The Role of *grossone* in Nonlinear Programming and Exact Penalty Methods

Renato De Leone

Abstract Exact penalty methods form an important class of methods for solving constrained optimization problems. Using penalty functions, the original constrained optimization problem can be transformed in an "equivalent" unconstrained problem. In this chapter we show how *grossone* can be utilized in constructing exact differentiable penalty functions for the case of only equality constraints, the general case of equality and inequality constraints, and quadratic problems. These new penalty functions allow to recover the solution of the unconstrained problem from the finite term (in its *grossone* expansion) of the optimal solution of the unconstrained problem. Moreover, Lagrangian duals associated to the constraints are also automatically obtained from the infinitesimal terms. Finally a new algorithmic scheme is presented.

1 Introduction

Penalty methods represent an important class of methods for solving constrained optimization problems. These functions, in their simplest form, are composed of two parts: the original objective function and a penalty function, usually multiplied by a positive scalar called penalty parameter, which penalize the violation of the constraints. Using penalty functions, the original constrained optimization problem can be transformed in an "equivalent" unconstrained problem. Two major issues must be taken into account when constructing such penalty functions: exactness and differentiability. Roughly speaking, exactness requires that the global or local solution of the uncon-

R. De Leone (✉)
School of Science and Technology, University of Camerino, Camerino, MC, Italy
e-mail: renato.deleone@unicam.it

strained problem or a stationary point of it, must correspond to a global or local minimum of the constrained problem or to a point satisfying Karush–Kuhn–Tucker conditions. Using the Euclidean norm, it is quite simple to construct continuously differentiable, or at least differentiable penalty functions, as long as the functions in the original problem are sufficiently smooth. However, in these cases the penalty parameter must be driven to $+\infty$, thus generating numerical instability. Using the 1–norm, instead, it is possible to construct penalty functions that are exact, that is they do not require that the penalty parameter goes to $+\infty$. The difficulty in this case arises from the non–differentiability of the function.

Recently, Sergeyev introduced a new approach to infinite and infinitesimals. The proposed numeral system is based on ①, the number of elements of ℕ, the set of natural numbers. We refer the reader to Chap. 1 for insights on the arithmetic of infinity and the properties of ①. Here we want to stress that ① is not a symbol and is not used to make symbolic calculation. In fact, the new numeral ① is a natural number, and it has both cardinal and ordinal properties, exactly as the "standard", finite natural numbers. Moreover, the new proposed approach is far apart from non–Standard Analysis, as clearly shown in [30]. A comprehensive description of the grossone–based methodology can also be found in [29].

In this chapter we discuss the use of ① in constructing exact differentiable penalty functions for the case of only equality constraints, the general cases of equality and inequality constraints, and quadratic problems. Using this novel penalty function, it is possible to recover the solution of the unconstrained problem from the finite term (in its ① expansion) of the optimal solution of the unconstrained problem. Moreover, Lagrangian duals are also automatically and at no additional cost obtained just considering the $①^{-1}$ grossdigits in their expansion in term ①.

While this chapter only concentrates the attention on the use of ① to define novel exact penalty functions for constrained optimization problems, it must be noted that the use of ① has been beneficial in many other areas in optimization. Already in [8], the authors demonstrated how the classical simplex method for linear programming can be modified, using ① to overcome the difficulties due to degenerate steps. Along this line of research, more recently in [4], the authors proposed the Infinitely-Big-M method, a re–visitation of the Big–M method for the Infinity Computer. Various different optimization problems have been successfully tackled using this new methodology: multiobjective optimization problems [3, 5, 6, 21], the use of negative curvature directions in large-scale unconstrained optimization [11, 12], variable metric methods in nonsmooth optimization [16]. Recently, this computational methodology has also been also utilized in the field of Machine Learning

allowing to construct new spherical separations for classification problems
[2], and novel sparse Support Vector Machines [10]. Furthermore, the use
of ① has given rise to a variety of applications in several fields of pure and
applied mathematics, providing new and alternative approaches. Here we
only mention numerical differentiation [26], ODE [1, 20, 31], hyperbolic
geometry [23], infinite series and the Riemann zeta function [25, 27], biol-
ogy [28], and cellular automata [7].

We briefly describe our notation now. All vectors are column vectors and
will be indicated with lower case Latin letter (x, y, ...). Subscripts indicate
components of a vector, while superscripts are used to identify different vec-
tors. Matrices will be indicated with upper case roman letter (A, B, ...). The
set of real numbers and the set of nonnegative real numbers will be denoted
by $\mathbb{R}$ and $\mathbb{R}_+$ respectively. The space of the n–dimensional vectors with real
components will be indicated by $\mathbb{R}^n$. Superscript ^{T} indicates transpose. The
scalar product of two vectors x and y in $\mathbb{R}^n$ will be denoted by $x^T y$. The
norm of a vector x will be indicated by $\|x\|$. The space of the $m \times n$ matrices
with real components will be indicated by $\mathbb{R}^{m \times n}$. Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, the
gradient $\nabla f(x)$ of $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ is a column vector with
$[\nabla f(x)]_j = \frac{\partial f(x)}{\partial x_j}$.

