Computational Optimization and Applications, 25, 57–83, 2003

An Exact Augmented Lagrangian Function for Nonlinear Programming with Two-Sided Constraints

GIANNI DI PILLO
GIAMPAOLO LIUZZI
STEFANO LUCIDI
LAURA PALAGI
Dipartimento di Informatica e Sistemistica "Antonio Ruberti," Università di Roma "La Sapienza"
via Buonarroti, 12, 00185 Roma, Italy

Received February 15, 2002; Accepted July 22, 2002

Dedicated to: This paper is our modest tribute to Elijah Polak, an eminent scholar who greatly influenced the development of optimization theory and practice for over forty years. Several of his contributions in unconstrained and constrained nonlinear programming, both in the smooth and in the non smooth case, in the analysis and optimization of control systems, in the implementation of effective design methods in engineering, are milestones. We are in particular grateful to Eljiah for his work on exact penalty algorithms, from which we derived much inspiration, including some at the basis of this paper.

Abstract. This paper is aimed toward the definition of a new exact augmented Lagrangian function for two-sided inequality constrained problems. The distinguishing feature of this augmented Lagrangian function is that it employs only one multiplier for each two-sided constraint. We prove that stationary points, local minimizers and global minimizers of the exact augmented Lagrangian function correspond exactly to KKT pairs, local solutions and global solutions of the constrained problem.

Keywords: nonlinear programming, augmented Lagrangian function, two-sided constraints

1. Introduction and assumptions

In this paper we are concerned with the inequality constrained problem

$$\begin{aligned} \min \quad & f(x) \\ & l \le g(x) \le u, \end{aligned} \tag{P}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$, $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $l, u \in \mathbb{R}^m$ are such that $l_i < u_i$ for $i = 1, \dots, m$. We denote the feasible set of Problem (P) by

$$\mathcal{F} = \{x \in \mathbb{R}^n : l \le g(x) \le u\}.$$

Let $\mathcal{S}$ be an open set such that $\mathcal{F} \subset \mathcal{S}$. We assume $f(x)$ and $g(x)$ to be twice continuously differentiable functions over $\mathcal{S}$. We denote by $\mathring{\mathcal{S}}$ the interior of $\mathcal{S}$, by $\bar{\mathcal{S}}$ its closure and by $\partial\mathcal{S}$ its boundary.

Frequently real world problems present constraints with both lower and upper bounds, so that it is of great interest to treat them directly rather than to split them into two singlesided constraints, thus obtaining

$$\begin{aligned} \min \quad & f(x) \\ & g(x) - u \leq 0 \\ & l - g(x) \leq 0. \end{aligned} \tag{1}$$

It is well-known that Problem (1) can be transformed into an unconstrained minimization problem by employing a continuously differentiable exact merit function. In particular if this transformation is carried out by means of an exact augmented Lagrangian function, we come up with a problem on the augmented space $\mathbb{R}^{n+2m}$ of the primal and dual variables. In the case of large number of constraints, this approach may lead us to tackle huge problems. In this context, we show that it is possible to introduce an exact augmented Lagrangian function for Problem (P) depending only on $n + m$ variables.

The starting point of our approach consists in conveniently rewriting the KKT conditions for Problem (P) by introducing only m multipliers; as in [1]. Specifically, we say that a pair $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a KKT pair if it satisfies:

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} = 0 \tag{2a}$$

$$\forall i = 1, \ldots, m \begin{cases} g_i(\bar{x}) = u_i & \text{and} \quad \bar{\lambda}_i \geq 0, \quad \text{or} \\ g_i(\bar{x}) = l_i & \text{and} \quad \bar{\lambda}_i \leq 0, \quad \text{or} \\ l_i < g_i(\bar{x}) < u_i & \text{and} \quad \bar{\lambda}_i = 0. \end{cases} \tag{2b}$$

Before going into details, we introduce the following definitions.

The *Lagrangian function* associated with Problem (P) is the function $L : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ given by

$$L(x, \lambda) = f(x) + \lambda^\top g(x),$$

and we denote its gradient and Hessian by

$$\nabla_x L(x, \lambda) = \nabla f(x) + \nabla g(x)\lambda$$
$$\nabla_x^2 L(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x).$$

Given a vector $x \in \mathbb{R}^n$, we define

$$A_0(x) = \{i : g_i(x) = u_i \text{ or } g_i(x) = l_i\},$$

as the set of the indices of the constraints which are active at their lower or upper bound. If the gradients $\nabla g_i(\bar{x})$, $i \in A_0(\bar{x})$ are linearly independent at a feasible point $\bar{x}$, then the KKT conditions (2) are *first order necessary optimality conditions* for $\bar{x}$ to be a local solution of Problem (P) with associated multiplier $\bar{\lambda}$.

Given a KKT pair $(\bar{x}, \bar{\lambda})$, we define the following set

$$A_+(\bar{x}, \bar{\lambda}) = \{i : g_i(\bar{x}) = u_i, \bar{\lambda}_i > 0 \text{ or } g_i(\bar{x}) = l_i, \bar{\lambda}_i < 0\}.$$

We say that a KKT pair $(\bar{x}, \bar{\lambda})$ satisfies the *strict complementarity condition* if it holds that

$$A_+(\bar{x}, \bar{\lambda}) = A_0(\bar{x}).$$

We say that a KKT pair $(\bar{x}, \bar{\lambda})$ satisfies the *strong second order sufficient optimality condition* if it happens that

$$y^\top \nabla_x^2 L(\bar{x}, \bar{\lambda})\, y > 0, \quad \forall y : \nabla g_i(\bar{x})^\top y = 0 \quad \text{with } i \in A_+(\bar{x}, \bar{\lambda}). \tag{3}$$

In the sequel we make use of the following assumptions.

Assumption A1. The open set $\mathcal{S}$ and the constraint functions $g_i(x)$, $i = 1, \ldots, m$ are such that for every sequence $\{x^k\}$ converging toward $\bar{x} \in \partial\mathcal{S}$, an index $i \in \{1, \ldots, m\}$ exists such that

$$\lim_{k\to\infty} |g_i(x^k)| = +\infty.$$

Assumption A2. At least one of the two following conditions is satisfied:

(a) there exists a known feasible point $\hat{x} \in \mathcal{F}$ and $f(x)$ is coercive on $\bar{\mathcal{S}}$ (that is for any $\{x^k\} \subseteq \mathcal{S}$ such that $\|x^k\| \to \infty$ we have $f(x^k) \to \infty$);
(b) the set $\bar{\mathcal{S}}$ is compact and at every point $x \in \mathcal{S} \setminus \mathcal{F}$

$$\sum_{i=1}^{m} r_i(x) \nabla g_i(x) \neq 0$$

where

$$r_i(x) \begin{cases} >0 & \text{if } g_i(x) > u_i \\ =0 & \text{if } l_i \le g_i(x) \le u_i \quad \forall i = 1, \ldots, m \\ <0 & \text{if } g_i(x) < l_i \end{cases}$$

Assumption A3. For every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in A_0(x)$, are linearly independent.

