Theoretical aspects of adopting exact penalty elements within sequential methods for nonlinear programming*

A. A. Ribeiro[†] M. Sachine[‡] S. A. Santos[§]

July 18, 2013

Abstract

In the context of sequential methods for solving general nonlinear programming problems, it is usual to work with augmented subproblems instead of the original ones, tackled by the ℓ_1-penalty function together with the shortcut usage of a convenient penalty parameter. This paper addresses the theoretical reasoning behind handling the original subproblems by such an augmentation strategy, by means of the differentiable reformulation of the ℓ_1-penalized problem. The convergence properties of related sequences of problems are analyzed. Furthermore, examples that elucidate the interrelations among the obtained results are presented.

Keywords: nonlinear programming; exact penalty function; smooth reformulation; feasibility; KKT conditions; Mangasarian-Fromovitz constraint qualification.

*This work was partially supported by FAPESP (10/09773-4, 13/05475-7, 13/07375-0), CNPq (304032/2010-7) and PRONEX-CNPq/FAPERJ (E-26/111.449/2010-APQ1).

[†]Department of Mathematics, Federal University of Paraná, CP 19081, 81531-980, Curitiba, PR, Brazil. E-mail: ademir.ribeiro@ufpr.br

[‡]Department of Mathematics, Federal University of Paraná, CP 19081, 81531-980, Curitiba, PR, Brazil. E-mail: mael@ufpr.br

[§]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, 13083-859, Campinas, SP, Brazil. E-mail: sandra@ime.unicamp.br

1 Introduction

This work focuses on theoretical properties of the smooth reformulation for the ℓ_1-penalty function. The target is the inequality constrained nonlinear programming problem with simple bounds, referred to as the *original problem* (OP). Basically resting upon the continuous differentiability of all the functions that define the OP, we intend to infer about the feasibility and the regularity features of the constraints of such a problem by means of a related, and always feasible, *augmented problem* (AP). As detailed in the sequel, it turns out that every feasible point of the AP fulfils a constraint qualification.

On one hand, the augmentation has been conveniently adopted in the context of practical methods that approximate the feasible set, such as the Sequential Quadratic Programming (SQP) [5, 13], in which the consistency of the subproblems is a crucial issue for the generation of the iterative scheme. Among the structural optimization community, the augmentation is also largely employed to prevent a breakdown of the optimization process. Badly chosen starting points might generate infeasible subproblems either for the Convex Linearization Method (CONLIN) [11, 12] or for the Method of the Moving Asymptotes (MMA) [19] and its extensions and modifications [14, 20, 21, 22], to name a few.

On the other hand, the augmentation is a powerful device for deducing intrinsic features of the OP as well. Firstly, whenever a local minimizer of the AP is obtained within a finite, or a nonnull, threshold for the penalty parameter, as depending on the formulation it could be either driven to infinity or to zero, the exact nature of such a parameter allows the achievement of a local minimizer of the OP. In this context, the existence of a local minimum for the related exact penalty function may be interpreted as a constraint qualification, so that the Karush-Kuhn-Tucker (KKT) conditions are fulfilled at the local minima of the OP. This will be resumed later. Secondly, the AP may be used to locally detect infeasibility of the OP. This idea has been recently exploited in [5], where the authors have devised an SQP algorithm that achieves superlinear convergence to infeasible stationary points that satisfy regularity, strict complementarity and second-order sufficiency conditions. In a slightly distinct context, [17] is another recent reference concerned with fast detection of infeasibility. It is based upon an augmented Lagrangian algorithm that employs adaptive tolerances, depending on the acceptable degree of infeasibility and complementarity at each iterate, taking advantage of the fact that one does not need great accuracy in the absence of near-feasibility, as long as the overall convergence properties remain under control.

Without convexity or particular intrinsic properties (e.g. [4]), obtaining a global minimizer of a nonlinear programming problem is a challenging task. However, within a sequential method for the general problem, such as SQP, CONLIN or MMA, particular advantage might be taken from the structure of the subproblems upon consideration, so that a global solution is indeed attainable. In such a context, what matters is the consistent generation of a sequence of subproblems, circumventing any possibility of premature stopping due to failure. Our analysis applies to the pair *original problem* and the associated *augmented problem*, being the latter a useful device in the aforementioned context. Besides, a shortcut penalty strategy might work if a bound for the penalty parameter is known. This is the case of applications in which the user is able to overestimate, even roughly, how much the objective value would improve as the effect of accepting a unit increase of the right-hand side of a certain constraint of the problem, i.e. the effect of the shadow-prices are known, see, e.g. [3]. In other words, the physical meaning of the inherent Lagrange multipliers, or dual variables, might offer additional elements for computing a solution (cf. [21]).

Our aim is to analyze the features of the sequence of *solutions* generated by the ℓ_1-penalty framework, encompassing stationary points and local minimizers, and relate them to the corresponding counterparts of the OP. To preserve the potential differentiability of the functions that define the OP, a smooth reformulation is employed as the main tool for the analysis. We stress that the nondifferentiable setting is solely adopted as a theoretical resource, and our research is not computational nor experimental. Moreover, our analysis differs from the classical exact penalty approach, as no assumptions are made concerning the feasibility or the fulfilment of a constraint qualification by the OP. Indeed, the primary question that has motivated our investigation was: from a *solution* of the AP, what can be inferred about the OP?

This work is organized as follows: in Section 2 the main ingredients are stated, including notation and preliminary results. Section 3 is devoted to examine the convergence properties of the sequence of stationary points generated by the framework under consideration. Section 4 highlights particular features of the generated sequences and the associated limit points by means of two examples. The convergence results of local minimizers for the sequence of augmented problems and its potential association with the OP are studied in Section 5. Finally, concluding remarks close our text in Section 6.