For what is necessary in this chapter, in this new positional numeral system
with base ① a value C is expressed as

$$C = C^{(1)}① + C^{(0)} + C^{(-1)}①^{-1} C^{(-2)}①^{-2} + \cdots$$

Here and throughout the symbols := and =: denote definition of the term on
the left and the right sides of each symbol, respectively.

2 Exact Penalty Methods

In this section we will utilize the novel approach to infinite and infinitesimal
numbers proposed by Sergeyev[1] to construct exact differentiable penalty
functions for nonlinear optimization problems.

Consider the constrained optimization problem

---

[1] We refer the reader to Chap. 1 for an in–depth description of this new applied approach
to infinite and infinitesimal quantities and the arithmetics of infinity.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & c_i(x) \leq 0 \qquad i = 1, \ldots, m \\ & c_i(x) = 0 \qquad i = m+1, \ldots, m+h \end{aligned} \qquad (1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function and $c_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m+h$ are the constraints defining the feasible region X

$$X := \left\{ x \in \mathbb{R}^n : c_i(x) \leq 0, \ i = 1, \ldots, m \text{ and } c_i(x) = 0, \ i = m+1, \ldots, m+h \right\}.$$

For simplicity, we will assume that all the functions are at least twice continuously differentiable.

Let $\bar{x}$ be a feasible point for the above problem. The set of active constraints at $\bar{x}$ is defined as

$$\mathcal{A}(\bar{x}) := \left\{ i : 1 \leq i \leq m, c_i(\bar{x}) = 0 \right\} \cup \left\{ m+1, \ldots, m+h \right\}.$$

In nonlinear optimization a key role is played by the Constraint Qualification conditions that ensure that the tangent cone and the cone of linearized feasible directions coincide and allow to express necessary and sufficient optimality conditions in terms of the well known Karush–Kuhn–Tucker conditions.[2] For reader's easiness we recall here the fundamental Linear Independence Constraint Qualification that will be heavily utilized in this form or in a modified form in this chapter.

Definition 1 Linear independence constraint qualification (LICQ) condition is said to hold true at $\bar{x} \in X$ if the gradients of the active constraints at $\bar{x}$ are linearly independent.

Note that weaker Constraint Qualification conditions can be imposed [32]. In [34] various Constraint Qualification conditions are stated and categorized into four levels by their relative strengths from weakest (less stringent) to strongest (more stringent, but easier to check).

The Lagrangian function associated to Problem (1) is given by

$$L(x, \lambda) := f(x) + \sum_{i=1}^{m+h} \lambda_i c_i(x). \qquad (2)$$

Necessary and sufficient optimality conditions can be written in terms of the Lagrangian function. If $x^* \in X$ is a local minimum of Problem (1) at which LICQ condition holds true, then, there exist a vector λ^* such that the pair (x^*, λ^*) is a Karush–Kuhn–Tucker point.

---

[2] For further details we refer the reader to Chap. 2 and references therein.

Different algorithms have been proposed in literature for finding a local minimum of Problem (1). Among the most effective methods, penalty methods play a crucial role, obtaining the solution of the constrained problem by solving a single or a sequence of unconstrained optimization problems.

Let $\delta : \mathbb{R}^n \to \mathbb{R}_+$ be a function, when possible continuously differentiable, such that

$$\delta(x) \begin{cases} = 0 \text{ if } x \in X \\ > 0 \text{ otherwise.} \end{cases}$$

Then the constrained optimization problem (1) can be replaced by the following unconstrained problem

$$\min_x \psi(x, \mu). \tag{3}$$

where

$$\psi(x, \mu) := f(x) + \mu \delta(x), \tag{4}$$

and μ is a positive real number. Different choices for the function $\delta(x)$ conduct to different penalty methods. A convenient, highly utilized, choice is

$$\delta(x) = \frac{1}{2} \sum_{i=1}^{m} \max\{c_i(x), 0\}^2 + \frac{1}{2} \sum_{i=m+1}^{m+h} (c_i(x))^2 \tag{5}$$

which is differentiable but not twice differentiable. Therefore, this choice for $\delta(x)$ does not allow to utilize some of the most effective algorithms available for unconstrained optimization.

One key issue in exterior penalty methods is exactness. Roughly speaking, the penalty function is exact if, for some finite value of the parameter μ, a local (global) minimum of it corresponds to a local (global) minimum point of the constrained problem. As noted in [14] this characterization is satisfactory only when both the constrained problem and the penalty function are convex, while in the non–convex case a more precise definition of exactness is necessary [13, 14].

