z(z_j; \epsilon_j) = \lim_{j \to \infty} \max\{0, \hat{g}_1(z_j; \epsilon_j), \dots, \hat{g}_m(z_j; \epsilon_j)\} = +\infty. \tag{42}$$

On the contrary,

$$\lim_{j \to \infty} \epsilon_j Z(\tilde{x}; \epsilon_j) + \epsilon_j \rho_{j-1} \ln m = \max\{0, g_1(\tilde{x}), \dots, g_m(\tilde{x})\} < +\infty. \tag{43}$$

Thus, (41), (42) and (43) prove that $\bar{z}$ cannot be on $\partial \mathcal{D}_\alpha$ therefore $\bar{z} \in \mathcal{S}_\alpha$.
Now, the test and the instructions at Step 3 of Algorithm DeFCon yield that for every index j,

$$\frac{(\sigma_j^{max})^{q_2}}{\rho_j} < \min\Big\{\epsilon, \max\{0, g_1(z_j), \dots, g_m(z_j)\}\Big\} \le \epsilon = \epsilon_j, \tag{44}$$

which, recalling that the sequence $\{\rho_j\}$ is bounded above, implies that $\lim_{j \to \infty} (\sigma_j^{max})^{q_2} = 0$, hence

$$\lim_{j \to \infty} \sigma_j^{max} = 0. \tag{45}$$

Now we show that $\lim_{j\to\infty} w_j^i = \bar{z}$, for all $i = 1, \dots, t_j$. To this aim, recalling the definition of σ_j^{max} and the instructions at Step 2.3 of the iteration $\mathcal{M}$, we can write

$$\|w_j^i - z_j\| \le t_j \sigma_j^{max}, \qquad \text{for all } i \in \{1, \dots, t_j\},$$

which, by (45) and by the boundedness of t_j, by Assumption 3, yield

$$\lim_{j\to\infty} \|w_j^i - z_j\| = 0, \qquad \text{for all } i \in \{1, \dots, t_j\}. \tag{46}$$

Hence, by the fact that $z_j \to \bar{z}$ we obtain that

$$w_j^i \to \bar{z}, \qquad \forall\ i = 1, \dots, t_j. \tag{47}$$

By (47), (45) and the fact that $\bar{z} \in \mathcal{S}_\alpha$, we have that, for sufficiently large values of j, $w_j^i + \tilde{\sigma}_j^i d_j^i \in \mathcal{S}_\alpha$ and $w_j^i + \sigma_j^i d_j^i \in \mathcal{S}_\alpha$. We recall that point (ii) of Proposition 6 in reference [22] holds. Therefore, by the instructions of Step 3 and (46), we obtain that, for sufficiently large values of j, either

$$w_j^i + \frac{\sigma_j^i}{\delta} d_j^i \in \mathcal{S}_\alpha \text{ and } Z\left(w_j^i + \frac{\sigma_j^i}{\delta} d_j^i; \rho_j, \epsilon_j\right) \ge Z(w_j^i; \rho_j, \epsilon_j) - \gamma \left(\frac{\sigma_j^i}{\delta}\right)^2$$

or

$$w_j^i + \tilde{\sigma}_j^i d_j^i \in \mathcal{S}_\alpha \text{ and } Z\left(w_j^i + \tilde{\sigma}_j^i d_j^i; \rho_j, \epsilon_j\right) \ge Z(w_j^i; \rho_j, \epsilon_j) - \gamma \left(\tilde{\sigma}_j^i\right)^2.$$

are satisfied. Now, setting $\xi_j^i = \frac{\sigma_j^i}{\delta}$ in the first case and $\xi_j^i = \tilde{\sigma}_j^i$ in the second one, we have, for sufficiently large valules of j,

$$w_j^i + \xi_j^i d_j^i \in \mathcal{S}_\alpha \text{ and } Z\left(w_j^i + \xi_j^i d_j^i; \rho_j, \epsilon_j\right) \ge Z(w_j^i; \rho_j, \epsilon_j) - \gamma \left(\xi_j^i\right)^2. \tag{48}$$

From the updating formula for y^i in Step 2.4 of the iteration $\mathcal{M}$, we note that

$$\|w_j^i - z_j\| \le \sum_{l=1}^{i-1} \sigma_j^l \le \delta \sum_{l=1}^{i-1} \xi_j^l \le \delta t_j \max_{l=1,\dots,t_j} \{\xi_j^l\}, \tag{49}$$

from which we get that

$$\max_{i=1,\dots,t_j} \{\xi_j^i, \|z_j - w_j^i\|\} \le \max\{1, \delta t_j\} \max_{i=1,\dots,t_j} \{\xi_j^i\} \le t_j \max_{i=1,\dots,t_j} \{\tilde{\sigma}_j^i, \sigma_j^i\}. \tag{50}$$

From (44), we have that, for every index j,

$$\max_{i=1,\dots,t_j} \{(\tilde{\sigma}_j^i)^{q_2}, (\sigma_j^i)^{q_2}\} = (\sigma_j^{max})^{q_2} < \rho_j \epsilon_j,$$

which implies that

$$(\epsilon_j \rho_j)^{1/q_2} > \sigma_j^{max}, \tag{51}$$

so that, by (50) and (51), we obtain $\max_{i=1,\dots,t_j} \{\xi_j^i, \|z_j - w_j^i\|\} < t_j (\epsilon_j \rho_j)^{1/q_2}$, from which, recalling that $0 < q_2 < 1$, we get

$$\lim_{j\to\infty} \frac{\max_{i=1,\ldots,t_j}\{\xi_j^i, \|z_j - w_j^i\|\}}{\epsilon_j\rho_j} = 0. \tag{52}$$