2 Preliminaries

The *original problem* (OP) is established next:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X}, \end{array} \tag{1}$$

where the functions f_i, $i = 0, \ldots, m$, are continuously differentiable and

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n \mid x_j^{\min} \le x_j \le x_j^{\max},\ j = 1, \ldots, n \right\} \tag{2}$$

is a compact set, that is, $-\infty < x_j^{\min} < x_j^{\max} < +\infty$ for all $j = 1, \ldots, n$.

In case problem (1) admits feasible points that fulfil a constraint qualification, by applying any method of choice for nonlinear programming, one expects to find a stationary point. This means that such a point should satisfy the KKT conditions:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) - \nu + \zeta = 0, \tag{3a}$$

$$f_i(x) \leq 0, \quad i = 1, \ldots, m \tag{3b}$$

$$x_j^{\min} - x_j \leq 0 \quad \text{and} \quad x_j - x_j^{\max} \leq 0, \quad j = 1, \ldots, n \tag{3c}$$

$$\lambda_i \geq 0, \quad i = 1, \ldots, m \tag{3d}$$

$$\nu_j \geq 0 \quad \text{and} \quad \zeta_j \geq 0, \quad j = 1, \ldots, n \tag{3e}$$

$$\lambda_i f_i(x) = 0, \quad i = 1, \ldots, m \tag{3f}$$

$$\nu_j(x_j^{\min} - x_j) = 0 \quad \text{and} \quad \zeta_j(x_j - x_j^{\max}) = 0, \quad j = 1, \ldots, n. \tag{3g}$$

However, if the feasible set of the OP is empty, obtaining a minimizer of some infeasibility related problem may be declared a successful achievement.

A classical way [9, 10] to deal with the constrained problem (1) is to consider the ℓ_1-penalty function and the nonsmooth problem

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} c_i f_i^+(x) \\ \text{subject to} & x \in \mathcal{X}, \end{array} \tag{4}$$

where $c \in \mathbb{R}_+^m$ and the function $f^+ : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$f_i^+(x) = \max\{0, f_i(x)\}, \; i = 1, \ldots, m.$$

Alternatively to the formulation (4), one may address its differentiable counterpart

$$\begin{array}{ll} \text{minimize} & f_0(x) + c^T y \\ \text{subject to} & f_i(x) - y_i \leq 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X}, y \geq 0, \end{array} \tag{5}$$

referred to as the *augmented problem* AP(c).

Despite demanding artificial variables that increase its dimension, it is precisely due to this augmentation that problem (5) has the feature of always being feasible. Indeed, given an arbitrary $x \in \mathcal{X}$, the point $(x, f^+(x))$ is feasible for the AP(c). So, we have an easy way to select a starting point for solving this problem.

This approach is employed by [16] in the context of constraint reduction, so that the easily initialized extended problem is dealt with, instead of the original one. A worth-mentioning related strategy, known as the *big-M method* (see e.g. [3, §4.3]), is used in linear programming to initiate the simplex method with a promptly available basic feasible solution for the associated augmented problem. Another recent work that rests upon the differentiable augmented problem is [8], in the development of a penalty-interior-point algorithm for general nonlinear programming.

The next three results, proved by Svanberg in Propositions 2.2, 2.3 and 2.4 of [21], respectively, establish some properties of the augmented problem (5). The first one shows the equivalence between the solutions of the problems (4) and (5).

Lemma 2.1. *If $(\hat{x}, \hat{y})$ is a global solution of the problem (5), then $\hat{x}$ is a global solution of the problem (4). On the other hand, if $\hat{x}$ is a global solution of the problem (4), then $(\hat{x}, f^+(\hat{x}))$ is a global solution of the problem (5).*

The following lemma ensures that the augmented problem has a global minimizer.

Lemma 2.2. *There is at least one global solution to problem (5).*

The next result concerns the qualification of the constraints of the augmented problem. It was mentioned by Han and Mangasarian as a remark along the proof of Theorem 4.8 in [15].

Lemma 2.3. *Every feasible point of the problem (5) satisfies the Mangasarian-Fromovitz Constraint Qualification (MFCQ).*

The main tool for the analysis is to parameterize the augmented problem by a *sequence* of penalty parameters. First, it offers conditions for the threshold to be reached, in case it exists, so that the ℓ_1-penalty acts in its exact sense. Second, as will be shown, an unlimited increasing of the penalty parameters might indicate infeasibility of the OP or that it does not satisfy a constraint qualification. We summarize the framework underlying our analysis in the following steps:

Framework I. *Let $(c^k) \subset \mathbb{R}^m_+$ be a sequence such that*

$$c_i^k \to \infty, \quad \forall i = 1, \ldots, m.$$

For all $k \in \mathbb{N}$, obtain a solution (x^k, y^k) *of the subproblem*

$$\begin{array}{ll} \text{minimize} & f_0(x) + (c^k)^T y \\ \textit{subject to} & f_i(x) - y_i \leq 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X}, y \geq 0. \end{array}$$

In the next section we shall examine the convergence of the sequences produced by the Framework I, assuming first that the so-called *solutions* are stationary points.

3 Convergence to stationary points

In order to obtain our theoretical results, we shall consider a scalar sequence $(c_k) \subset \mathbb{R}_+$ such that $c_k \to \infty$ and define

$$c^k = c_k \mathbb{1} \in \mathbb{R}^m, \tag{6}$$

where $\mathbb{1}$ is the vector with all entries equal to 1.