Unfortunately, for the penalty function (5) this exactness property does not hold, not even in the case of only equality constraints, that is when $m = 0$. In [15, p. 279] a simple 1–dimensional counter-example is reported, showing that the solution of the constraint problem is only obtained when $\mu \to +\infty$. Exact penalty functions can be constructed using 1-norm instead of the 2-norm utilized in (5) to penalize the violation of the constraints. However, in these cases the resulting function is nondifferentiable, and ad-hoc methods must be utilized for which convergence is often quite slow. Furthermore, exact differentiable penalty functions can be constructed [13] by introducing

---

Algorithm 1: Generic Sequential Minimization Algorithm

---

1 **Choose** $x^0 \in \mathbb{R}^n$. **Let** $\{\mu_k\}$ be a monotonically increasing sequence of positive real values, **Set** $k = 0$, $x^{\mu_0} = x^0$;

2 **while** $\delta(x^{\mu_k}) > \epsilon$ **do**

3   **Set** $k = k + 1$;

4   **Compute** x^{μ_k} an optimal solution of the unconstrained differentiable problem

$$\min_x \psi(x, \mu_k).$$

5 **end**

---

in the objective function additional terms related to first order optimality conditions, thus making the objective function more complicate to manage.

Sequential penalty methods require to solve a sequence of minimization problems with increasing values of the parameter μ as shown in Algorithm 1.

In the algorithm above, the point x^{μ_k} obtained at iteration k can be used as starting point for the minimization problem at iteration $k + 1$. Note that Step 4 cannot be, in general, completed in a finite number of steps and, hence, the algorithm must be modified requiring to calculate, at each iteration, only an approximation of the optimal solution.

In [8] a novel exact differentiable penalty method is introduced using the numeral grossone. In the next three sections, the equality constraints, the general equality and inequality constraints and quadratics case will be discussed. Finally, a simple new non–monotone algorithmic scheme is also proposed for the solution of penalty functions based on ①.

3   Equality Constraints Case

Consider the constrained optimization problem with only equality constraints (that is $m = 0$):

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0 \qquad i = 1, \dots, h. \end{aligned} \tag{6}$$

Let

$$\delta(x) = \frac{1}{2} \sum_{i=1}^{h} (c_i(x))^2$$

and let $\psi(x, \mu)$ be given by (4). In this case, it is possible to show [15, Theorem 12.1.1] that for the sequence constructed by Algorithm 1

1. $\{\psi(x^{\mu_k}, \mu_k)\}$ is monotonically non–decreasing,
2. $\{\delta(x^{\mu_k})\}$ is monotonically non–increasing,
3. $\{f(x^{\mu_k})\}$ is monotonically non–decreasing.

Moreover, $\lim_k c_i(x^{\mu_k}) = 0, i = 1, \ldots, h$ and each accumulation point x^* of the sequence $\{x^{\mu_k}\}_k$ solves Problem (6).

For the equality constrained case, the penalty function proposed in [8] is defined as follows:

$$\min_x \quad \psi(x) := f(x) + \frac{①}{2} \sum_{i=1}^h (c_i(x))^2 . \tag{7}$$

Under the hypothesis that any generic point x, the function value $f(x)$, and the generic constraint $c_i(x)$ as well as the gradient $\nabla f(x)$ and $\nabla c_i(x)$ have a finite part (i.e., grossdigits corresponding to grosspower 0) and only infinitesimal terms, it is possible to show that [8, Theorem 3.3] if x^* is a stationary point for (7) then $(x^{*(0)}, \lambda^*)$ is a KKT point for (6), where $x^{*(0)}$ is the finite term in the representation of x^*, and for $i = 1, \ldots, h, \lambda_i^* = c_i^{(-1)}(x^{*(0)})$ where $c_i^{(-1)}(x^*)$ is the grossdigit corresponding to the grosspower -1 in the representation of $c_i(x^*)$.

The above result strongly relies on the fact that at x^* the LICQ condition is satisfied. The proof is based on the observation that grossdigits "do not mix", that is, they are kept well separated in the computations. Therefore, setting to 0 a grossnumber is equivalent to set to zero all grossdigits in its representation.

The fundamental aspect of this result is that, by solving the unconstrained optimization problem (where the objective function is twice continuously differentiable), an optimal solution of the constrained problem is obtained. Moreover, the multipliers associated to the equality constraints are automatically recovered at no additional cost, just from the representation of $c_i(x^*)$ in terms of powers of ①.

In [9] two simple examples are discussed, showing the importance of constraint qualification conditions. Here, we propose a novel example, similar in spirit to the first example discussed in [9]. The problem is originally studied in [15, pp. 279–280] to show the effectiveness and weakness of sequential penalty method.

Consider the problem

Fig. 1 Feasible region
and optimal solution for
Problem (8)
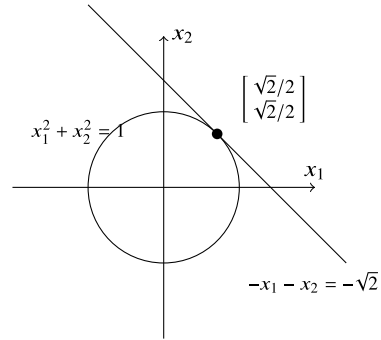
$$\begin{array}{rl} \min_{x} & -x_1 - x_2 \\ \text{subject to } & x_1^2 + x_2^2 - 1 = 0. \end{array} \tag{8}$$

The feasible region and the optimal solution for this problem is shown in
Fig. 1.