By considering (38), (48) and (52) we have that the hypotheses of point (b) of Proposition 3.4 are satisfied so that a $\bar{\jmath} \geq 0$ exists such that, for all $j \geq \bar{\jmath}$, z_j is feasible for Problem (1).
On the other hand, by the instruction of Step 3 of Algorithm DeFCon, we have that, for every index j

$$\frac{(\sigma_j^{max})^{q_2}}{\rho_j} < \max\{0, g_1(z_j), \ldots, g_m(z_j)\},$$

which means that $z_j \in \mathcal{S}_\alpha$ and $g(z_j) \not\leq 0$, for every index j. The latter contradicts what has just been proved, namely that z_j is feasible for Problem (1) for j sufficiently large, thus completing the proof. □

The above proposition guarantees that, after finitely many times, the test at Step 3 of Algorithm DeFCon is never satisfied so that the penalty parameter ϵ stays fixed at its last value, say $\epsilon_{\bar{\jmath}}$, and $\{\epsilon_j\}$, $\{z_j\}$, $\{\rho_j\}$ are all finite sequences. Therefore, from now on, we shall assume that $\epsilon = \epsilon_{\bar{\jmath}}$.
The following proposition describes some properties concerning the sequences of points and of objective function values generated by Algorithm DeFCon and the sampling technique adopted.

Proposition 4.2 (See [22]) *Let $\{x_k\}$, $\{\mu_k\}$ be the sequences generated by Algorithm* DeFCon *when $\epsilon = \epsilon_{\bar{\jmath}}$. Then*

(a) *$\{x_k\}$ is well defined;*

(b) *the sequence $\{x_k\}$ is bounded;*

(e) *the following limits hold:*

$$\lim_{k\to\infty} \max_{i=1,\ldots,r_k} \{\alpha_k^i\} = 0, \tag{53}$$

$$\lim_{k\to\infty} \max_{i=1,\ldots,r_k} \{\tilde{\alpha}_k^i\} = 0, \tag{54}$$

$$\lim_{k\to\infty} \max_{i=1,\ldots,r_k} \|x_k - y_k^i\| = 0. \tag{55}$$

As showed in reference [22] carrying out the convergence analysis, a significant role is played by the index set K defined as follows

$$K = \{k : \mu_{k+1} < \mu_k\}. \tag{56}$$

Indeed, the following proposition shows that every accumulation point of the sequence $\{x_k\}_K$ is a KKT point for Problem (1).

Proposition 4.3. *Let $\{x_k\}$ be the sequence generated by Algorithm DeFCon when $\epsilon = \epsilon_{\bar{\jmath}}$. Then the sequence $\{x_k\}$ is bounded and every accumulation point $\bar{x}$ of $\{x_k\}_K$, where K is defined by (56), is a KKT point for Problem (1).*

Proof. First of all we prove that $\bar{x}$ is feasible for Problem (1). Since ϵ is no longer updated, from the instructions of Step 3 of Algorithm DeFCon we know that, for every index $k \in K$

$$0 \leq \min\{\epsilon_{\bar{j}}, \max\{0, g_1(x_{k+1}), \ldots, g_m(x_{k+1})\}\} \leq \frac{(\tilde{\alpha}_{k+1}^{max})^{q_2}}{\mu_{k+1}},$$

and, by the instruction at Step 3 of the iteration $\mathcal{M}$,

$$\frac{(\tilde{\alpha}_{k+1}^{max})^{q_2}}{\mu_{k+1}} = (\tilde{\alpha}_{k+1}^{max})^{q_2 - q_1},$$

which, taking the limit for $k \to \infty, k \in K$, recalling the results of Proposition 4.2 and the fact that $q_1 < q_2$, implies that

$$\lim_{k\to\infty, k\in K} \max\{0, g_1(x_{k+1}), \ldots, g_m(x_{k+1})\} = 0. \tag{57}$$

Let now $\bar{x}$ be an accumulation point of sequence $\{x_{k+1}\}_K$, that is, an infinite index set $\hat{K} \subseteq K$ exists such that

$$\lim_{k\to\infty, k\in \hat{K}} x_{k+1} = \bar{x}.$$

On account of relation (57) we have that

$$\lim_{k\to\infty, k\in \hat{K}} \max\{0, g_1(x_{k+1}), \ldots, g_m(x_{k+1})\} = 0,$$

which means that $\bar{x}$ is such that $\bar{x} \in \mathcal{F}$.
By employing (55) of Proposition 4.2 and the definition of the iteration $\mathcal{M}$, we have that

$$\lim_{k\to\infty} \|x_k - x_{k+1}\| = 0.$$

Hence, we know that

$$\lim_{k\to\infty, k\in \hat{K}} x_k = \lim_{k\to\infty, k\in \hat{K}} x_{k+1},$$

so that $\lim_{k\to\infty, k\in \hat{K}} x_k = \bar{x} \in \mathcal{F}$.
Finally, we show that $\bar{x}$ is a KKT point for Problem (1). By the instruction of the iteration $\mathcal{M}$, every point $x_k \in \mathcal{S}_\alpha$ whose closure is compact by Assumption 1. Hence the sequence $\{x_k\}$ is bounded and therefore it admits limit points.
Let now $\bar{x}$ be any accumulation point of the subsequence $\{x_k\}_K$, where K is defined by (56). By the first part of the proof, we know that $\bar{x} \in \mathcal{F}$. Furthermore, by Corollary 1 of reference [22], we have that $\bar{x}$ is a stationary point of the exact penalty function $Z(x;\epsilon)$, so that, by Proposition 2.5, $\bar{x}$ is a KKT point for Problem (1). □