Let

$$\left(x^k, y^k, \lambda^k, \mu^k, \nu^k, \zeta^k\right) \tag{7}$$

be a sequence of *stationary points* of the augmented problems AP(c^k) for which (6) holds. This means that the following relations are satisfied:

$$\nabla f_0(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k) - \nu^k + \zeta^k = 0, \tag{8a}$$

$$c^k - \lambda^k - \mu^k = 0, \tag{8b}$$

$$f_i(x^k) - y_i^k \leq 0 \quad \text{and} \quad y_i^k \geq 0, \quad i = 1, \ldots, m \tag{8c}$$

$$x_j^{\min} - x_j^k \leq 0 \quad \text{and} \quad x_j^k - x_j^{\max} \leq 0, \quad j = 1, \ldots, n \tag{8d}$$

$$\lambda_i^k \geq 0 \quad \text{and} \quad \mu_i^k \geq 0, \quad i = 1, \ldots, m \tag{8e}$$

$$\nu_j^k \geq 0 \quad \text{and} \quad \zeta_j^k \geq 0, \quad j = 1, \ldots, n \tag{8f}$$

$$\lambda_i^k (f_i(x^k) - y_i^k) = 0 \quad \text{and} \quad \mu_i^k y_i^k = 0, \quad i = 1, \ldots, m \tag{8g}$$

$$\nu_j^k (x_j^{\min} - x_j^k) = 0 \quad \text{and} \quad \zeta_j^k (x_j^k - x_j^{\max}) = 0, \quad j = 1, \ldots, n. \tag{8h}$$

Remark 1. *The Framework I may produce a KKT point of the augmented problem for which, for some finite iterate $k \in \mathbb{N}$, the vector $y^k = 0$ is obtained. In such a case, it follows directly from (8a)-(8h) that the point $(x^k, \lambda^k, \nu^k, \zeta^k)$ satisfies the relations (3a)-(3g), being thus stationary for the OP.*

In the sequel, the sequence of stationary points of the AP(c^k) is assumed to be infinite. From (8d), the sequence (x^k) lies in the compact set $\mathcal{X}$. So, without loss of generality, we assume that

$$x^k \to x^\star \in \mathcal{X}. \tag{9}$$

Nevertheless, our analysis differs from the classical exact penalty approach, because there is no guarantee, *a priori*, that the OP is either feasible or verifies a constraint qualification. Thus, the penalty parameter has to be driven to infinity in the Framework I, resembling, in some sense, the ℓ_2-penalty approach (see e.g. [7]). In the differentiable augmented reformulation (5) for the ℓ_1-penalty approach, the boundedness of the Lagrange multipliers cannot be ensured, as shown in the development that follows.

We start by proving that the artificial variable y_i^k measures the infeasibility of x^k with respect to the i-th constraint of the original problem (see also [21, Prop. 2.1]).

Lemma 3.1. *The sequence (7) is such that*

$$y_i^k = f_i^+(x^k)$$

for all $k \in \mathbb{N}$ and $i = 1, \ldots, m$. Furthermore, $y^k \to f^+(x^\star)$.

Proof. By (8c), it follows that, for all $i = 1, \dots, m$,

$$y_i^k \geq \max\{0, f_i(x^k)\}. \tag{10}$$

Moreover, since $c_k \to \infty$, we can assume that $c_k > 0$. Thus, the conditions (8b) and (8e) ensure that $\lambda_i^k > 0$ or $\mu_i^k > 0$. So, using (8g), we conclude that $y_i^k = 0$ or $y_i^k = f_i(x^k)$. This, together with (10), yields $y_i^k = \max\left\{0, f_i(x^k)\right\} = f_i^+(x^k)$. Finally, by the continuity of f^+, we obtain

$$y^k = f^+(x^k) \to f^+(x^\star),$$

completing the proof. □

Now, we prove that the multipliers of the simple-bounded constraints are orthogonal and fulfill a bound on their norm related to the stationary conditions, depending on the optimal combination of the gradients of the objective function and of the general constraints.

Lemma 3.2. *For the sequence (7) it holds $\nu_j^k \zeta_j^k = 0$ for all $k \in \mathbb{N}$ and $j = 1, \dots, n$. In particular, $(\nu^k)^T \zeta^k = 0$. Furthermore,*

$$\|\nu^k\|_2 \leq \left\| \nabla f_0(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k) \right\|_2 \tag{11}$$

and

$$\|\zeta^k\|_2 \leq \left\| \nabla f_0(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k) \right\|_2. \tag{12}$$

Proof. By (8d) and due to the fact that $x_j^{\min} < x_j^{\max}$, we have that either $x_j^k > x_j^{\min}$ or $x_j^k < x_j^{\max}$ holds. Thus, the complementarity condition (8h) ensures that $\nu_j^k = 0$ or $\zeta_j^k = 0$, proving the first part of the result. To prove that (11) is valid, take the inner product of ν^k by the expression in (8a), obtaining

$$(\nu^k)^T \nu^k = (\nu^k)^T \left(\nabla f_0(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k) \right).$$

So, by the Cauchy-Schwarz inequality, it follows that

$$\|\nu^k\|_2 \leq \left\| \nabla f_0(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k) \right\|_2.$$

The inequality (12) can be obtained by the same reasoning. □

In the next two lemmas we relate the multipliers of the simple bounds with the multipliers of the general augmented constraints. The second result is valid for an arbitrary norm.