The Lagrangian function is given by

$$L(x, \lambda) = -x_1 - x_2 + \lambda \left(x_1^2 + x_2^2 - 1 \right)$$

and the optimal solution is $x^* = \begin{bmatrix} \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \end{bmatrix}$. Moreover, it is not difficult to show

that the pair $\left(x^*, \lambda^* = \dfrac{1}{\sqrt{2}} \right)$ satisfies the KKT conditions.

In Fig. 2 the contour plots for the function $\psi(x, \mu)$ for different values of
μ are shown.

The penalty function we construct is:

$$\psi(x, \textcircled{1}) := -x_1 - x_2 + \frac{\textcircled{1}}{2} \left(x_1^2 + x_2^2 - 1 \right)^2. \tag{9}$$

The First–Order Optimality Conditions $\nabla \psi(x, \textcircled{1}) = 0$ are:

$$\begin{cases} -1 + 2\textcircled{1}x_1 \left(x_1^2 + x_2^2 - 1 \right) = 0, \\ -1 + 2\textcircled{1}x_2 \left(x_1^2 + x_2^2 - 1 \right) = 0. \end{cases} \tag{10}$$

By symmetry,

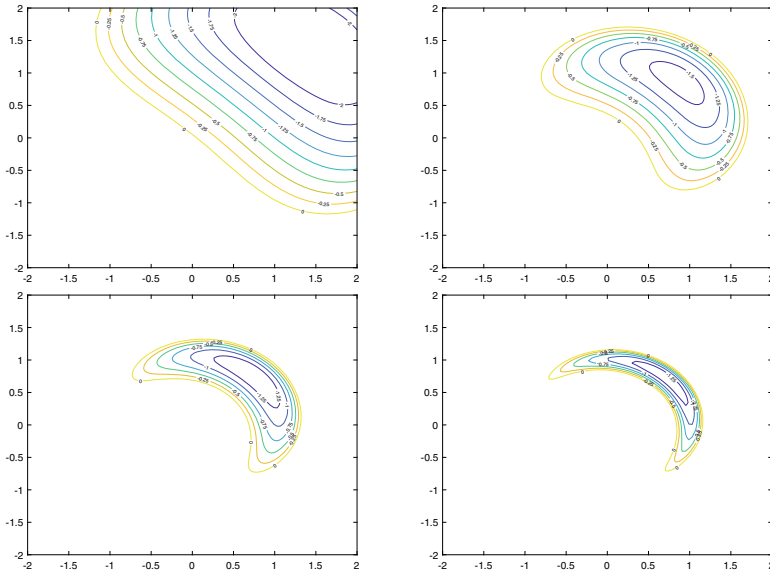$$x_1^* = x_2^* = R = R^{(0)} + \textcircled{1}^{-1} R^{(-1)} + \textcircled{1}^{-2} R^{(-2)} + \dots$$

Fig. 2 Contour plots of $\psi(x, \mu)$ for $\mu = 0.1, 1, 5, 20$

and from (10) we have that

$$-1 + 4①R\left[R^2 - \frac{1}{2}\right] = 0. \tag{11}$$

From (11), by equating the term of order ① to 0, we obtain:

$$4R^{(0)}\left[\left(R^{(0)}\right)^2 - \frac{1}{2}\right] = 0$$

from which either $R^{(0)} = \frac{1}{\sqrt{2}}$ or $R^{(0)} = -\frac{1}{\sqrt{2}}$ or $R^{(0)} = 0$. The first choice for $R^{(0)}$ corresponds to a minimum of the unconstrained problem (and also for the constrained problem (8)), while the second corresponds to a maximum. Finally, the third choice of $R^{(0)}$ corresponds to a spurious solution whose presence is due to the fact that, for this point $\hat{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, LICQ are not satisfied. In fact $\nabla h(\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Fixing now $R^{(0)} = \frac{1}{\sqrt{2}}$, and equating to 0 the finite terms in (11), i.e., the term corresponding to $①^0$, we obtain:

$$-1 + 4\frac{1}{\sqrt{2}}2\frac{1}{\sqrt{2}}R^{(-1)} = 0,$$

from which $R^{(-1)} = \frac{1}{4}$ and hence

$$x_1^* = x_2^* = R = \frac{1}{\sqrt{2}} + \frac{1}{4}①^{-1} + \cdots$$

where all remaining terms are infinitesimal of higher order.

Moreover,

$$\left(x_1^*\right)^2 + \left(x_2^*\right)^2 - 1 = 2R^2 - 1 = 2\left[\frac{1}{2} + \frac{2}{4\sqrt{2}}①^{-1} + \frac{1}{16}①^{-2} + \cdots\right] - 1$$

$$= \frac{1}{\sqrt{2}}①^{-1} + \cdots$$

where again we neglect infinitesimal of higher order than $①^{-1}$.

As expected from the theory, the $①^{-1}$ terms in the representation of the (unique) constraint provides automatically, and at no additional costs, the value of the Lagrangian multiplier.