5. Case study: constrained parameter estimation for glucose kinetics model

The aim of the paper is mainly theoretical. Thus the development of an efficient code based on the proposed algorithm and the analysis of its numerical performances are beyond the scope of the present paper. We refer to [11] for a numerical experimentation on standard test problems. In [11] many different derivative-free approaches for constrained optimization problems are

analyzed from a computational point of view. An extensive numerical experimentation is carried out on a large set of test problems from the CUTEr collection [10]. The conclusion of [11] is that the use of a smooth approximation of an ℓ_∞ exact penalty function leads to a derivative-free algorithm which shows a good compromise between quality of the final point and number of function evaluations required to get convergence.
Encouraged by these results we used a rough Matlab implementation of the proposed algorithm to solve a real world problem connected with the study of an insulin-glucose model of the human body. The study and understanding of circulatory models of glucose kinetics are of great importance in medicine and biology. Such models study the response of body tissues to an impulsive injection of a glucose bolus of known quantity. The intravenous glucose tolerance test (IVGTT) is a simple and standardized test that allows to measure the reaction of the organism to the mentioned impulsive perturbation of the steady state. IVGTT has a documented ability to asses the functioning of the key organs involved in glucose homeostasis. Moreover, it is a powerful tool in the study of diabetes mellitus in that it is able to provide information on beta-cell function and insulin sensitivity (both peripheral and hepatic).
The IVGTT experimental protocol used prescripts the following operations [15].

- Collection of $3ml$ blood basal samples at -30, -15 and 0 minutes to glucose injection.
- Injection of a $300mg/kg$ glucose bolus at 0 min. immediately after the collection of the last basal sample.
- Infusion, at 20 min. from injection, of $0.03U/kg$ insulin at a constant rate for 5 minutes.
- Collection of $3ml$ blood samples at 2, 3, 4, 5, 6, 8, 10, 15, 20, 25, 30, 40, 60, 80, 100, 120, 140, 160, 180, 210 and 240 min. from injection of glucose, for measurements of glucose, glucose tracer, insulin and C-peptide concentrations.

In the circulatory model of glucose kinetics studied in [17, 18, 19], the body tissues are lumped into two blocks. The heart-lungs block represents the heart chambers and the lungs, i.e. the tissues in between the right atrium and left ventricle. The periphery block represents all the remaining tissues, nourished by the entire arterial tree originating from the left ventricle (including the heart tissues nourished by the coronaries).
The dynamics of the exogenous glucose concentration during the IVGTT can be modelled by the following system of ordinary differential equations.

$$\frac{dG_A(t)}{dt} = -\lambda G_A(t) + \lambda\Big[G_1(t) + G_2(t) + J/F\Big], \tag{58a}$$

$$\frac{dG_1(t)}{dt} = -\alpha_1 G_1(t) + \alpha_1\vartheta\Big[1 - E_b - \gamma Z(t)\Big]G_A(t), \tag{58b}$$

$$\frac{dG_2(t)}{dt} = -\alpha_2 G_2(t) + \alpha_2(1-\vartheta)\Big[1 - E_b - \gamma Z(t)\Big]G_A(t), \tag{58c}$$

$$\frac{dZ(t)}{dt} = -\beta Z(t) + \beta\Big[I(t) - I_b\Big], \quad Z(0) = 0. \tag{58d}$$

Here, $G_A(t)$ denotes the exogenous glucose concentration, $I(t)$ is the insulin concentration during the exam, $J = 300mg/kg$ is the known intravenous glucose infusion, $F = 2688ml\,\cdot$

$min^{-1} \cdot m^{-2}$ is the cardiac output, $\lambda = 3.84 min^{-1}$ is the reciprocal of the mean heart-lungs transit time [20] and is patient-independent, I_b is the basal value of insulin concentration and is patient-specific, we used $I_b = 50 pmol/l$. The numerical solution of the system of ODE (58) requires that the insulin concentration increment $I(t) - I_b$ in (58d) is available at every time instant. For this purpose, the measured values of $I(t) - T_b$ have been smoothed and interpolated by a continuous function of time as detailed in [16]. Finally, $\alpha_1, \alpha_2, \beta, \gamma, \vartheta, E_b$ are the model parameters to be estimated in such a way that $G_A(t)$ approximates as well as possible the measurements gathered during the IVGTT. The model parameters, as suggested by biomedical engineers, can be sensibly bounded both from below and above as follows.

$$0.5 \le \alpha_1 \le 5, \qquad 0.01 \le \alpha_2 \le 0.5, \qquad 0.3 \le \vartheta \le 0.9, \qquad 0.01 \le \beta \le 1$$
$$0.01 \le E_b \le 0.1, \qquad 5 \cdot 10^{-5} \le \gamma \le 5 \cdot 10^{-4}.$$