Assumption A1 is not as restrictive as it may appear. It allows us to considerably simplify the definition of the new exact augmented Lagrangian function. This assumption can be easily satisfied by appropriately choosing the open perturbation $\mathcal{S}$ of the feasible set and by scaling the constraints consequently. In fact, let the feasible set of the original problem be

$$\tilde{\mathcal{F}} = \{x \in \mathbb{R}^n : \tilde{l} \leq \tilde{g}(x) \leq \tilde{u}\}$$

and assume that Assumption A1 is not satisfied. For every $\alpha_i > 0, i = 1, \dots, m$, a first choice for the set $\mathcal{S}$ is the following:

$$\mathcal{S} = \{x \in \mathbb{R}^n : (\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x))) > 0, i = 1, \dots, m\},$$

Consequently let us modify the constraints as follows:

$$g_i(x) = \frac{\tilde{g}_i(x) - \tilde{l}_i}{(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x)))}, \quad i = 1, \dots, m$$

and the corresponding bounds as $l_i = 0$ and $u_i = (\tilde{u}_i - \tilde{l}_i)/\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)$. A second choice corresponds to setting

$$\mathcal{S} = \{x \in \mathbb{R}^n : \alpha_i - \max\{\tilde{l}_i - \tilde{g}_i(x), 0\}^3 - \max\{\tilde{g}_i(x) - \tilde{u}_i, 0\}^3 > 0, i = 1, \dots, m\},$$

and

$$g_i(x) = \frac{\tilde{g}_i(x) - \tilde{l}_i}{\alpha_i - \max\{\tilde{l}_i - \tilde{g}_i(x), 0\}^3 - \max\{\tilde{g}_i(x) - \tilde{u}_i, 0\}^3},$$

being $l_i = 0$ and $u_i = (\tilde{u}_i - \tilde{l}_i)/\alpha_i$ the corresponding bounds.

We can prove that $\tilde{\mathcal{F}} = \mathcal{F}$ and that the modified function $g(x)$ satisfies Assumptions A2 and A3, if the original function $\tilde{g}(x)$ does. The proof is reported in Appendix A.1.

As concerns Assumptions A2 and A3, they are usual assumptions in the field of continuously differentiable exact merit functions (see, e.g., [1, 3, 14]).

We remark that the above assumptions are *a priori* assumptions in the sense that they regard the problem statement rather than the behaviour of an algorithm. We refer the interested reader to [15] for a discussion on similar assumptions. We just note that by Assumption A2 (a) constrained optimization problems with unbounded feasible set can be tackled, provided that a feasible point is known and that $f(x)$ is coercive on $\bar{\mathcal{S}}$. Assumption A2 (b) is a weaker form of the Mangasarian-Fromowitz constraint qualification condition ([16] and [5]).

In Sections 2–6 we shall suppose that Assumptions A1–A3 hold.

We conclude this section by introducing some basic notation. Given a vector v, we denote by V the diagonal matrix $V = \text{diag}(v)$. Let S and T be two index subsets, then we denote by v_S the subvector of a vector v with components $v_i, i \in S$ and by Q_{ST} the submatrix of a matrix Q made up by elements q_{ij} with $i \in S$ and $j \in T$. Given two vectors u, v of same dimension, we denote by $\max\{u, v\}$ the vector with components $\max\{u_i, v_i\}$. Moreover, we denote by $\|v\|$ the Euclidean norm of v.

2. The augmented Lagrangian function

In this section we introduce a continuously differentiable exact augmented Lagrangian function whose expression stems from the KKT conditions written as in (2). It is well-known [3, 6, 14] that a continuously differentiable exact augmented Lagrangian function can be obtained by adding to the objective function terms that provide a "smooth" penalization of the violation of the KKT conditions.

A general expression of an exact augmented Lagrangian function is the following:

$$G(x, \lambda; \epsilon) = f(x) + \frac{1}{2\epsilon}\psi(x, \lambda; \epsilon) + \eta(x, \lambda),$$

where ϵ is a strictly positive penalty parameter (see e.g., [6]). Roughly speaking the first penalty term $\psi(x, \lambda; \epsilon)$ forces the feasibility and the complementarity condition, while $\eta(x, \lambda)$ is a positive term that penalizes the distance between the variable λ and a KKT multiplier $\bar{\lambda}$ and that, in a neighborhood of KKT pair, convexifies in some sense $G(x, \lambda; \epsilon)$ with respect to λ.

The term $\psi(x, \lambda; \epsilon)$

In order to define $\psi(x, \lambda; \epsilon)$, we draw our inspiration from [1]. First, let us introduce the function $\gamma : \mathcal{S} \times \mathbb{R}^m \to \mathbb{R}^m$, defined componentwise as

$$\gamma_i(x, \lambda; \epsilon p(\lambda)) = \begin{cases} g_i(x) - u_i & \text{if } g_i(x) - u_i \geq -\epsilon p(\lambda)\lambda_i \\ g_i(x) - l_i & \text{if } g_i(x) - l_i \leq -\epsilon p(\lambda)\lambda_i \\ -\epsilon p(\lambda)\lambda_i & \text{otherwise,} \end{cases} \quad i = 1, \dots, m$$

where $p(\lambda)$ is a strictly positive scalar function. We note that, in a more compact notation, γ can be written as

$$\gamma(x, \lambda; \epsilon p(\lambda)) = \min\{g(x) - l, -\epsilon p(\lambda)\lambda\} + \max\{g(x) - u, -\epsilon p(\lambda)\lambda\} + \epsilon p(\lambda)\lambda.$$

The following property holds.

Proposition 2.1. *It results $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$ if and only if the pair $(\bar{x}, \bar{\lambda})$ satisfies* (2b).

Now we are ready to define the term ψ. In particular we set:

$$\psi(x, \lambda; \epsilon p(\lambda)) = 2\epsilon\lambda^\top \gamma(x, \lambda; \epsilon p(\lambda)) + \frac{1}{p(\lambda)}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2.$$

It is easily verified (reasoning as in [6]) that the function $\psi(x, \lambda; \epsilon p(\lambda))$ satisfies the following properties:

- if $p(\lambda)$ is a continuously differentiable function, then $\psi(x, \lambda; \epsilon p(\lambda))$ is continuously differentiable with respect to (x, λ);

– if $x \in \mathcal{F}$, then $\psi(x, \lambda; \epsilon p(\lambda)) = 0$ if and only if (x, λ) satisfy conditions (2b);
– $\lim_{\epsilon \to 0} \psi(x, \lambda; \epsilon p(\lambda)) = \|\rho(x)\|^2 / p(\lambda)$ where $\rho_i(x)$, for $i = 1, \ldots, m$, are given by

$$\rho_i(x) = \max\{g_i(x) - u_i, 0\} + \min\{g_i(x) - l_i, 0\}. \tag{4}$$

We point out that $\rho(x) = 0$ if and only if $x \in \mathcal{F}$. Therefore, when ϵ tends toward zero, the term ψ becomes a measure of the constraints violation.