Lemma 3.3. *If the sequence (λ^k) is bounded, so are (ν^k) and (ζ^k).*

Proof. It follows from the inequalities established in Lemma 3.2 and due to $x^k \to x^\star$ and $f_i \in C^1$, $i = 0, \dots, m$. ◻

Lemma 3.4. *If the sequence* (λ^k) *is unbounded, then*

$$\|\nu^k\| = O(\|\lambda^k\|) \quad \text{and} \quad \|\zeta^k\| = O(\|\lambda^k\|).$$

Proof. From Lemma 3.2 we obtain

$$\frac{\|\nu^k\|_2}{\|\lambda^k\|} \leq \left\| \frac{\nabla f_0(x^k)}{\|\lambda^k\|} + \sum_{i=1}^{m} \frac{\lambda_i^k}{\|\lambda^k\|} \nabla f_i(x^k) \right\|_2 .$$

As $\dfrac{\nabla f_0(x^k)}{\|\lambda^k\|} \to 0$ and the sequences $\left(\dfrac{\lambda_i^k}{\|\lambda^k\|}\right)$ and $\left(\nabla f_i(x^k)\right)$ are bounded, it follows that $\|\nu^k\|_2 = O(\|\lambda^k\|)$. By the equivalence of norms, we conclude the first claim. Analogously, we prove that $\|\zeta^k\| = O(\|\lambda^k\|)$, completing the proof. ◻

Now, we present our first stationarity result, assuming the boundedness of the sequence (λ^k).

Theorem 3.5. *If the sequence* (λ^k) *is bounded, then the limit point* $x^\star$ *of the sequence generated by the Framework I is a KKT point of the original problem* (OP).

Proof. By Lemma 3.3, the sequences (ν^k) and (ζ^k) are bounded. So, there exists an infinite set $\mathbb{N}' \subset \mathbb{N}$ and vectors $\lambda^\star \in \mathbb{R}_+^m$, $\nu^\star, \zeta^\star \in \mathbb{R}_+^n$ such that

$$\lambda^k \overset{\mathbb{N}'}{\to} \lambda^\star \,,\ \nu^k \overset{\mathbb{N}'}{\to} \nu^\star \quad \text{and} \quad \zeta^k \overset{\mathbb{N}'}{\to} \zeta^\star.$$

Therefore, taking the limit in (8a), in this subsequence, we obtain

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) - \nu^\star + \zeta^\star = 0,$$

which, together with the fact that $x^\star \in \mathcal{X}$, proves (3a), (3c)-(3e). Now, let us see that (3b) holds. If $f_i(x^\star) > 0$ for some i, then $f_i^+(x^\star) > 0$. By Lemma 3.1, $y_i^k > 0$ for all $k \in \mathbb{N}$ sufficiently large. Thus, from (8g), $\mu_i^k = 0$, which in turn implies that $\lambda_i^k = c_k$, because of (8b). But (λ^k) is bounded and (c_k) is unbounded. This contradiction means that $f_i(x^\star) \leq 0$. It remains to prove the complementarity condition. If $f_i(x^\star) < 0$, then, for all $k \in \mathbb{N}$ sufficiently large, we have $f_i(x^k) < 0$, which yields

$$y_i^k = \max\{0, f_i(x^k)\} = 0.$$

So, by (8g), we obtain $\lambda_i^k f_i(x^k) = 0$. Since $f_i(x^k) < 0$, it follows that $\lambda_i^k = 0$. Taking the limit, we have $\lambda_i^\star = 0$, proving (3f). Finally, we get (3g) taking the limit in (8h). ◻

In order to analyze the case in which (λ^k) is unbounded, we shall consider the following problem of minimizing the infeasibility measure associated with the functional constraints of (1):

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} f_i^+(x) \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{13}$$

The nonsmooth problem (13) can be rewritten as

$$\begin{array}{ll} \underset{(x,z)}{\text{minimize}} & \sum_{i=1}^{m} z_i \\ \text{subject to} & f_i(x) - z_i \leq 0,\ i = 1, \ldots, m \\ & x \in \mathcal{X},\ z \geq 0, \end{array} \tag{14}$$

whose KKT conditions are

$$\sum_{i=1}^{m} \gamma_i \nabla f_i(x) - \alpha + \beta = 0, \tag{15a}$$

$$\mathbb{1} - \gamma - \delta = 0, \tag{15b}$$

$$f_i(x) \leq z_i \quad \text{and} \quad z_i \geq 0, \quad i = 1, \ldots, m \tag{15c}$$

$$x_j^{\min} - x_j \leq 0 \quad \text{and} \quad x_j - x_j^{\max} \leq 0, \quad j = 1, \ldots, n \tag{15d}$$

$$\gamma_i \geq 0 \quad \text{and} \quad \delta_i \geq 0, \quad i = 1, \ldots, m \tag{15e}$$

$$\alpha_j \geq 0 \quad \text{and} \quad \beta_j \geq 0, \quad j = 1, \ldots, n \tag{15f}$$

$$\gamma_i(f_i(x) - z_i) = 0 \quad \text{and} \quad \delta_i z_i = 0, \quad i = 1, \ldots, m \tag{15g}$$

$$\alpha_j(x_j^{\min} - x_j) = 0 \quad \text{and} \quad \beta_j(x_j - x_j^{\max}) = 0, \quad j = 1, \ldots, n. \tag{15h}$$

The next stationarity result states that $(x^\star, f^+(x^\star))$, with $x^\star$ as in (9), is a KKT point of the problem (14).