Let us define $t = (2, 3, 4, 5, 6, 8, 10, 15, 20, 25, 30, 40, 60, 80, 100, 120, 140, 160, 180, 210, 240)^\top$ and let g_i denote the exogenous glucose concentration measured at time t_i during the IVGTT. Then

$$f(\alpha_1, \alpha_2, \beta, \gamma, \vartheta, E_b) = \sum_{i=1}^{21} (G_A(t_i) - g_i)^2,$$

is the sum of squared errors between model prediction and actual measurements. Moreover, among all the possible models we are interested in those for which the mean periphery transit time $\vartheta/\alpha_1 + (1 - \vartheta)/\alpha_2$ is greater than or equal to 2.5 minutes. Thus, we end up with the following constrained problem

$$\begin{array}{ll} \min f(\alpha_1, \alpha_2, \beta, \gamma, \vartheta, E_b) & \\ s.t. \quad g(\alpha_1, \alpha_2, \vartheta) = \vartheta/\alpha_1 + (1 - \vartheta)/\alpha_2 \ge 2.5 & \\ \quad 0.5 \le \alpha_1 \le 5 & \\ \quad 0.01 \le \alpha_2 \le 0.5 & \\ \quad 0.3 \le \vartheta \le 0.9 & \\ \quad 0.01 \le \beta \le 1 & \\ \quad 0.01 \le E_b \le 0.1 & \\ \quad 5 \cdot 10^{-5} \le \gamma \le 5 \cdot 10^{-4}. & \end{array} \tag{59}$$

The circulatory model of glucose has been implemented in Matlab/Simulink [28]. The solution of the system of ODE (59) which is at the basis of the model is done numerically with a precision ξ that can be set by the user.

We started our experimentation by comparing the outcomes of DeFCon and the constrained nonlinear minimization routine of Matlab, fmincon, selecting a precision level of the ODE solver $\xi = 10^{-3}$.

Initial values for the parameters, as suggested by biomedical engineers, are

α_1	α_2	β	γ	ϑ	E_b
1.0089	0.40794	0.16481	$3.932 \cdot 10^{-4}$	0.84496	0.020991

In Figure 1 we report the actual measurements of exogenous glucose during IVGTT as crosses and we plot the curve $G_A(t)$ obtained in correspondence to the initial values of the parameters as listed above.

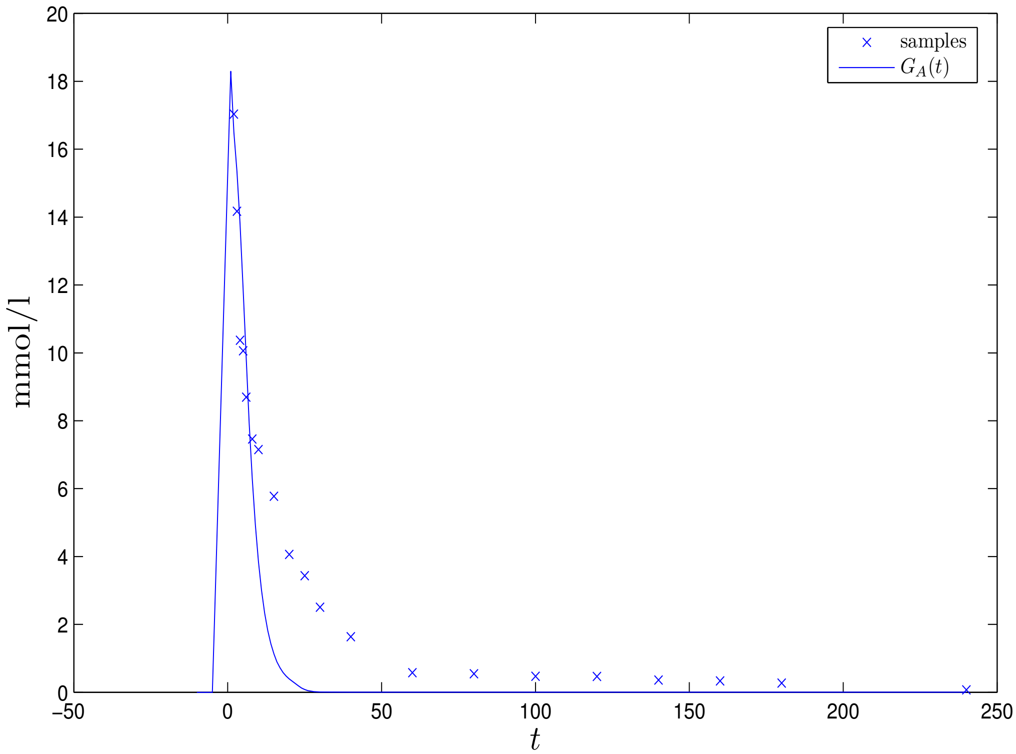

Figure 1: Initial configuration

Starting from this initial point, the proposed derivative-free Algorithm (DeFCon) and the Matlab routine fmincon yield the results reported in the Table 1. where (n.it.) and (n.f.) denote, respectively, number of iterations and number of functions evaluations. $\alpha_1^\star, \alpha_2^\star, \beta^\star, \gamma^\star, \vartheta^\star$ and $E_b^\star$ represent the final values of the model parameters to be estimated. Finally, $f^\star$ and $g^\star$ represent, respectively, the final valules of the objective function and of the nonlinear constraint (which should be greater than or equal to 2.5). Looking at the results it can be noted that, even though both methods manage to achieve feasibility of the final point, DeFCon is able to produce a point whose objective function value is much better than that produced by fmincon. Indeed, the parameter values obtained by DeFCon and fmincon yield two different curves $G_A(t)$ which are substantially different in terms of approximation of the glucose measurements (see Figure 2). This better behavior of DeFCon over fmincon is achieved at the expense of an higher computational burden.