It is worth noting that, if the function $p(\lambda)$ is poorly chosen (e.g. a positive constant) then the term $\psi(x, \lambda; \epsilon p(\lambda))$ can be unbounded from below with respect to λ. In order to avoid this possibility, we define the function $p(\lambda)$ as follows

$$p(\lambda) = \frac{1}{1 + \|\lambda\|^2}. \tag{5}$$

It is evident that the term $1/p(\lambda)$ penalizes the fact that the norm of vector λ is too large thus preventing function $\psi(x, \lambda; \epsilon p(\lambda))$ from being unbounded from below with respect to λ.

The term $\eta(x, \lambda)$

As already discussed, the key role of this term is to weight the distance between the actual value of the multiplier λ and the KKT multiplier $\bar{\lambda}$. This is obtained by minimizing with respect to λ a function penalizing condition (2a) and the complementarity conditions $(g_i(x) - l_i)(g_i(x) - u_i)\lambda_i = 0$, $i = 1, \ldots, m$, which are implied by (2b). In particular, we consider the function

$$\|\nabla f(x) + \nabla g(x)\lambda\|^2 + \|(G(x) - L)(G(x) - U)\lambda\|^2.$$

By minimizing the above function with respect to λ, we get the condition

$$\varphi(x, \lambda) = \nabla g(x)^\top (\nabla f(x) + \nabla g(x)\lambda) + (G(x) - U)^2 (G(x) - L)^2 \lambda = 0. \tag{6}$$

We set $\eta(x, \lambda) = \|\varphi(x, \lambda)\|^2$, that is

$$\eta(x, \lambda) = \|M(x)\lambda + \nabla g(x)^\top \nabla f(x)\|^2 \tag{7}$$

where

$$M(x) = \nabla g(x)^\top \nabla g(x) + (G(x) - U)^2 (G(x) - L)^2. \tag{8}$$

The function $\eta(x, \lambda)$ satisfies the following properties [6]:

– it is continuously differentiable with respect to (x, λ);
– it is a convex function with respect to λ;
– if $(\bar{x}, \bar{\lambda})$ is a KKT pair, then $\eta(\bar{x}, \lambda) = 0$ if and only if $\lambda = \bar{\lambda}$.

It is worth noting that, at points where $M(x)$ is nonsingular, $\eta(x, \lambda)$ can be rewritten as

$$\eta(x, \lambda) = \|M(x)(\lambda - \lambda(x))\|^2$$

with $\lambda(x)$ given by

$$\lambda(x) = -M(x)^{-1}\nabla g(x)^\top \nabla f(x).$$

The above function is an extension to double-sided constraints of the multiplier function first introduced by Glad and Polak in [12] for one-sided constraints. Following analogous reasoning as in [12], it is easy to check that, under Assumption A3, the matrix $M(x)$ given by (8) is nonsingular for all $x \in \mathcal{F}$.

We are now in a position to define an exact augmented Lagrangian function for Problem (P) as follows

$$\begin{aligned} L_b(x, \lambda; \epsilon) = {} & f(x) + \lambda^\top \gamma(x, \lambda; \epsilon p(\lambda)) + \frac{\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2}{2\epsilon p(\lambda)} \\ & + \|M(x)\lambda + \nabla g(x)^\top \nabla f(x)\|^2. \end{aligned} \tag{9}$$

We note that, if we let $p(\lambda) = 1$ and we remove the last term in (9) we obtain the augmented Lagrangian function proposed by Bertsekas in [1], which is an extension to two-sided constraints of the Hestenes-Powell-Rockafeller augmented Lagrangian function [18].

One-sided inequality constraints

Before entering into the analysis of the properties of $L_b(x, \lambda; \epsilon)$ we briefly discuss the case of one-sided inequality constrained problems:

$$\begin{aligned} \min \quad & f(x) \\ & g(x) \le 0. \end{aligned}$$

The path which led us to the definition of the augmented Lagrangian (9) can be adapted to this case thus resulting in the definition of the following exact augmented Lagrangian function:

$$\begin{aligned} L_a(x, \lambda; \epsilon) = {} & f(x) + \lambda^\top \max\{g(x), -\epsilon p(\lambda)\lambda\} + \frac{\|\max\{g(x), -\epsilon p(\lambda)\lambda\}\|^2}{2\epsilon p(\lambda)} \\ & + \|\nabla g(x)^\top \nabla_x L(x, \lambda) + G(x)^2\lambda\|^2. \end{aligned}$$

It is worth noting the differences between L_a and other exact augmented Lagrangian functions previously proposed for one-sided constraints [7, 14]. Indeed, in [7, 14], the compactness property of the level sets of the exact merit function is guaranteed by introducing some barrier term in x in the expression of the merit function itself. On the contrary, as we shall see in the following, it is possible to assure the compactness of the level sets of L_a by employing Assumption A1 which, as already mentioned, can be satisfied by appropriately scaling the constraint functions.

Under Assumptions A1–A3 (where in Assumption A2(b) we can set $r(x) = \max\{g(x), 0\}$) the study of the properties of L_a follows, with minor modifications, from the one which will be carried out in subsequent sections for L_b.

3. Preliminary properties of L_b

In this section we point out some properties of the function $L_b(x, \lambda; \epsilon)$.

Proposition 3.1. *For any value of* $\epsilon > 0$, *we have*:
(a) *at every* KKT *pair* $(\bar{x}, \bar{\lambda})$ *of Problem* (P):

$$L_b(\bar{x}, \bar{\lambda}; \epsilon) = f(\bar{x});$$

(b) *for all* $(x, \lambda) \in \mathcal{F} \times \mathbb{R}^m$ *it results*:

$$L_b(x, \lambda; \epsilon) \leq f(x) + \eta(x, \lambda); \tag{10}$$

(c) *for all* $(x, \lambda) \in \mathcal{S} \times \mathbb{R}^m$ *it results*:

$$L_b(x, \lambda; \epsilon) \geq f(x) - \frac{\epsilon}{2} + \frac{1}{2\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 + \eta(x, \lambda) \geq f(x) - \frac{\epsilon}{2}. \tag{11}$$

Proof: Point (a) immediately follows from Proposition 2.1 and expression (9) of L_b.