Here we want to stress the importance of Constraint Qualification conditions that, when not satisfied, could bring to spurious results, as shown by this example.

This situation is even better clarified in the second example from [9]:

$$\min_{x} \quad x_1 + x_2 \tag{12}$$
$$\text{subject to } \left(x_1^2 + x_2^2 - 2\right)^2 = 0.$$

Here the objective function and the feasible region are the same as in the first example in [9] and the optimal solution is $x^* = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ with Lagrangian multiplier $\lambda^* = \frac{1}{2}$. However, in this case

$$\nabla c_1(x) = \begin{bmatrix} 2x_1\left(x_1^2 + x_2^2 - 2\right) \\ 2x_1\left(x_1^2 + x_2^2 - 2\right) \end{bmatrix}$$

which, calculated at the optimal solution, gives:

$$\nabla c_1\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Therefore, the LICQ condition is not satisfied and the Karush–Kuhn–Tucker conditions

$$\begin{cases} 1 - 4\lambda x_1 \left(x_1^2 + x_2^2 - 2\right) = 0 \\ 1 - 4\lambda x_2 \left(x_1^2 + x_2^2 - 2\right) = 0 \\ \left(x_1^2 + x_2^2 - 2\right)^2 = 0 \end{cases} \tag{13}$$

have no solution. In this case (see [9]) the optimal solution $x^* = \begin{bmatrix} c - 1 \\ -1 \end{bmatrix}$ of Problem (12) cannot be recovered from the exact penalty function

$$\psi(x, ①) = x_1 + x_2 + \frac{①}{2} \left(x_1^2 + x_2^2 - 2\right)^4$$

using first order optimality conditions.

4 Equality and Inequality Constraints Case

Consider now the more general nonlinear optimization problem (1) that includes equality and inequality constraints. In this case, the exact penalty function proposed in [8] is given by

$$\psi(x, ①) = f(x) + \frac{①}{2} \sum_{i=1}^{m} \max\{0, c_i(x)\}^2 + \frac{①}{2} \sum_{i=m+1}^{m+h} c_i(x)^2. \tag{14}$$

In order to derive the correspondence between stationary points of (14) and KKT points for Problem (1) a Modified LICQ condition is introduced in [8].

Definition 2 Let $\bar{x} \in \mathbb{R}^n$. The Modified LICQ (MLICQ) condition is said to hold true at $\bar{x}$ if the vectors

$$\left\{\nabla c_i(\bar{x}), i : 1 \leq i \leq m \text{ and } c_i(\bar{x}) \geq 0\right\} \cup \left\{\nabla c_i(\bar{x}), i = m + 1, \ldots, m + h\right\}$$

are linearly independent.

If the above conditions are satisfied, and x^* is a stationary point for the unconstrained problem

$$\min_x \psi(x, ①)$$

then, it possible, again, to show [8, Theorem 3.4] that the pair $\left(x^{*(0)}, \lambda^*\right)$ is a KKT point for Problem (1), where $x^{*(0)}$ is the finite term in the representation

of x^*, and for $i = 1, \ldots, m + h$, $\lambda_i^* = c_i^{(-1)}(x^{*(0)})$ where $c_i^{(-1)}(x^*)$ is the grossdigit corresponding to the grosspower -1 in the representation of $c_i(x^*)$.

5 Quadratic Case

In this section we apply the new exact penalty function to the quadratic problem

$$
\begin{aligned}
\min_{x} \quad & \tfrac{1}{2}x^T M x + q^T x \\
\text{subject to} \quad & Ax = b \\
& x \geq 0
\end{aligned}
\tag{15}
$$

where $M \in \mathbb{R}^{n \times n}$ is positive definite, $A \in \mathbb{R}^{m \times n}$ with rank$(A) = m$, $q \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$. We assume that the feasible region is not empty.

For this linearly constrained problem, Constraint Qualification conditions are always satisfied and the Karush–Kuhn–Tucker conditions [22] are:

$$
\begin{aligned}
Mx + q - A^T u - v &= 0, \\
Ax - b &= 0, \\
x &\geq 0, \\
v &\geq 0, \\
x^T v &= 0.
\end{aligned}
\tag{16}
$$

For this quadratic problem, the new exact penalty function using ① is given by:

$$
\psi(x, ①) := \frac{1}{2}x^T M x + q^T x + \frac{①}{2}\|Ax - b\|_2^2 + \frac{①}{2}\|\max\{0, -x\}\|_2^2
\tag{17}
$$

and the corresponding unconstrained problem is:

$$
\min_{x} \psi(x, ①).
\tag{18}
$$

The first order optimality conditions can be written as follows

$$
0 = \nabla \psi(x, ①) = Mx + q + ① A^T (Ax - b) - ① \max\{0, -x\}.
\tag{19}
$$

Lemma 1 in [9] shows that the function

$$
\delta(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\|\max\{0, -x\}\|_2^2
$$

is convex, and $\nabla\delta(x) = 0$ if and only if $Ax = b$ and $x \geq 0$.