The inefficiency of the Matlab solver fmincon along with the modest number of iterations and function evaluations to get convergence might indicate that the ODE solver tolerance is to high for estimation of first order derivatives by finite differences to be reliable. Hence, we tried to solve the problem with increasing precision levels for the ODE solver, namely we set the precision $\xi = 10^{-4}, 10^{-5}, 10^{-6}$, and compared the results in Table 2.

As concerns the above comparison, we first note that there is only a slight change in the points

	DeFCon	fmincon
n.it.	46	12
n.f.	627	245
$\alpha_1^\star$	2.0228	0.80479
$\alpha_2^\star$	0.11881	0.011788
$\beta^\star$	0.055959	1.0
$\gamma^\star$	$1.6529 \cdot 10^{-4}$	$5.0 \cdot 10^{-5}$
$\vartheta^\star$	0.7409	0.9
$E_b^\star$	0.01	0.01
$f^\star$	2.0959	27.6094
$g^\star$	2.547	9.6017

Table 1: Results obtained by DeFCon and fmincon for $\xi = 10^{-3}$.

ξ	10^{-6}		10^{-5}		10^{-4}	
	DeFCon	fmincon	DeFCon	fmincon	DeFCon	fmincon
n.it.	97	21	68	9	57	9
n.f.	1410	316	994	157	827	146
$\alpha_1^\star$	2.1468	2.2779	2.1659	0.58594	2.1261	0.5
$\alpha_2^\star$	0.12418	0.049484	0.12401	0.012527	0.12318	0.036089
$\beta^\star$	0.067333	0.95153	0.067373	1.0	0.06504	0.01
$\gamma^\star$	$1.5059 \cdot 10^{-4}$	$5.0 \cdot 10^{-5}$	$1.5038 \cdot 10^{-4}$	$5.0 \cdot 10^{-5}$	$1.5314 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$
$\vartheta^\star$	0.73186	0.86005	0.73182	0.9	0.73407	0.89948
$E_b^\star$	0.01	0.033761	0.01	0.01	0.01	0.01
$f^\star$	2.0061	63.7352	2.0075	69.183	2.0146	103.9875
$g^\star$	2.5002	3.2058	2.5005	9.5193	2.504	4.5842

Table 2: Results of fmincon and DeFCon for different values of precision parameter ξ

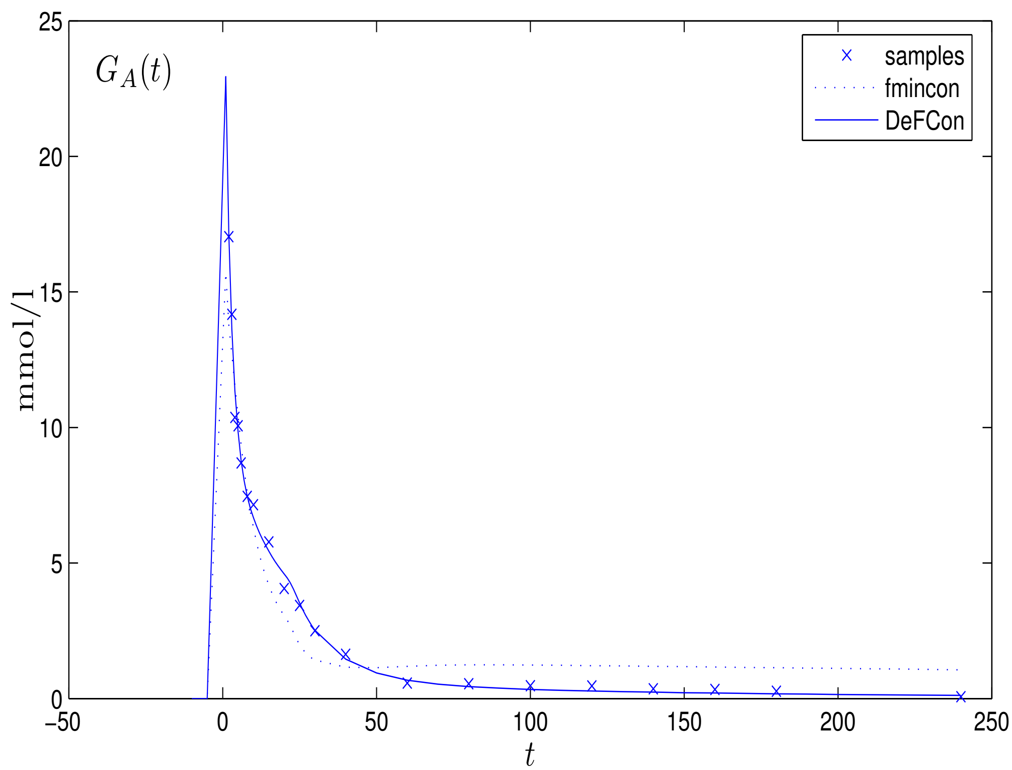

Figure 2: Optimal curves

produced by DeFCon. Namely, as the ODE solver precision ξ increases, DeFCon, though requiring more iterations and function evaluations to converge, produces points which are very close to each other. This is confirmed by the objective and constraint function values which gain more and more accuracy as the precision ξ becomes finer.
On the contrary, fmincon exhibits a more unpredictable behavior converging to points that are largely different from each other both in terms of parameter, objective and constraint function values. The outcomes of fmincon seem to be unrelated with the precision level of the ODE solver apart for the fact that the computational burden increases as ξ gets lower. This inefficiency of fmincon can be due to the lack of derivative knowledge on the problem which fmincon tries to overcome by computing gradients by finite difference approximation. This, in turn makes fmincon more subject to the numerical noise introduced by the ODE solver thus explaining the apparent instability of the code.