Point (b). From expression (9) of L_b we have that

$$L_b(x, \lambda; \epsilon) - f(x) - \eta(x, \lambda) = \sum_{i=1}^{m} \left[\lambda_i \gamma_i(x, \lambda; \epsilon p(\lambda)) + \frac{1}{2\epsilon p(\lambda)}\gamma_i(x, \lambda; \epsilon p(\lambda))^2\right]. \tag{12}$$

We show that, for $x \in \mathcal{F}$, the i-th term of the summation in (12) is non positive. Indeed, it can be rewritten as

$$\sigma_i = \frac{\gamma_i(x, \lambda; \epsilon p(\lambda))}{2\epsilon p(\lambda)}(2\epsilon p(\lambda)\lambda_i + \gamma_i(x, \lambda; \epsilon p(\lambda))).$$

Now, we have the following three cases:

(i) $-\epsilon p(\lambda)\lambda_i \leq g_i(x) - u_i \leq 0$, that is, $\gamma_i(x, \lambda; \epsilon p(\lambda)) = g_i(x) - u_i \leq 0$ and

$$\sigma_i = \frac{g_i(x) - u_i}{2\epsilon p(\lambda)}(2\epsilon p(\lambda)\lambda_i + g_i(x) - u_i) \leq 0;$$

(ii) $-\epsilon p(\lambda)\lambda_i \geq g_i(x) - l_i \geq 0$, that is, $\gamma_i(x, \lambda; \epsilon p(\lambda)) = g_i(x) - l_i \geq 0$ and

$$\sigma_i = \frac{g_i(x) - l_i}{2\epsilon p(\lambda)}(2\epsilon p(\lambda)\lambda_i + g_i(x) - l_i) \leq 0;$$

(iii) $\gamma_i(x, \lambda; \epsilon p(\lambda)) = -\epsilon p(\lambda)\lambda_i$ and $\sigma_i = -\frac{\epsilon p(\lambda)\lambda_i^2}{2} \leq 0.$

Hence we can conclude that the summation (12) is non positive and hence that $L_b(x, \lambda; \epsilon) \leq f(x) + \eta(x, \lambda)$ for every $x \in \mathcal{F}$ and $\lambda \in \mathbb{R}^m$.

Point (c). Recalling the expression (5) of $p(\lambda)$ we can write function $L_b(x, \lambda; \epsilon)$ as:

$$L_b(x, \lambda; \epsilon) = f(x) + \lambda^\top \gamma(x, \lambda; \epsilon p(\lambda)) + \frac{1}{2\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2(\|\lambda\|^2 + 1) + \eta(x, \lambda),$$

from which we obtain:

$$\begin{aligned} L_b(x, \lambda; \epsilon) \geq\ & f(x) - \|\lambda\|\|\gamma(x, \lambda; \epsilon p(\lambda))\| + \frac{1}{2\epsilon}\|\lambda\|^2\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 \\ & + \frac{1}{2\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 + \eta(x, \lambda). \end{aligned}$$

Now, taking into account that the quadratic form $-u + \frac{1}{2\epsilon}u^2$ attains its minimum value $-\frac{\epsilon}{2}$ when $u = \epsilon$, we get:

$$L_b(x, \lambda; \epsilon) \geq f(x) - \frac{\epsilon}{2} + \frac{1}{2\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 + \eta(x, \lambda) \geq f(x) - \frac{\epsilon}{2},$$

which proves (11). □

Now we put in evidence some interesting properties of the level sets of the augmented Lagrangian function L_b. Let us define

$$\Omega(x^\circ, \lambda^\circ; \epsilon) = \{(x, \lambda) \in \mathcal{S} \times \mathbb{R}^m : L_b(x, \lambda; \epsilon) \leq L_b(x^\circ, \lambda^\circ; \epsilon)\},$$

where $(x^\circ, \lambda^\circ) \in \mathcal{S} \times \mathbb{R}^m$.

From points (a) and (b) of Proposition 3.1, it easily follows a first property of the level set Ω. In fact, if a feasible point x° is known, it is possible to select properly λ° so that KKT points of Problem (P) belonging to the set $\Omega(x^\circ, \lambda^\circ; \epsilon)$ have an objective function value smaller than or equal to $f(x^\circ)$. In particular we have:

Proposition 3.2. *Let x° be a feasible point and λ° be a solution of*

$$\eta(x^\circ, \lambda) = \nabla g(x^\circ)^\top \nabla f(x^\circ) + M(x^\circ)\lambda = 0, \tag{13}$$

where $M(x)$ is given by (8). *Then, for every $\epsilon > 0$, any* KKT *pair $(\bar{x}, \bar{\lambda})$ of Problem* (P) *belonging to $\Omega(x^\circ, \lambda^\circ; \epsilon)$ is such that $f(\bar{x}) \leq f(x^\circ)$.*

It is easy to check that, under Assumption A3, a solution λ° of system (13) exists. Then, the proof follows from the expression of L_b and (10). Indeed, $L_b(x^\circ, \lambda^\circ; \epsilon) \le f(x^\circ)$ and $L_b(\bar{x}, \bar{\lambda}; \epsilon) = f(\bar{x})$.

In order to prove the results on the compactness of the level sets of L_b, we need the following technical lemma.

Lemma 3.3 [4]. *Let $\{s_k^{(i)}\}$, $i = 1, \ldots, p$, be p sequences of positive numbers. Then, there exists an index i^* and a sequence of integers $K = \{k_j\}$ such that*

$$\lim_{j\to\infty} \frac{s_{k_j}^{(i^*)}}{s_{k_j}^{(i)}} = t_i < +\infty, \quad i = 1, \ldots, p.$$

In particular $t_{i^*} = 1$.

In the next two propositions, we state the compactness properties of the level set Ω.

Proposition 3.4. *For every $\epsilon_M > 0$, there exists a compact set $\mathcal{C} \subset \mathcal{S}$ such that $\Omega(x^\circ, \lambda^\circ; \epsilon) \subseteq \mathcal{C} \times \mathbb{R}^m$ for all $\epsilon \in (0, \epsilon_M]$.*

Proof: Let

$$\Pi_x = \{x \in \mathbb{R}^n : (x, \lambda) \in \Omega(x^\circ, \lambda^\circ; \epsilon) \text{ for some } \epsilon \in (0, \epsilon_M]\}$$

and $\mathcal{C} = \bar{\Pi}_x$. We first show that $\mathcal{C}$ is compact. Since it is closed by definition, we show that $\mathcal{C}$ is a bounded set. By definition of Ω, we have that

$$\mathcal{C} \subseteq \bar{\mathcal{S}}. \tag{14}$$

We have to distinguish if either point (a) of Assumption A2 or point (b) of Assumption A2 holds.