Based on this lemma it is possible to show that, if x^* is a stationary point for the unconstrained problem (18), then $\left(x^{*(0)}, \pi^*, \mu^*\right)$ is a Karush–Kuhn–Tucker points for Problem (15), where again $x^{*(0)}$ is the finite term in the representation of x^*, and, taking into account the linearity of the original constraints,

- $\pi^* = Ax^{*(-1)} + b^{(-1)}$, where $x^{*(-1)}$ (resp. $b^{(-1)}$) is the grossdigit corresponding to the grossterm $①^{-1}$ in the representation of x^* (resp. b), and
- $\mu_j^* = \max\{0, -x^{*(-1)}\}$.

Once again, the proof is based on the fact that during the computation grossdigits do not mix. Then, the results follow from (19)

- by setting first to 0 the terms with grosspower $①$, in this way the feasibility of $x^{*(0)}$ is provided (second and third Karush–Kuhn–Tucker conditions in (16)),
- equating to zero the finite terms (that is the terms with grosspower $①^0$) thus obtaining the first, fourth and last Karush–Kuhn–Tucker conditions in (16).

6 A General Scheme for the New Exact Penalty Function

In the previous sections, we have shown how the solution of the constrained minimization problem (1) can be obtained by solving an unconstrained minimization problem that uses $①$. Then, any standard optimization algorithm can be utilized to obtain a stationary point for these problems from which the solution of the constrained problems as well the multipliers can be easily obtained.

In this section a simple novel general algorithmic scheme is proposed to solve the unconstrained minimization problem arising from penalizing the constraints using $①$, as constructed in the previous sections. The algorithms belongs to the class of non–monotone descend algorithms [17–19, 33, 35], and does not require that the new calculated points be necessary better than the current one. This property is necessary since for a descent algorithm, when applied to determine the minimum of $\psi(x, ①)$, once a feasible point is obtained, then all remaining points generated by the algorithm should remain feasible. This, in general, could be quite complicate to ensure in practice.

For simplicity, consider the generic minimization problem:

$$\min_x f(x)$$

where f is continuously differentiable and

$$f(x) = ① f^{(1)}(x) + f^{(0)}(x) + ①^{-1} f^{(-1)}(x) + \dots \tag{20}$$

$$\nabla f(x) = ① \nabla f^{(1)}(x) + \nabla f^{(0)}(x) + ①^{-1} \nabla f^{(-1)}(x) + \dots \tag{21}$$

The proposed algorithm utilizes (as customary in non–monotone algorithms) two sequences $\{t_k\}_k$ and $\{l_k\}_k$ of positive integers such that

$$t_0 = 0, \quad t_{k+1} \le \max\{t_k + 1, T\},$$

$$l_0 = 0, \quad l_{k+1} \le \max\{l_k + 1, L\},$$

where T and L are two fixed positive integers.

At the generic iteration k, having x^k, it is necessary first to verify if the stopping criterion

$$\nabla f^{(1)}(x^k) = 0 \text{ and } \nabla f^{(0)}(x^k) = 0$$

is satisfied. Otherwise, the next iterate x^{k+1} is calculated in the following way:

• if $\nabla f^{(1)}(x^k) \ne 0$, choose x^{k+1} such that

$$f^{(1)}(x^{k+1}) \le f^{(1)}(x^k) + \sigma \left(\left\| \nabla f^{(1)}(x^k) \right\| \right),$$

and
$$f^{(0)}(x^{k+1}) \le \max_{0 \le j \le l_k} f^{(0)}(x^{k-j}) + \sigma \left(\left\| \nabla f^{(0)}(x^k) \right\| \right);$$

• If $\nabla f^{(1)}(x^k) = 0$, choose x^{k+1} such that

$$f^{(0)}(x^{k+1}) \le f^{(0)}(x^k) + \sigma \left(\left\| \nabla f^{(0)}(x^k) \right\| \right),$$

$$f^{(1)}(x^{k+1}) \le \max_{0 \le j \le t_k} f^{(1)}(x^{k-j})$$

where $\sigma(.)$ is a forcing function [24].

In other words, when $\nabla f^{(1)}(x^k) \ne 0$ the infinite term cannot grow more than a quantity that depends on the norm of $\nabla f^{(1)}(x^k)$. For the finite term, instead, the new value $f^{(0)}(x^{k+1})$ cannot be bigger than the worst value of $f^{(0)}(x^{k-j})$ at l_k previous steps plus a quantity that depends on the norm of $\nabla f^{(0)}(x^k)$.

Instead, when $\nabla f^{(1)}(x^k) = 0$, the finite term $f^{(0)}(x^{k+1})$ cannot grow bigger than a quantity that depends on the norm of $\nabla f^{(0)}(x^k)$, and the infinite term $f^{(1)}(x^{k+1})$ must be better than the worst of the previous t_k values $f^{(1)}(x^{k-j})$.

In order to demonstrate convergence of the above scheme, we need to consider two different cases.