6. Conclusions

In the paper we presented a derivative-free algorithm for the solution of inequality constrained nonlinear programming problems. The method is based on the derivative-free minimization of a smooth approximation of a non-differentiable exact penalty function. We proved the method

to be globally convergent toward KKT point of the constrained problem. In order to stress the efficiency of our method to tackle real world problems, we report the results obtained on a constrained problem concerning the parameter estimation of an insulin-glucose model of the human body. A comparison with a standard optimization routine show the effectiveness of the proposed method.
The convergence properties and theoretical analysis of the proposed method has been carried out in the case when only inequality constraints are present. The method can be adapted to handle both equality and inequality constraints preserving its convergence properties but at the expense of some non-trivial technicalities which considerably complicate the analysis.

Acknowledgments

The authors wish to express thankfulness to Dr. Andrea Mari for providing them the applicative problem along with many useful insights that greatly helped in understanding and using the Matlab/Simulink model.

Appendix

Proof of Proposition 2.5. Since $\bar{x} \in \mathcal{F}$, then $B(\bar{x};\epsilon) = I_0(\bar{x})$. Therefore, by Proposition 2.3, we have

$$\left(\nabla f(\bar{x}) + \sum_{i\in I_0(\bar{x})} \frac{\lambda_i((\alpha_i - g_i(x))^2 + \epsilon\alpha_i)}{\epsilon(\alpha_i - g_i(x))^2}\nabla g_i(\bar{x})\right)^\top d \geq 0, \quad \text{for all } d \in T(\bar{x}).$$

Then, by setting $\bar{\lambda}_i = \frac{\lambda_i((\alpha_i - g_i(x))^2 + \epsilon\alpha_i)}{\epsilon(\alpha_i - g_i(x))^2}$, $i \in I_0(\bar{x})$, and $\bar{\lambda}_i = 0$, $i \in \{1,\dots,m\} \setminus I_0(\bar{x})$, we have that there does not exist any direction $d \in \Re^n$ such that

$$\begin{aligned} &\left(\nabla f(\bar{x}) + \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x})\right)^\top d < 0, \\ &a_j^\top d \leq 0, \quad \forall j \in J(\bar{x}). \end{aligned}$$

Hence, by using the Motzkin theorem [27], we have that $y_0 > 0$ and $\mu_j \geq 0$, $j \in J(\bar{x})$, exist such that

$$y_0\left(\nabla f(\bar{x}) + \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x})\right) + \sum_{j\in J(\bar{x})} a_j \mu_j = 0.$$

The result follows, by taking $\bar{\mu}_j = \mu_j/y_0$, for $j \in J(\bar{x})$, and $\bar{\mu}_j = 0$, for $j \notin J(\bar{x})$. □

In order to complete the proof of the exactness results of the penalty function $Z(x;\epsilon)$, we need some technical result which are reported in the following propositions.

Proposition .1. *Let $\hat{x} \in \hat{\mathcal{S}}_\alpha$; then there exists numbers $\epsilon(\hat{x}) > 0$ and $\sigma(\hat{x}) > 0$, such that, for all $\epsilon \in (0, \epsilon(\hat{x})]$ and for all $x \in \mathcal{B}(\hat{x}, \sigma(\hat{x})) \cap \mathcal{S}_\alpha$ and $g(x) \not\leq 0$, there exists a direction $d \in T(x)$ satisfying $DZ(x, \hat{d}; \epsilon) < 0$.*

Proof. By Assumption 2, we have that the hypotheses of Lemma 2.4 are satisfied at $\hat{x}$ for $I = I_\pi(\hat{x})$. Let $\mathcal{B}(\hat{x}, \rho)$ and $d \in T(\hat{x})$ be the neighborhood and the direction considered in Lemma 2.4, we have that $d \in T(\hat{x})$ is such that

$$\nabla \hat{g}_i(x; \epsilon)^\top d \leq -1, \tag{60}$$

for all $i \in I_\pi(\hat{x})$. By continuity, we can find a neighborhood $\mathcal{B}(\hat{x}, \sigma(\hat{x})) \subseteq \mathcal{B}(\hat{x}, \rho)$ such that, for $i \notin I_\pi(\hat{x})$ and $x \in \mathcal{B}(\hat{x}, \sigma(\hat{x})) \cap \mathcal{S}_\alpha$, we have $g_i(x) < 0$; it follows that $I_\pi(x) \subseteq I_\pi(\hat{x})$, for $x \in \mathcal{B}(\hat{x}, \sigma(\hat{x})) \cap \mathcal{S}_\alpha$.
Let now $x \in \mathcal{B}(\hat{x}, \sigma(\hat{x})) \cap \mathcal{S}_\alpha$ be an infeasible point, that is, $g(x) \not\leq 0$. Then, there must exist at least an index $i \in I_\pi(\hat{x})$ such that $g_i(x) > 0$ and $\hat{g}_i(x; \epsilon) > 0$, so that, it results $B(x; \epsilon) \subseteq I_\pi(x)$. Therefore, recalling the expression of the directional derivative of $Z(x; \epsilon)$ and (60), we get