If Assumption A2(b) holds, $\bar{\mathcal{S}}$ is compact, so that by (14) we get that $\mathcal{C}$ is bounded. Otherwise, if Assumption A2(a) holds, a feasible point $\hat{x}$ is known. Then, letting $x^\circ = \hat{x}$, by (10) and (11) we have, for every $\epsilon \in (0, \epsilon_M]$ that

$$f(x) - \frac{\epsilon_M}{2} \le L_b(x, \lambda; \epsilon) \le L_b(x^\circ, \lambda^\circ; \epsilon) \le f(x^\circ) + \eta(x^\circ, \lambda^\circ).$$

Now, by definition,

$$\mathcal{C} \subseteq \left\{x \in \bar{\mathcal{S}} : f(x) \le f(x^\circ) + \frac{\epsilon_M}{2} + \eta(x^\circ, \lambda^\circ)\right\},$$

which, taking into account that $f(x)$ is coercive on $\bar{\mathcal{S}}$, implies that $\mathcal{C}$ is compact.

Now we prove that $\mathcal{C} \subset \mathcal{S}$. Assume by contradiction that sequences $\{(x^k, \lambda^k)\}$ and $\{\epsilon^k\}$ exist such that

$$\begin{aligned}
&0 < \epsilon^k \le \epsilon_M, \quad \forall k,\\
&(x^k, \lambda^k) \in \Omega(x^\circ, \lambda^\circ; \epsilon^k),\\
&x^k \to \tilde{x} \in \partial\mathcal{S}.
\end{aligned}$$

By Assumption A1, the index set $J = J_+ \cup J_-$ with

$$J_+ = \{i : g_i(x^k) \to +\infty\}, \quad J_- = \{i : g_i(x^k) \to -\infty\},$$

is not empty. By Lemma 3.3 with $s_k^{(i)} = 1/|g_i(x^k)|$, an index i^* exists such that for every $i \in J$

$$\lim_{k\to\infty} \frac{|g_i(x^k)|}{|g_{i^*}(x^k)|} = t_i < \infty \quad (t_{i^*} = 1).$$

Recalling that $(x^k, \lambda^k) \in \Omega(x^\circ, \lambda^\circ; \epsilon^k)$ and the expression (9) of L_b, it results:

$$\begin{aligned}
\frac{(\epsilon^k)^2}{|g_{i^*}(x^k)|^2}\left(f(x^k) + \frac{1}{2\epsilon^k}\psi(x^k, \lambda^k; \epsilon^k)\right) &\le \frac{(\epsilon^k)^2}{|g_{i^*}(x^k)|^2} L_b(x^k, \lambda^k; \epsilon^k)\\
&\le \frac{(\epsilon^k)^2}{|g_{i^*}(x^k)|^2} L_b(x^\circ, \lambda^\circ; \epsilon^k).
\end{aligned}$$

Taking the limit for $k \to \infty$, recalling that $\{f(x^k)\}$ is bounded, we have

$$\lim_{k\to\infty} \frac{\epsilon^k}{2|g_{i^*}(x^k)|^2}\psi(x^k, \lambda^k; \epsilon^k) \le \lim_{k\to\infty} \frac{(\epsilon^k)^2}{|g_{i^*}(x^k)|^2} L_b(x^\circ, \lambda^\circ; \epsilon^k) = 0.$$

Now, taking into account that, for $i = 1, \dots, m$, $\{\epsilon^k p(\lambda^k)\lambda_i^k\}$ is a bounded sequence, we can write

$$\begin{aligned}
0 &\ge \lim_{k\to\infty} \frac{\epsilon^k}{|g_{i^*}(x^k)|^2}\psi(x^k, \lambda^k; \epsilon^k)\\
&= \lim_{k\to\infty} \frac{1}{p(\lambda^k)}\Bigg[\sum_{i\in\{1,\dots,m\}\setminus J} \frac{\epsilon^k}{|g_{i^*}(x^k)|^2}\left(p(\lambda^k)\lambda_i^k \gamma_i(x^k, \lambda^k; \epsilon^k) + \frac{1}{2\epsilon^k}\gamma_i(x^k, \lambda^k; \epsilon^k)^2\right)\\
&\quad + \sum_{i\in J_+} \frac{\epsilon^k}{|g_{i^*}(x^k)|^2}\left(p(\lambda^k)\lambda_i^k(g_i(x^k) - u_i) + \frac{1}{2\epsilon^k}(g_i(x^k) - u_i)^2\right)\\
&\quad + \sum_{i\in J_-} \frac{\epsilon^k}{|g_{i^*}(x^k)|^2}\left(p(\lambda^k)\lambda_i^k(g_i(x^k) - l_i) + \frac{1}{2\epsilon^k}(g_i(x^k) - l_i)^2\right)\Bigg]\\
&= \lim_{k\to\infty} (1 + \|\lambda^k\|^2) \sum_{i\in J} \frac{1}{2} t_i^2.
\end{aligned}$$

which contradicts the fact that $\sum_{i\in J} t_i > 0$ $(t_{i^*} = 1)$. □

We now prove the compactness of the level sets of L_b for every value of the penalty parameter ϵ.

Proposition 3.5. *For every $\epsilon > 0$ the level set $\Omega(x^\circ, \lambda^\circ; \epsilon)$ is compact.*

Proof: We first show that $\Omega(x^\circ, \lambda^\circ; \epsilon)$ is bounded. The proof is by contradiction. Therefore, we assume that there exists a sequence $\{x^k, \lambda^k\}$ such that $x^k \in \mathcal{C}$, $\|\lambda^k\| \to \infty$ and

$$L_b(x^k, \lambda^k; \epsilon) \leq L_b(x^\circ, \lambda^\circ; \epsilon), \tag{15}$$

where $\mathcal{C}$ is the compact set defined in Proposition 3.4. Since $x^k \in \mathcal{C}$, a subsequence exists, that we relabel $\{x^k, \lambda^k\}$, such that:

$$x^k \to \tilde{x}, \qquad \frac{\lambda^k}{\|\lambda^k\|} \to \tilde{\lambda}.$$

Taking into account the expression (5) of $p(\lambda)$, we have:

$$\lim_{k\to\infty} p(\lambda^k) = 0, \qquad \lim_{k\to\infty} \|\lambda^k\| p(\lambda^k) = 0, \tag{16}$$

$$\lim_{k\to\infty} \|\lambda^k\|^2 p(\lambda^k) = 1. \tag{17}$$

Now, dividing (15) by $\|\lambda^k\|^2$, recalling the expression of L_b and taking the limit for $k \to \infty$, by (16), we obtain:

$$\limsup_{k\to\infty} \frac{1}{\|\lambda^k\|^2} L_b(x^k, \lambda^k; \epsilon) \leq \limsup_{k\to\infty} \frac{\|\gamma(x^k, \lambda^k; \epsilon p(\lambda^k))\|^2}{2\epsilon \|\lambda^k\|^2 p(\lambda^k)} + \limsup_{k\to\infty} \left\| M(k) \frac{\lambda^k}{\|\lambda^k\|} \right\|^2 \leq 0,$$

which, by (17), yields

$$\lim_{k\to\infty} \gamma(x^k, \lambda^k; \epsilon p(\lambda^k)) = 0 \tag{18}$$

$$\lim_{k\to\infty} \left\| M(x^k) \frac{\lambda^k}{\|\lambda^k\|} \right\|^2 = \|M(\tilde{x})\tilde{\lambda}\|^2 = 0. \tag{19}$$

From (18) and the properties of $\gamma(x, \lambda; \epsilon)$ we get $\tilde{x} \in \mathcal{F}$. Moreover, from (19), we get $M(\tilde{x})\tilde{\lambda} = 0$ (with $\|\tilde{\lambda}\| = 1$). Hence, the matrix $M(\tilde{x})$ should be singular, but this contradicts Assumption A3. Therefore, we can conclude that for every $\epsilon > 0$, the level set $\Omega(x^\circ, \lambda^\circ; \epsilon)$ is bounded.