Theorem 3.6. *Suppose that the sequence (λ^k) is unbounded. Then there exist vectors $\alpha^\star, \beta^\star \in \mathbb{R}_+^n$, $\gamma^\star, \delta^\star \in \mathbb{R}_+^m$, such that $(x^\star, f^+(x^\star), \alpha^\star, \beta^\star, \gamma^\star, \delta^\star)$ satisfies the conditions (15a)-(15h).*

Proof. Consider the vector $c^k = c_k \mathbb{1}$ defined by (6). From (8b) and (8e),

$$\lambda_i^k \leq \lambda_i^k + \mu_i^k = c_k \quad \text{and} \quad \mu_i^k \leq \lambda_i^k + \mu_i^k = c_k, \ \text{ for } i = 1, \ldots, m.$$

Thus,

$$\|\lambda^k\| = \mathcal{O}(c_k) \quad \text{and} \quad \|\mu^k\| = \mathcal{O}(c_k). \tag{16}$$

Define

$$\alpha^k = \frac{\nu^k}{c_k},\ \beta^k = \frac{\zeta^k}{c_k},\ \gamma^k = \frac{\lambda^k}{c_k} \quad \text{and} \quad \delta^k = \frac{\mu^k}{c_k}.$$

By Lemma 3.4 and (16), the sequences (α^k) and (β^k) are bounded. Again using (16), we conclude that (γ^k) and (δ^k) are also bounded. Therefore, there exists an infinite set $\mathbb{N}'' \subset \mathbb{N}$ and vectors $\alpha^\star, \beta^\star \in \mathbb{R}^n_+$, $\gamma^\star, \delta^\star \in \mathbb{R}^m_+$, such that

$$\alpha^k \xrightarrow{\mathbb{N}''} \alpha^\star \ , \ \beta^k \xrightarrow{\mathbb{N}''} \beta^\star \ , \ \gamma^k \xrightarrow{\mathbb{N}''} \gamma^\star \quad \text{and} \quad \delta^k \xrightarrow{\mathbb{N}''} \delta^\star.$$

Dividing (8a) by c_k and taking the limit in this subsequence, we obtain

$$\sum_{i=1}^{m} \gamma_i^\star \nabla f_i(x^\star) - \alpha^\star + \beta^\star = 0,$$

proving (15a). Dividing (8b) by c_k and taking the limit for $k \in \mathbb{N}''$, we get (15b). The conditions (15c)-(15f) are straightforward. To obtain (15g), note first that by Lemma 3.1, $y^k \to f^+(x^\star)$. Now it is enough to divide (8g) by c_k and take the limit for $k \in \mathbb{N}''$. Finally, we can see that (15h) follows from (8h). □

We stress that, for a fixed $k \in \mathbb{N}$, the set of Lagrange multipliers for the augmented problem AP(c^k) is compact. This follows from Lemma 2.3. However, as $k \to \infty$, the control over the multipliers λ^k is lost. Thus, the ℓ_1-penalty function plays an *exact role* solely under the scenario of Remark 1, that is, if the augmented variables of the generated sequence happen to reach the value zero for a finite iterate k. Consequently, in such a case, the OP is not only feasible, but also fulfils a constraint qualification in the sense interpreted by Han and Mangasarian [15, p.267].

4 Examples

In this section we present two examples to highlight specific features of the sequences generated by the Framework I.

The first problem has no feasible points and illustrates the convergence of the sequence to the global minimizer of (13), so that the infeasibility is measured by the ℓ_1-norm of the constraint violation.

Example 4.1. Consider the original problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = x \\ \textit{subject to} & f_1(x) = x + 3 \leq 0, \\ & f_2(x) = (x-1)^2 + 2 \leq 0, \\ & x \in [-2, 2]. \end{array}$$

The augmented problem for this example, for which (6) holds, is

$$\begin{array}{ll} \text{minimize} & x + c_k(y_1 + y_2) \\ \text{subject to} & x + 3 - y_1 \leq 0, \\ & (x-1)^2 + 2 - y_2 \leq 0, \\ & x \in [-2, 2], \ y \geq 0. \end{array} \tag{17}$$

Consider a sequence of stationary points generated by the Framework I

$$\left(x^k, y^k, \lambda^k, \mu^k, \nu^k, \zeta^k\right). \tag{18}$$

Note that $y_i^k = f_i^+(x^k) = f_i(x^k) > 0$, which by (8g) yields $\mu^k = 0$ and so, $\lambda_i^k = c_k$. Furthermore, $x^k \neq -2$ and $x^k \neq 2$, which by (8h) and (8a), yield $1 + c_k + 2c_k(x^k - 1) = 0$. Solving this equation, we obtain

$$x^k = \frac{c_k - 1}{2c_k} \to \frac{1}{2}.$$

Note that the infeasibility measure is given by $f_1^+(x) + f_2^+(x) = x^2 - x + 6$, for $x \in [-2, 2]$, whose global minimizer is just the limit $1/2$ of the sequence (x^k). Besides, by Lemma 2.2, the augmented problem (17) has a global solution, which must satisfy (8a)-(8h), because of Lemma 2.3. Thus, the only KKT point (x^k, y^k) obtained here is, indeed, the global solution of (17).

Note that Theorem 3.5 implies, in particular, that any limit point of the sequence (x^k) is feasible for the original problem, when (λ^k) is bounded. The next example shows that the converse is not true.

Example 4.2. Consider the original problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = x \\ \textit{subject to} & f_1(x) = \dfrac{x^4}{4} - \dfrac{4x^3}{3} + \dfrac{3x^2}{2} + \dfrac{9}{4} \leq 0, \\ & x \in [-1, 4]. \end{array}$$

The augmented problem for Example 4.2 is

$$\begin{array}{ll} \text{minimize} & x + c_k y \\ \text{subject to} & \dfrac{x^4}{4} - \dfrac{4x^3}{3} + \dfrac{3x^2}{2} + \dfrac{9}{4} - y \leq 0, \\ & x \in [-1, 4], \ y \geq 0. \end{array}$$