$$DZ(x, d; \epsilon) = \nabla f(x)^\top d + \frac{1}{\epsilon} \max_{i \in B(x;\epsilon)} \{\nabla \hat{g}_i(x; \epsilon)^\top d\} \leq \nabla f(x)^\top d - \frac{1}{\epsilon},$$

from which it follows that a value $\epsilon(\hat{x}) > 0$ exists such that, for all $\epsilon \in (0, \epsilon(\hat{x})]$ and $x \in \mathcal{B}(\hat{x}, \sigma(\hat{x})) \cap \mathcal{S}_\alpha$ with $x \notin \mathcal{F}$, it must hold that

$$DZ(x, d; \epsilon) < 0$$

which concludes the proof. □

Proposition .2. *Let $\bar{\lambda} \in \Re^m$ and $\bar{\mu} \in \Re^p$ be multipliers such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple for Problem (1). Then the following bound holds*

$$\|\bar{\lambda}\|_q \leq \nabla f(\bar{x})^\top z,$$

where $z \in T(\bar{x})$ is a vector such that

$$\nabla g_i(\bar{x})^\top z \leq -1, \quad i \in I_0(\bar{x}). \tag{61}$$

Proof. From the fact that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple, it follows that, for any $z \in T(\bar{x})$ satisfying (61), we have

$$\nabla f(\bar{x})^\top z = -\sum_{i \in I_0(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x})^\top z - \sum_{j \in J(\bar{x})} \bar{\mu}_j a_j^\top z \geq 0.$$

Therefore, the following linear program and its dual are both feasible and bounded

$$\begin{array}{lll} \min\limits_z & \nabla f(\bar{x})^\top z & \\ & \nabla g_i(\bar{x})^\top z \leq -1, & i \in I_0(\bar{x}) \\ & a_j^\top z \leq 0, & j \in J(\bar{x}), \end{array} \tag{62}$$

$$\begin{array}{ll} \max\limits_{u,v} & \sum\limits_{i \in I_0(\bar{x})} u_i \\ & \sum\limits_{i \in I_0(\bar{x})} \nabla g_i(\bar{x}) u_i + \sum\limits_{i \in I_0(\bar{x})} a_j v_j = -\nabla f(\bar{x}), \\ & u, v \geq 0. \end{array} \tag{63}$$

Let $z^\star$ and $(u^\star, v^\star)$ be optimal solutions of (62) and (63) respectively. Recalling that every KKT multipliers (λ, μ) of Problem (1) satisfy the constraints of Problem (63), then we have

$$\|\bar{\lambda}\|_q \leq \|\bar{\lambda}\|_1 \leq \sum_{i\in I_0(\bar{x})} u_i^\star = \nabla f(\bar{x})^\top z^\star \leq \nabla f(\bar{x})^\top z,$$

for any $z \in T(\bar{x})$ satisfying (61). ◁

Proposition .3 ([7], Proposition 8) *A number Λ exists such that $\|\bar{\lambda}\|_\infty \leq \Lambda$ for all KKT triples $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of Problem (1).*

Now, we can finally prove Propositions 2.6 and 2.7.

Proof of Proposition 2.6.
"If"-part: it follows from Proposition 10 in [8].
"Only if"-part: as $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple for Problem 1 we can write:

$$\nabla f(\bar{x}) = -\left(\sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x}) + \sum_{j\in J(\bar{x})} \bar{\mu}_j a_j \right).$$

Recalling the expression of the directional derivative of $Z(x;\epsilon)$

$$DZ(x, d; \epsilon) = \nabla f(x)^\top d + \frac{1}{\epsilon} \max_{i\in B(x;\epsilon)} \{\nabla \hat{g}_i(x;\epsilon)^\top d\},$$

and the fact that $\bar{x} \in \mathcal{F}$ so that $B(\bar{x};\epsilon) = I_0(\bar{x})$, for all $\epsilon > 0$, we have that

$$DZ(\bar{x}, d; \epsilon) \geq \frac{1}{\epsilon} \max_{i\in I_0(\bar{x})} \{\max\{\nabla \hat{g}_i(\bar{x};\epsilon)^\top d, 0\}\} - \left(\sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \max\{\nabla g_i(\bar{x})^\top d, 0\} + \sum_{j\in J(\bar{x})} \bar{\mu}_j a_j^\top d \right).$$

Whenever $d \in T(\bar{x})$, by definition of $T(\bar{x})$, it results

$$DZ(x, d; \epsilon) \geq \frac{1}{\epsilon} \max_{i\in I_0(x)} \{\max\{\nabla \hat{g}_i(x;\epsilon)^\top d, 0\}\} - \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \frac{\alpha_i}{\alpha_i + \epsilon} \max\{\nabla \hat{g}_i(\bar{x};\epsilon)^\top d, 0\}$$

Now, recalling Proposition .3, we have that

$$\begin{aligned} \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \frac{\alpha_i}{\alpha_i + \epsilon} \max\{\nabla \hat{g}_i(\bar{x};\epsilon)^\top d, 0\} &\leq \max_{i\in I_0(\bar{x})} \{\max\{\nabla \hat{g}_i(\bar{x};\epsilon)^\top d, 0\}\} \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i \\ &\leq \max_{i\in I_0(\bar{x})} \{\max\{\nabla \hat{g}_i(\bar{x};\epsilon)^\top d, 0\}\} m\Lambda, \end{aligned}$$

so that we can say that $\bar{x}$ is a critical point of Problem (4) for all $\epsilon \in (0, \epsilon^\star]$ where $\epsilon^\star = 1/m\Lambda$. ◁

Proof of Proposition 2.7. The proof follows by considering Propositions 2.5, .1 and references [6, 8]. ◁

References

[1] C. Audet and J. E. Dennis Jr., *A pattern search filter method for nonlinear programming without derivatives*, SIAM J. Optimization, 14, 4 (2004), pp. 980–1010.