Now, we have to prove that $\Omega(x^\circ, \lambda^\circ; \epsilon)$ is also closed. To this aim we show that every limit point $(\tilde{x}, \tilde{\lambda})$ of every sequence $\{(x^k, \lambda^k)\} \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ belongs to $\Omega(x^\circ, \lambda^\circ; \epsilon)$. Suppose, by contradiction, that $\{(\tilde{x}, \tilde{\lambda})\} \notin \Omega(x^\circ, \lambda^\circ; \epsilon)$; then, by the definition of $\Omega(x^\circ, \lambda^\circ; \epsilon)$

and by continuity of L_b it results that $\tilde{x} \in \partial\mathcal{S}$ which contradicts Proposition 3.4, that is, $\Omega(x^\circ, \lambda^\circ; \epsilon) \subseteq \mathcal{C} \times \mathbb{R}^m$ where $\mathcal{C} \subset \mathcal{S}$. □

The fact, shown by Proposition 3.5, that the continuously differentiable function $L_b(x, \lambda; \epsilon)$ has compact level sets for every value of the penalty parameter ϵ is quite relevant. It implies, on the one hand, that L_b admits a global minimum point, and hence a stationary point, on $\mathcal{S} \times \mathbb{R}^m$; on the other hand, that any globally convergent unconstrained minimization algorithm, using only first order derivatives of the objective function, can be employed to compute the stationary points of L_b.

4. First order analysis

In this section we consider the relationships between KKT pairs of Problem (P) and stationary points of $L_b(x, \lambda; \epsilon)$. First we prove that, for any $\epsilon > 0$:

– every KKT pair of Problem (P) is a stationary point of L_b;
– every stationary point $(\bar{x}, \bar{\lambda})$ of L_b such that $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$ is a KKT pair of Problem (P).

Finally we prove that for sufficiently small values of ϵ, every stationary point $(\bar{x}, \bar{\lambda})$ of L_b is such that $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$ and, hence, that it is a KKT pair of Problem (P).

From the definition and under the differentiability assumptions on f and g, it follows that the function $L_b(x, \lambda; \epsilon)$ is a C^1 function for all $(x, \lambda) \in \mathcal{S} \times \mathbb{R}^m$. The gradient of L_b is obtained from (9) as:

$$\nabla_x L_b(x, \lambda; \epsilon) = \nabla_x L(x, \lambda) + \frac{1}{\epsilon p(\lambda)} \nabla g(x) \gamma(x, \lambda; \epsilon p(\lambda)) + Q(x, \lambda)\varphi(x, \lambda) \tag{20}$$

$$\nabla_\lambda L_b(x, \lambda; \epsilon) = \gamma(x, \lambda; \epsilon p(\lambda)) + \frac{1}{\epsilon} \|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 \lambda + 2M(x)\varphi(x, \lambda) \tag{21}$$

where

$$\begin{aligned} Q(x, \lambda) = 2\Bigg[& \nabla_x^2 L(x, \lambda) \nabla g(x) + \sum_{i=1}^m \nabla^2 g_i(x) \nabla_x L(x, \lambda) e_i^\top \\ & + 2\nabla g(x)(G(x) - U)(2G(x) - U - L)(G(x) - L)\Lambda \Bigg], \end{aligned} \tag{22}$$

$M(x)$ is given by (8) and $\varphi(x, \lambda)$ is defined in (6).

Proposition 4.1. *Let $(\bar{x}, \bar{\lambda})$ be a* KKT *pair of Problem* (P). *Then, for any $\epsilon > 0$, the pair $(\bar{x}, \bar{\lambda})$ is a stationary point of $L_b(x, \lambda; \epsilon)$.*

Proof: The proof is straightforward using (20) and (21). □

Proposition 4.2. *Let* $(\bar{x}, \bar{\lambda}) \in \mathcal{S} \times \mathbb{R}^m$ *be a stationary point of* $L_b(x, \lambda; \epsilon)$ *such that* $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$. *Then* $(\bar{x}, \bar{\lambda})$ *is a* KKT *pair for Problem* (P).

Proof: By using $\nabla_\lambda L_b(x, \lambda; \epsilon) = 0$ and $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$, we have

$$M(\bar{x})\varphi(\bar{x}, \bar{\lambda}) = 0.$$

Premultiplying by $\varphi(\bar{x}, \bar{\lambda})^\top$ and recalling the definition (8) of $M(x)$, we get

$$\varphi(\bar{x}, \bar{\lambda})^\top \begin{bmatrix} \nabla g(\bar{x}) \\ (G(\bar{x}) - U)(G(\bar{x}) - L) \end{bmatrix}^\top \begin{bmatrix} \nabla g(\bar{x}) \\ (G(\bar{x}) - U)(G(\bar{x}) - L) \end{bmatrix} \varphi(\bar{x}, \bar{\lambda}) = 0,$$

from which

$$\begin{bmatrix} \nabla g(\bar{x}) \\ (G(\bar{x}) - U)(G(\bar{x}) - L) \end{bmatrix} \varphi(\bar{x}, \bar{\lambda}) = 0.$$

Premultiplying again by $[\nabla_x L(\bar{x}, \bar{\lambda})^\top \vdots \bar{\lambda}^\top (G(\bar{x}) - U)(G(\bar{x}) - L)]$ we get

$$\varphi(\bar{x}, \bar{\lambda}) = 0. \tag{23}$$

Taking into account $\nabla_x L_b(\bar{x}, \bar{\lambda}; \epsilon) = 0$ and $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$, and using (20) and (23), we obtain

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0,$$

which, along with $\gamma(\bar{x}, \bar{\lambda}; \epsilon p(\bar{\lambda})) = 0$, yelds that $(\bar{x}, \bar{\lambda})$ is a KKT pair of Problem (P). □

In order to prove that a strictly positive value $\bar{\epsilon}$ of the penalty parameter exists such that every stationary point of the merit function $L_b(x, \lambda; \epsilon)$ corresponds to a KKT pair of Problem (P), we need the following technical proposition. As shown in [7], the proposition turns out to be also useful in the definition of an algorithm for the solution of Problem (P), and more precisely in the definition of an updating rule for the automatic adjustment of the penalty parameter.