Case 1: there exists $\bar{k}$ such that $\nabla f^{(1)}(x^k) = 0$, for all $k \geq \bar{k}$. Then

$$f^{(1)}(x^{k+1}) \leq \max_{0 \leq j \leq t_k} f^{(1)}(x^{k-j}), \qquad k \geq \bar{k}.$$

Therefore, in this case:

$$\max_{0 \leq i \leq T} f^{(1)}(x^{\bar{k}+Tj+i}) \leq \max_{0 \leq i \leq T} f^{(1)}(x^{\bar{k}+T(j-1)+i})$$

and the sequence $\left\{ \max_{0 \leq i \leq T} f^{(1)}(x^{\bar{k}+Tj+i}) \right\}_j$ is monotonically decreasing.

Moreover,

$$f^{(0)}(x^{k+1}) \leq f^{(0)}(x^k) + \sigma \left(\left\| \nabla f^{(0)}(x^k) \right\| \right), \qquad k \geq \bar{k}.$$

Assuming that the level sets for $f^{(1)}(x^0)$ and $f^{(0)}(x^0)$ are compact sets, the sequence $\left\{ \max_{0 \leq i \leq T} f^{(1)}(x^{\bar{k}+Tj+i}) \right\}_j$ has at least one accumulation point x^* and any accumulation point (according to the second condition) satisfies $\nabla f^{(0)}(x^*) = 0$ in addition to $\nabla f^{(1)}(x^*) = 0$.

Case 2: there exists a subsequence j_k such that $\nabla f^{(1)}(x^{j_k}) \neq 0$. In this case we have that

$$f^{(1)}(x^{j_k+1}) \leq f^{(1)}(x^{j_k}) + \sigma \left(\left\| \nabla f^{(1)}(x^{j_k}) \right\| \right).$$

Again, taking into account that it is possible that for some index i bigger than j_k, $\nabla f^{(1)}(x^i) = 0$

$$\max_{0 \leq i \leq M} f^{(1)}(x^{j_k+Tj+i}) \leq \max_{0 \leq i \leq M} f^{(1)}(x^{j_k+T(j-1)+i}) + \sigma \left(\left\| \nabla f^{(1)}(x^{j_k}) \right\| \right)$$

and hence, assuming again that the level sets for $f^{(1)}(x^0)$ and $f^{(0)}(x^0)$ are compact sets, we have that $\nabla f^{(1)}(x^{j_k})$ goes to 0.

Moreover,

$$\max_{0 \leq i \leq L} f^{(0)}(x^{j_k+Lj+i}) \leq \max_{0 \leq i \leq L} f^{(0)}(x^{j_k+L(j-1)+i}) + \sigma \left(\left\| \nabla f^{(0)}(x^{j_k}) \right\| \right)$$

and hence also $\nabla f^{(0)}(x^{j_k})$ goes to 0.

7 Conclusions

Penalty methods are an important and widely studied class of algorithms in nonlinear optimization. Using penalty functions the solution of the original constrained optimization problem could be obtained by solving an unconstrained problem. The main issues with this class of methods are exactness and differentiability. In this chapter we presented some recent development on penalty functions based on the use of ①. The solution of the proposed unconstrained problem provides not only the solution of the original problem but also the Lagrangian dual variables associated to the constraints at no additional cost, from the expansion of the constraints in terms of ①. Some simple examples are also reported, showing the effectiveness of the method. Finally a general non–monotone scheme is presented for the minimization of functions that include ① grossterms.

References

1. Amodio, P., Iavernaro, F., Mazzia, F., Mukhametzhanov, M.S., Sergeyev, Y.D.: A generalized Taylor method of order three for the solution of initial value problems in standard and infinity floating-point arithmetic. Math. Comput. Simul. **141**, 24–39 (2017)
2. Astorino, A., Fuduli, A.: Spherical separation with infinitely far center. Soft. Comput. **24**(23), 17751–17759 (2020)
3. Cococcioni, M., Cudazzo, A., Pappalardo, M., Sergeyev, Y.D.: Solving the lexicographic multi-objective mixed-integer linear programming problem using branch-and-bound and grossone methodology. Commun. Nonlinear Sci. Numer. Simul. **84**, 105177 (2020)
4. Cococcioni, M., Fiaschi, L.: The Big-M method with the numerical infinite M. Optim. Lett. **15**, 2455–2468 (2021)
5. Cococcioni, M., Pappalardo, M., Sergeyev, Y.D.: Towards lexicographic multi-objective linear programming using grossone methodology. In: Y.D. Sergeyev, D.E. Kvasov, F. Dell'Accio, M.S. Mukhametzhanov (eds.) Proceedings of the 2nd International Conference. Numerical Computations: Theory and Algorithms, vol. 1776, p. 090040. AIP Publishing, New York (2016)
6. Cococcioni, M., Pappalardo, M., Sergeyev, Y.D.: Lexicographic multi-objective linear programming using grossone methodology: theory and algorithm. Appl. Math. Comput. **318**, 298–311 (2018)