[2] C. Audet and J. E. Dennis Jr., *Mesh Adaptive Direct Search Algorithms for Constrained Optimization*, SIAM J. Optimization, 17, 1 (2006), pp. 188–217.

[3] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, New York, 1999.

[4] D. P. Bertsekas, *Constrained Optimization and Lagrange Multipliers Methods*, Academic Press, New York, 1982.

[5] A.R. Conn, N.I.M. Gould and Ph.L. Toint, *globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM J. Numer. Anal., vol. 28, (1989) pp. 545-572.

[6] G. Di Pillo and L. Grippo, *On the Exactness of a Class of Nondifferentiable penalty Functions*, J. of Optim. Theory and Appl., 57, no. 3, pp. 399-410 (1988)

[7] G. Di Pillo and L. Grippo, *Globally exact nondifferentiable exact penalty functions*, Department of Computer and Systems science, University of Rome "La Sapienza" R.10.87 (1987)

[8] G. Di Pillo and L. Grippo, *Exact penalty functions in constrained optimization*, SIAM J. on Control and Optimization, vol. 27, no. 3, pp. 1333-1360 (1989)

[9] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Math. Programming, 91 (2002), pp. 239–269.

[10] N. I. M. Gould, D. Orban and Ph. L. Toint, *CUTEr and SifDec: a constrained and unconstrained testing environment, revisited*, ACM T. Math. Software, 29 (2003), pp. 373-394.

[11] J.D. Griffin and T.G. Kolda, *Nonlinearly-constrained optimization using asynchronous parallel generating set search*, Technical Report SAND2007-3257, Sandia National Laboratories, Albuquerque, NM and Livermore, CA. (2007)

[12] S.P. Han and O.L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Programming, 17 (1979), pp. 251–269.

[13] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, in Lecture notes in Economics and Mathematical Systems, n. 187, Springer-Verlag, Berlin, Heidelberg, New York, 1981.

[14] T. G. Kolda, R. M. Lewis and V. Torczon, *Stationarity results for generating set search for linearly constrained optimization*, Sandia Tech. rep. SAND2003-8550 (2003).

[15] A. Mari, *Assessment of insulin sensitivity and secretion with the labelled intravenous glucose tolerance test: improved modelling analysis*, Diabetologia, 41, pp. 1029–1039 (1998).

[16] A. Mari and A. Valerio, *A circulatory model for the estimation of insulin sensitivity*, Control Eng. Practice, 5, pp. 1747–1752 (1997).

[17] A. Mari, *Circulatory models of intect-body kinetics and their relationship with compartmental and noncompartmental analysis*, J. Theo. Biol., 160, pp. 509–531 (1993).

[18] A. Mari, *Determination of the single-pass impulse response of the body tissues with circulatory models*, IEEE Trans. Biomedical Eng., 42, pp. 304–312 (1995).

[19] A. Mari, *Calculation of organ and whole-body uptake and production with the impulse response approach*, J. Theo. Biol., 174, pp. 341–353 (1995).

[20] N.A. Lassen and W. Perl, *Tracer kinetic methods in medical physiology*, Raven Press, New York (1979).

[21] T.G. Kolda, R.M. Lewis and V. Torczon, *Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods*, SIAM Review, 45 (2003), pp. 385–482.

[22] G. Liuzzi, S. Lucidi and M. Sciandrone, *A derivative-free algorithm for linearly constrained finite minimax problems*, SIAM J. Optimization, 16 (2006), pp. 1054–1075.

[23] S. Lucidi and M. Sciandrone, *On the global convergence of derivative free methods for unconstrained optimization*, SIAM J. Optimization, 13 (2002), pp. 97–116.

[24] S. Lucidi, M. Sciandrone and P. Tseng, *Objective-derivative-free methods for constrained optimization*, Math. Programming Ser. A, 92 (2002), pp. 37–59.

[25] R.M. Lewis and V. Torczon, *Pattern search methods for linearly constrained minimization*, SIAM J. Optimization, 10 (2000), pp. 917–941.

[26] R.M. Lewis and V. Torczon, *A globally convergent augmented lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, SIAM J. Optimization, 12 (2002), pp. 1075–1089.

[27] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[28] The MathWorks inc. *Matlab, the language of technical computing.*

[29] J. H. May, *Linearly constrained nonlinear programming: a solution method that does not require analytic derivatives*, Ph.D. Thesis, Yale University, 1974.

[30] S. Xu, *Smoothing method for minimax problems*, Computational Optimization and Appl., 20 (2001), pp. 267–279.

[31] W. I. Zangwill, *Nonlinear Programming via Penalty Functions*, Management Science, 13, pp. 344-358 (1967).