Proposition 4.3. *There exists an* $\bar{\epsilon} > 0$ *such that for all* $\epsilon \in (0, \bar{\epsilon}]$ *and all* $(x, \lambda) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ *we have*

$$\|\nabla L_b(x, \lambda; \epsilon)\| \geq \|\gamma(x, \lambda; \epsilon p(\lambda))\|. \tag{24}$$

Proof: This proof is similar to that concerning an analogous property of the merit function in [7]. For the sake of completeness it is reported in Appendix A.2. □

Employing Proposition 4.3, we can state the following result, that together with Proposition 4.1, completes our analysis and allows us to establish a complete correspondence between stationary points of the augmented Lagrangian $L_b(x, \lambda; \epsilon)$ and KKT pairs of Problem (P).

Proposition 4.4. *There exists a positive number $\bar{\epsilon} > 0$ such that, for all $\epsilon \in (0, \bar{\epsilon}]$, if $(\bar{x}, \bar{\lambda}) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ is a stationary point of $L_b(x, \lambda; \epsilon)$, then $(\bar{x}, \bar{\lambda})$ is a* KKT *pair of Problem* (P).

To summarize, we have proved that for sufficiently small values of the penalty parameter ϵ, there exists a one-to-one correspondence between KKT pairs of Problem (P) and unconstrained stationary points of the new augmented Lagrangian function L_b in $\mathcal{S} \times \mathbb{R}^m$.

5. Second order analysis

In this section we assume that f and g_i, $i = 1, \ldots, m$ are three times continuously differentiable functions. Under these assumptions we perform an analysis of the second order properties of the augmented Lagrangian function L_b. This analysis allows us to prove additional exactness results, and provides the bases for the definition of algorithms which conciliate the global convergence with a superlinear convergence rate (see, e.g., [7]).

Under the differentiability assumption on f, g, it follows that L_b is an SC^1 function for all $(x, \lambda) \in \mathcal{S} \times \mathbb{R}^m$, that is a continuously differentiable function with a semismooth gradient (see [9]). Hence, we can define its generalized Hessian $\partial^2 L_b(x, \lambda; \epsilon)$, in Clarke's sense (see [2]). For the augmented Lagrangian function L_b it is possible to describe the structure of the generalized Hessian $\partial^2 L_b$ in a neighborhood of a KKT pair of Problem (P). To this aim we consider a partition of the index set $\{1, \ldots, m\}$ into the subsets A and $N = \{1, \ldots, m\} \setminus A$, and we partition the vectors and matrices accordingly. Then we introduce the following $(n + m) \times (n + m)$ symmetric matrix $H(x, \lambda; \epsilon, A)$, given block-wise by (for convenience we omit the arguments in the r.h.s):

$$H_{xx}(x, \lambda; \epsilon, A) = \nabla_x^2 L + \frac{1}{\epsilon p} \nabla g_A \nabla g_A^\top + 2\nabla_x^2 L \nabla g \nabla g^\top \nabla_x^2 L, \tag{25}$$

$$H_{\lambda\lambda}(x, \lambda; \epsilon, A) = -\epsilon p \begin{pmatrix} 0 & 0 \\ 0 & I_N \end{pmatrix} + 2M_N^2 \tag{26}$$

$$H_{x\lambda}(x, \lambda; \epsilon, A) = (\nabla g_A \quad 0) + 2\nabla_x^2 L \nabla g M_N, \tag{27}$$

where

$$M_N = \nabla g^\top \nabla g + \begin{pmatrix} 0 & 0 \\ 0 & (G - U)_N^2 (G - L)_N^2 \end{pmatrix}$$

and I_N is the identity matrix of order $|N|$. The following proposition states that, in a neighborhood of a KKT pair of Problem (P), the generalized Hessian $\partial^2 L_b(x, \lambda; \epsilon)$ can be described almost explicitly. In fact, by reasoning as in [7], we can state the following proposition whose proof is reported in Appendix A.3.

Proposition 5.1. *For every* KKT *pair* $(\bar{x}, \bar{\lambda})$ *of Problem* (P) *and for every* $\epsilon > 0$ *there exists a neighborhood* $\mathcal{B}$ *of* $(\bar{x}, \bar{\lambda})$ *such that, for all* $(x, \lambda) \in \mathcal{B}$, *it results* $\partial^2 L_b(x, \lambda; \epsilon) = \text{co}\{\partial^2_B L_b(x, \lambda; \epsilon)\}$ *with*

$$\partial^2_B L_b(x, \lambda; \epsilon) = \{H(x, \lambda; \epsilon, A) + K(x, \lambda; \epsilon, A) : A \in \mathcal{A}\},$$

where $\mathcal{A} = \{A : A_+(\bar{x}, \bar{\lambda}) \subseteq A \subseteq A_0(\bar{x})\}$, *and* $K(x, \lambda; \epsilon, A)$ *is a matrix such that* $\|K(x, \lambda; \epsilon, A)\| \leq \rho(x, \lambda)$, *with* $\rho(x, \lambda)$ *a nonnegative function such that* $\rho(\bar{x}, \bar{\lambda}) = 0$.

At a KKT pair where the strict complementarity holds, it results $A_+(\bar{x}, \bar{\lambda}) = A_0(\bar{x})$. In this case $\partial^2 L_b(\bar{x}, \bar{\lambda}; \epsilon)$ reduces to a singleton, and in a neighborhood of the KKT pair the generalized Hessian can be further characterized as stated in the following proposition.

Proposition 5.2. *For every* KKT *pair* $(\bar{x}, \bar{\lambda})$ *of Problem* (P) *where the strict complementarity holds, and for every* $\epsilon > 0$, *there exists a neighborhood* $\mathcal{B}$ *of* $(\bar{x}, \bar{\lambda})$ *such that, for all* (x, λ) *in* $\mathcal{B}$, L_b *is twice continuously differentiable, with Hessian matrix given by*:

$$\nabla^2 L_b(x, \lambda; \epsilon) = H(x, \lambda; \epsilon, A_0(\bar{x})) + K(x, \lambda; \epsilon, A_0(\bar{x})),$$

where H *and* K *are matrices like in Proposition* 5.1.

By employing the properties pointed out in Propositions 5.1 and 5.2, it is possible to establish some correspondences between second order stationary points of $L_b(x, \lambda; \epsilon)$ and points which satisfy second order optimality conditions for Problem (P).