Let $\left(x^k, y^k, \lambda^k, \mu^k, \nu^k, \zeta^k\right)$ be a sequence that satisfies (8a)-(8h). We have $x^k < 4$. Thus, $\zeta^k = 0$ and consequently,

$$1 + \lambda^k x^k (x^k - 1)(x^k - 3) = \nu^k. \tag{19}$$

Therefore, using (8h), we see that $x^k \neq 3$. So, since $f_1(x) \geq 0$ for all $x \in \mathbb{R}$ and its only root is $x = 3$, it follows that

$$y^k = f_1^+(x^k) = f_1(x^k) > 0.$$

By (8g), $\mu^k = 0$ and so, using (8b), $\lambda^k = c_k$. Furthermore, we have $x^k \neq -1$, which by (8h) and (19), yields $1 + c_k x^k (x^k - 1)(x^k - 3) = 0$, that is,

$$x^k (x^k - 1)(x^k - 3) = -\frac{1}{c_k}. \tag{20}$$

Note that the equation (20) has 3 solutions (for c_k greater than a constant $\bar{c} > 0$). So, for each index $k \in \mathbb{N}$, we obtain three stationary points of the corresponding augmented problem, namely,

$$(\bar{x}^k, f_1(\bar{x}^k)), \quad (\tilde{x}^k, f_1(\tilde{x}^k)) \quad \text{and} \quad (\hat{x}^k, f_1(\hat{x}^k)).$$

For all of them, we have $\lambda^k = c_k \to \infty$. Furthermore,

$$\bar{x}^k \to \bar{x} = 0\,,\ \tilde{x}^k \to \tilde{x} = 1 \quad \text{and} \quad \hat{x}^k \to \hat{x} = 3.$$

The infeasibility measure is $f_1^+(x) = f_1(x) = \dfrac{x^4}{4} - \dfrac{4x^3}{3} + \dfrac{3x^2}{2} + \dfrac{9}{4}$. Moreover, $\bar{x}$ is a local minimizer, $\tilde{x}$ is a local maximizer and $\hat{x}$ is the global minimizer of f_1^+. Figure 1 shows the level curves of the objective function of the augmented problem, denoted by $F_{c_k}(x, y) \equiv F_k(x, y) = x + c_k y$, as well as its feasible set. Additionally, the three stationary points are also depicted. Note that $(\tilde{x}^k, f_1(\tilde{x}^k))$ is not a local maximizer of F_k.

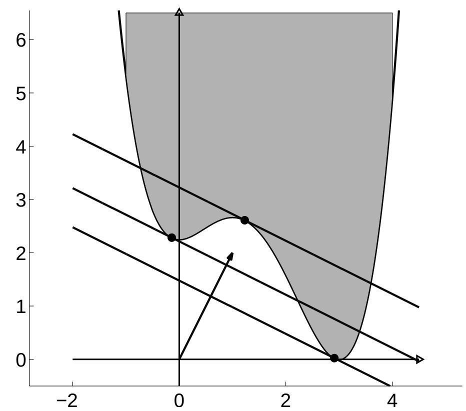

Figure 1: The stationary points of the augmented problem.

This example answers negatively the following questions:

- If the limit point of (x^k) is feasible, can we conclude that (λ^k) is bounded?
- If the feasible set of the OP is nonempty, can we conclude that the limit point of (x^k) belongs to this set?

The Examples 4.1 and 4.2 also illustrate an additional property of the Framework I: if we obtain global solutions of the augmented problems, then each limit point of (x^k) is a global solution of the infeasibility problem (13). In addition, if the feasible set of the original problem (1) is nonempty, then each limit point of (x^k) is a global solution of this problem. These results are similar to the ones obtained for the classical external penalty methods (cf. [9, 10]).

5 Convergence to local minimizers

Along the exact perspective of the ℓ_1-penalty function, under the regularity assumption of linearly independence of the gradients of the active constraints, Pietrzykowski [18] has proved the equivalence of local solutions of problems (1) and (4).

By assuming that a point $x^\star$ satisfies the second order sufficiency conditions to be a strong local minimizer for problem (1) and, using our notation, assuming that $\lambda_i^\star \leq c_i$ holds for all $i \ldots, m$, that is, the associated KKT multipliers of the nonlinear inequalities of the OP are bounded by the parameters c_i, Charalambous [6] has proved that $x^\star$ is a strong local minimizer of problem (4) as well.

We start this section by showing the relation between local solutions of problems (4) and (5), without any further assumption upon the problem (1).

Theorem 5.1. *If $(\hat{x}, \hat{y})$ is a local solution of the problem (5), then $\hat{x}$ is a local solution of the problem (4). On the other hand, if $\hat{x}$ is a local solution of the problem (4), then $(\hat{x}, f^+(\hat{x}))$ is a local solution of the problem (5).*

Proof. Suppose first that $(\hat{x}, \hat{y})$ is a local solution of (5). By Lemma 3.1, we have $\hat{y} = f^+(\hat{x})$. Furthermore, there exist neighborhoods $B_\delta \subset \mathbb{R}^n$ of $\hat{x}$ and $B_\varepsilon \subset \mathbb{R}^m$ of $\hat{y}$ such that

$$f_0(\hat{x}) + \sum_{i=1}^{m} c_i \hat{y}_i \leq f_0(x) + \sum_{i=1}^{m} c_i y_i \tag{21}$$

for all feasible pair $(x, y) \in B_\delta \times B_\varepsilon$. As f^+ is continuous, we can assume, without loss of generality, that $f^+(x) \in B_\varepsilon$ for all $x \in B_\delta$. Therefore, given $x \in \mathcal{X} \cap B_\delta$, we have that $(x, f^+(x))$ is feasible for (5) and belongs to the neighborhood $B_\delta \times B_\varepsilon$. Thus, using (21), we conclude that, for all $x \in B_\delta$,

$$f_0(\hat{x}) + \sum_{i=1}^{m} c_i f_i^+(\hat{x}) = f_0(\hat{x}) + \sum_{i=1}^{m} c_i \hat{y}_i \;\leq\; f_0(x) + \sum_{i=1}^{m} c_i f_i^+(x),$$

which means that $\hat{x}$ is a local minimizer of (4).