7. D'Alotto, L.: Cellular automata using infinite computations. Appl. Math. Comput. **218**(16), 8077–8082 (2012)
8. De Cosmis, S., De Leone, R.: The use of grossone in mathematical programming and operations research. Appl. Math. Comput. **218**(16), 8029–8038 (2012)
9. De Leone, R.: Nonlinear programming and grossone: quadratic programming and the role of constraint qualifications. Appl. Math. Comput. **318**, 290–297 (2018)
10. De Leone, R., Egidi, N., Fatone, L.: The use of grossone in elastic net regularization and sparse support vector machines. Soft. Comput. **24**(23), 17669–17677 (2020)
11. De Leone, R., Fasano, G., Roma, M., Sergeyev, Y.D.: How grossone can be helpful to iteratively compute negative curvature directions. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **11353 LNCS**, pp. 180–183 (2019)
12. De Leone, R., Fasano, G., Sergeyev, Y.D.: Planar methods and grossone for the conjugate gradient breakdown in nonlinear programming. Comput. Optim. Appl. **71**(1), 73–93 (2018)
13. Di Pillo, G., Grippo, L.: An exact penalty method with global convergence properties for nonlinear programming problems. Math. Program. **36**, 1–18 (1986)
14. Di Pillo, G., Grippo, L.: Exact penalty functions in constrained optimization. SIAM J. Control Optim. **27**(6), 1333–1360 (1989)
15. Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley (1990)
16. Gaudioso, M., Giallombardo, G., Mukhametzhanov, M.S.: Numerical infinitesimals in a variable metric method for convex nonsmooth optimization. Appl. Math. Comput. **318**, 312–320 (2018)
17. Grippo, L., Lampariello, F., Lucidi, S.: A truncated Newton method with nonmonotone line search for unconstrained optimization. J. Optim. Theory Appl. **60**(3), 401–419 (1989)
18. Grippo, L., Sciandrone, M.: Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. Comput. Optim. Appl. **23**(2), 143–169 (2002)
19. Grippo, L., Sciandrone, M.: Nonmonotone derivative-free methods for nonlinear equations. Comput. Optim. Appl. **37**(3), 297–328 (2007)
20. Iavernaro, F., Mazzia, F., Mukhametzhanov, M.S., Sergeyev, Y.D.: Computation of higher order Lie derivatives on the Infinity Computer. J. Comput. Appl. Math. **383**(113135) (2021)
21. Lai, L., Fiaschi, L., Cococcioni, M.: Solving mixed Pareto-lexicographic multi-objective optimization problems: the case of priority chains. Swarm Evol. Comput. **55** (2020)
22. Mangasarian, O.L.: Nonlinear programming. McGraw-Hill Series in Systems Science. McGraw-Hill, New York (1969)
23. Margenstern, M.: An application of grossone to the study of a family of tilings of the hyperbolic plane. Appl. Math. Comput. **218**(16), 8005–8018 (2012)
24. Ortega, J.M., Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables. Academic Press (1970)
25. Sergeyev, Y.D.: Numerical point of view on Calculus for functions assuming finite, infinite, and infinitesimal values over finite, infinite, and infinitesimal domains. Nonlinear Anal. Ser. A: Theory, Methods Appl. **71(12)**, e1688–e1707 (2009)
26. Sergeyev, Y.D.: Higher order numerical differentiation on the infinity computer. Optim. Lett. **5**(4), 575–585 (2011)
27. Sergeyev, Y.D.: On accuracy of mathematical languages used to deal with the Riemann zeta function and the Dirichlet eta function. p-Adic Numbers, Ultrametric Anal. Appl. **3(2)**, 129–148 (2011)

28. Sergeyev, Y.D.: Using blinking fractals for mathematical modelling of processes of growth in biological systems. Informatica **22**(4), 559–576 (2011)
29. Sergeyev, Y.D.: Numerical infinities and infinitesimals: methodology, applications, and repercussions on two Hilbert problems. EMS Surv. Math. Sci. **4**(2), 219–320 (2017)
30. Sergeyev, Y.D.: Independence of the grossone-based infinity methodology from non-standard analysis and comments upon logical fallacies in some texts asserting the opposite. Found. Sci. **24**(1), 153–170 (2019)
31. Sergeyev, Y.D., Mukhametzhanov, M.S., Mazzia, F., Iavernaro, F., Amodio, P.: Numerical methods for solving initial value problems on the infinity computer. Int. J. Unconv. Comput. **12**(1), 3–23 (2016)
32. Solodov, M.V.: Constraint qualifications. In: Wiley Encyclopedia of Operations Research and Management Science. Wiley Online Library (2010)
33. Sun, W., Han, J., Sun, J.: Global convergence of nonmonotone descent methods for unconstrained optimization problems. J. Comput. Appl. Math. **146**(1), 89–98 (2002)
34. Wang, Z., Fang, S.C., Xing, W.: On constraint qualifications: motivation, design and inter-relations. J. Ind. Manag. Optim. **9**, 983–1001 (2013)
35. Zhang, H.C., Hager, W.W.: A nonmonotone line search technique and its application to unconstrained optimization. SIAM J. Optim. **14**(4), 1043–1056 (2004)