Conversely, suppose that $\hat{x}$ is a local solution of (4). So, $\hat{x} \in \mathcal{X}$ and

$$f_0(\hat{x}) + \sum_{i=1}^{m} c_i f_i^+(\hat{x}) \leq f_0(x) + \sum_{i=1}^{m} c_i f_i^+(x) \tag{22}$$

for all $x \in \mathcal{X}$, x close enough to $\hat{x}$. Given (x, y) feasible for the problem (5) we have that $x \in \mathcal{X}$ and $f^+(x) \leq y$. Thus, if x is close to $\hat{x}$, it follows from (22) that

$$f_0(\hat{x}) + \sum_{i=1}^{m} c_i f_i^+(\hat{x}) \leq f_0(x) + \sum_{i=1}^{m} c_i y_i$$

and hence the pair $(\hat{x}, f^+(\hat{x}))$ is a local solution of (5). □

Note that, in the converse part of Theorem 5.1, the variable y is not restricted to be close to $f^+(\hat{x})$. This means that if $B_\delta \subset \mathbb{R}^n$ is a neighborhood of $\hat{x}$ in which this point minimizes the ℓ_1-penalty function, then $B_\delta \times \mathbb{R}^m$ is a neighborhood of $(\hat{x}, \hat{y})$ in which this augmented point minimizes the objective function of problem (5).

To analyze further the possible relationship between sequences of local solutions of problems (5) and a local solution of problem (4), consider the following example:

Example 5.1. Let the original problem be

$$\begin{array}{ll} \text{minimize} & f_0(x) = x \\ \textit{subject to} & f_1(x) = \dfrac{x^4}{4} - \dfrac{3x^3}{4} + \dfrac{1}{12}\left(\dfrac{9}{4}\right)^4 \leq 0, \\ & x \in [-1, 3]. \end{array}$$

Figure 2 illustrates Theorem 5.1 by showing a local minimizer of the augmented problem (5), associated with the OP stated in Example 5.1, and the corresponding local minimizer of the ℓ_1-penalty problem (4). It also depicts the feasible region of the associated AP (shaded region for x between -1 and 3, the curved border of which corresponds to the graph of $y = f_1^+(x)$), and the graph of the ℓ_1-penalty function $F_c(x) = f_0(x) + cf_1^+(x)$, in the dotted curve. The straight line is, locally, the best level curve of the augmented function $(x, y) \mapsto f_0(x) + cy$.

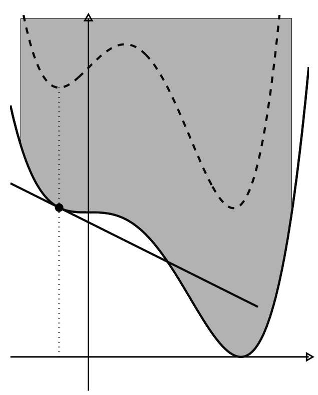

Figure 2: The augmented problem and the ℓ_1-penalty function (Example 5.1).

As established previously, the limit point of the sequence generated by the Framework I is either a *solution* of the original problem (1) or a *solution* of the problem of minimizing the infeasibility measure associated with the constraints of (1). The so-called *solution* stands for a stationary point or a global minimizer.

Unfortunately, the conclusions obtained for stationary points and global minimizers do not hold when we consider a sequence of local solutions. As we can see, suggested by Figure 2, the sequence of local minimizers of the ℓ_1-penalty function might have a limit point which is neither a local solution of the original problem nor a local minimizer of the infeasibility measure.

Despite being feasible, the limit point of (x^k) may not be a local minimizer of the original problem (1). This possibility is illustrated by the following example:

Example 5.2. Let the OP be

$$\begin{array}{ll} \text{minimize} & f_0(x) = x \\ \textit{subject to} & f_1(x) \leq 0, \\ & x \in [-0.1, 0.35], \end{array}$$

where $f_1(x) = \begin{cases} x^4 \sin^2\left(\frac{1}{x}\right), & \textit{if } x \neq 0 \\ 0, & \textit{if } x = 0. \end{cases}$

In Figure 3, by showing simultaneously the locally best level curves of the augmented function $(x, y) \mapsto f_0(x) + c^k y$, we have depicted a sequence of local minimizers of the AP(c^k) converging to the origin, a feasible point which is not a minimizer of the OP.

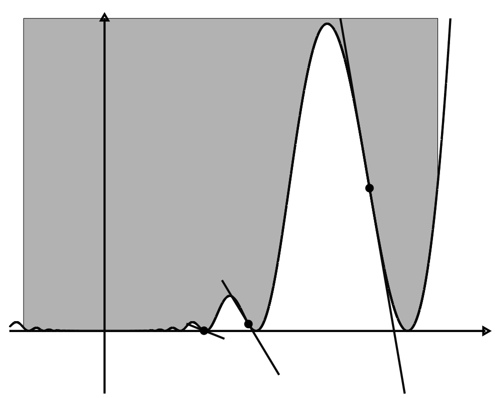

Figure 3: A sequence of local minimizers of the AP converging to a feasible point which is not a minimizer of the OP (Example 5.2).

6 Conclusions

We have analyzed the features of the sequence generated by the ℓ_1-penalty framework, associated with the inequality constrained nonlinear programming problem with simple bounds, the so-called original problem (OP). To preserve the potential differentiability of the functions that define the OP, a smooth reformulation was employed as the main tool for the analysis. Under the compactness of the set defined by the simple bounds, the continuous differentiability of all the original functions and the assumption that the vector of the penalty parameters is a scalar vector, we have examined not only stationary points, but also global and local minimizers of the augmented problem, with the aim of relating them to the corresponding counterparts of the OP.

Among the studied properties we emphasize that, in case the ℓ_1-penalty function does not play an exact role, as stated in Remark 1, the feasibility of the OP is related to the boundedness of the Lagrange multiplier. Moreover, the limit point of the sequence generated by the Framework I is either a stationary point of the OP (Theorem 3.5) or a stationary point of the problem of minimizing the infeasibility measure associated with the constraints of the OP (Theorem 3.6). Examples were provided to emphasize the intrinsic features: the convergence to a global minimizer of the infeasibility measure in case the OP has no feasible points (Example 4.1); the possibility of unbounded Lagrange multipliers, even for feasible original problems (Example 4.2); the fact that the limit points might not belong to a nonempty feasible set of the OP (Example 4.2) and the fact that the results obtained for stationary points, also valid for global minimizers, cannot be extended to local minimizers (Example 5.1), as a sequence of local minimizers of the ℓ_1-penalty function might have a limit point which is neither a local solution of the OP nor a local minimizer of the infeasibility measure (Example 5.2). Also, without any further assumption upon the OP, we have established the equivalence between the local solutions of the ℓ_1-penalty function formulation and its differentiable counterpart, the AP (Theorem 5.1).

When it comes to general nonlinear programming problems, to ensure that a local minimizer has been determined, second-order conditions must be verified, demanding further assumptions on the smoothness of the involved functions. In practical terms, even stationarity is not exactly achieved in general. Under this perspective, investigating the relationship between the satisfaction of sequential optimality conditions [1, 2] for the AP and for the OP are of interest, being the subject of ongoing research.

Acknowledgements. The authors are grateful to José Mario Martínez, for his fruitful suggestions that have encouraged us to pursue this research, and to Márcia A. Gomes-Ruggiero, for sharing our preliminary discussions.

References

[1] R. Andreani, G. Haeser, J.M. Martínez, On sequential optimality conditions for smooth constrained optimization, Optimization 60 (2011) 627–641.

[2] R. Andreani, J.M. Martínez, B.F. Svaiter, A new sequential optimality condition for constrained optimization and algorithmic consequences, SIAM Journal on Optimization 6 (2010) 3533–3554.

[3] M.S. Bazaraa, J.J. Jarvis, H.D. Sherali, Linear Programming and Network Flows, Third ed., John Wiley & Sons, Hoboken, New Jersey, 2005.

[4] A. Ben-Tal, M. Teboulle, Hidden convexity in some nonconvex quadratically constrained quadratic programming, Mathematical Programming 72 (1) (1996) 51–63.

[5] R. Byrd, F. Curtis, J. Nocedal, Infeasibility detection and SQP methods for nonlinear optimization, SIAM Journal on Optimization 20 (5) (2010) 2281–2299.

[6] C. Charalambous, On conditions for optimality of the nonlinear ℓ_1 problem, Mathematical Programming 17 (1979) 123–135.

[7] L. Chen, D. Goldfarb, Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties, Mathematical Programming 108 (2006) 1–36.

[8] F.E. Curtis, A penalty-interior-point algorithm for nonlinear constrained optimization, Mathematical Programming Computation 4 (2) (2012) 181–209.

[9] A.V. Fiacco, G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, SIAM, Philadelphia, 1990.

[10] R. Fletcher, Practical Methods of Optimization, Second ed., John Wiley & Sons, Chichester, England, 1987.

[11] C. Fleury, CONLIN: an efficient dual optimizer based on convex approximations concepts, Structural Optimization 1 (1989) 81–89.

[12] C. Fleury, V. Braibant, Structural optimization: a new dual method using mixed variables, International Journal for Numerical Methods in Engineering 23 (1986) 409–428.

[13] P.E. Gill, W. Murray, M.A. Saunders, SNOPT: An SQP algorithm for large-scale constrained optimization, SIAM Journal on Optimization 12 (4) (2002) 979–1006.

[14] M.A. Gomes-Ruggiero, M. Sachine, S.A. Santos, A spectral updating for the method of moving asymptotes, Optimization Methods & Software 25 (6) (2010) 883–893.

[15] S.-P. Han, O.L. Mangasarian, Exact penalty functions in nonlinear programming, Mathematical Programming 17 (1979) 251–269.

[16] J.H. Jung, D.P. O'Leary, A.L. Tits, Adaptive constraint reduction for convex quadratic programming, Computational Optimization and Applications 51 (2012) 125–157.

[17] J.M. Martínez, L.F. Prudente, Handling infeasibility in a large-scale nonlinear optimization algorithm, Numerical Algorithms 60 (2012) 263–277.

[18] T. Pietrzykowski, An exact potential method for constrained maxima, SIAM Journal on Numerical Analysis 6 (2) (1969) 299–304.

[19] K. Svanberg, The method of moving asymptotes - a new method for structural optimization, International Journal for Numerical Methods in Engineering 24 (1987) 359–373.

[20] K. Svanberg, The method of moving asymptotes (MMA) with some extensions, in Optimization of Large Structural Systems, Proc. NATO ASI, Berchtesgaden, Germany, 1991, G. I. N. Rozvany, ed., Kluwer Academic Publishers, Dordretch, 1993, 555–566.

[21] K. Svanberg, A class of globally convergent optimization methods based on conservative convex separable approximations, SIAM Journal on Optimization 12 (2002) 555–573.

[22] W.H. Zhang, C. Fleury, P. Duysinx, V.H. Nguyen, I. Laschet, A Generalized method of moving asymptotes (GMMA) including equality constraints, Structural Optimization 12 (1996) 143–146.