The first result shows that every stationary point of L_b satisfying second order necessary conditions correspond to a point satisfying second order necessary conditions for Problem (P). In particular we have:

Proposition 5.3. *Let* $(\bar{x}, \bar{\lambda})$ *be a* KKT *pair of Problem* (P) *and let* $\epsilon > 0$ *be given. If a positive semidefinite matrix* $H \in \partial^2_B L_b(\bar{x}, \bar{\lambda}; \epsilon)$ *exists, then the pair* $(\bar{x}, \bar{\lambda})$ *satisfies the second order necessary conditions for Problem* (P):

$$y^\top \nabla^2_x L(\bar{x}, \bar{\lambda}) y \geq 0, \quad \forall y : \nabla g_{A_0(\bar{x})}(\bar{x})^\top y = 0.$$

We refer to [8] for the proof of the above proposition.

The next proposition proves that, for sufficiently small values of ϵ, KKT pairs of Problem (P) satisfying the strong second order sufficient condition (3) are strict local minimizers of L_b which satisfy also the second order sufficient optimality condition for SC^1 functions (see [13]).

Proposition 5.4. *Let* $(\bar{x}, \bar{\lambda})$ *be a* KKT *pair of Problem* (P) *which satisfies the strong second order sufficient condition* (3). *Then, there exists a number* $\bar{\epsilon} > 0$ *such that, for* *every* $\epsilon \in (0, \bar{\epsilon}]$, $(\bar{x}, \bar{\lambda})$ *is an isolated local minimum point of* $L_b(x, \lambda; \epsilon)$ *and all matrices in* $\partial^2 L_b(\bar{x}, \bar{\lambda}; \epsilon)$ *are positive definite.*

Proof: We refer to [7] for the relevant details of this proof. □

6. Optimality results

In this chapter, we complete the analysis of the exactness properties of the new augmented Lagrangian function $L_b(x, \lambda; \epsilon)$ by establishing the relationships between local or global solutions of Problem (P) and local or global unconstrained minimum points of the augmented Lagrangian function. All the proofs of this section follow, with minor modifications, from the ones given in [7].

First of all we recall (see, e.g., [11], p. 46) a basic definition.

Definition 6.1. Given a set $\mathcal{M}$, a nonempty subset $\mathcal{M}^* \subset \mathcal{M}$ is called an isolated set of $\mathcal{M}$ if there exists a closed set $\mathcal{E}$ such that $\mathring{\mathcal{E}} \supset \mathcal{M}^*$ and such that, if $x \in \mathcal{E} \setminus \mathcal{M}^*$, then $x \notin \mathcal{M}$.

Now we can state that points in an isolated compact set of local minimizers of Problem (P) correspond to unconstrained local minimizers of L_b.

Proposition 6.2. *Let* $\mathcal{M}^*(\bar{f})$ *an isolated compact set of local solution of Problem* (P), *corresponding to the objective function value* $\bar{f}$; *then, there exists a number* $\bar{\epsilon} > 0$ *such that, for all* $\epsilon \in (0, \bar{\epsilon}]$, *if* $\bar{x} \in \mathcal{M}^*(\bar{f})$ *and* $\bar{\lambda}$ *is its associated* KKT *multiplier*, $(\bar{x}, \bar{\lambda})$ *is an unconstrained local minimum point of* $L_b(x, \lambda; \epsilon)$.

We have that also the converse is true.

Proposition 6.3. *There exists a number* $\bar{\epsilon} > 0$ *such that, for all* $\epsilon \in (0, \bar{\epsilon}]$, *if* $(\bar{x}, \bar{\lambda}) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ *is an unconstrained local minimum point of* $L_b(x, \lambda; \epsilon)$, *then* $\bar{x}$ *is a local solution of* (P) *and* $\bar{\lambda}$ *is its associated* KKT *multiplier.*

To conclude this chapter we can establish a bijective correspondence between global minimum points of the augmented Lagrangian function $L_b(x, \lambda; \epsilon)$ and global solutions of Problem (P).

Proposition 6.4. *Suppose that the feasible set* $\mathcal{F}$ *is not empty. Then, there exists a number* $\bar{\epsilon} > 0$ *such that, for all* $\epsilon \in (0, \bar{\epsilon}]$, *if* $\bar{x}$ *is a global solution of Problem* (P) *and* $\bar{\lambda}$ *is its associated* KKT *multiplier, the pair* $(\bar{x}, \bar{\lambda})$ *is a global minimum point of* $L_b(x, \lambda; \epsilon)$ *on* $\mathcal{S} \times \mathbb{R}^m$, *and conversely.*

It is thus evident that the augmented Lagrangian function $L_b(x, \lambda; \epsilon)$ enjoys the properties stated in Definitions 1, 2 and 3 of [5], namely, it is "globally" exact on the open set $\mathcal{S} \times \mathbb{R}^m$.

On this basis we can define a solution algorithm for Problem (P) based on the unconstrained minimization of L_b, and on an automatic adjustment rule for the penalty parameter of the type proposed by Polak in [17].

A. Appendix

A.1. Constraints transformation

We show that the constraints of Problem (P) can be modified in order to assure the satisfaction of Assumption A1 without perturbing the feasible set and the satisfaction of Assumptions A2 and A3. We consider only one of the two choices for $\mathcal{S}$ and the corresponding modification of g proposed in Section 1; the same type of arguments can be applied to the second choice.

First we note that we can write

$$\tilde{\mathcal{F}} = \{x \in \mathbb{R}^n : \tilde{l} \leq \tilde{g}(x) \leq \tilde{u}\} = \{x \in \mathbb{R}^n : 0 \leq \tilde{g}(x) - \tilde{l} \leq \tilde{u} - \tilde{l}\}.$$

Let us consider the first modification, corresponding to

$$g_i(x) = \frac{\tilde{g}_i(x) - \tilde{l}_i}{(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x)))}, \quad i = 1, \ldots, m,$$

and $l_i = 0$ and $u_i = (\tilde{u}_i - \tilde{l}_i)/\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)$.

By definition, we have that $(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x))) > 0$ for every $x \in \mathcal{S}$. We prove separately the two implications $x \in \tilde{\mathcal{F}} \Rightarrow x \in \mathcal{F}$ and $x \in \mathcal{F} \Rightarrow x \in \tilde{\mathcal{F}}$.

Let us assume that $x \in \tilde{\mathcal{F}} \subset \mathcal{S}$. By definition, $g(x) \geq 0$. Moreover, for every $x \in \tilde{\mathcal{F}}$, we can write the following inequality

$$\begin{aligned}(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x))) &= \alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i) + (\tilde{g}_i - \tilde{l}_i)(\tilde{u}_i - \tilde{g}_i) \\ &\geq \alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i) > 0.\end{aligned}$$

Hence we can write

$$g_i(x) \leq \frac{\tilde{g}_i(x) - \tilde{l}_i}{\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)} \leq \frac{\tilde{u}_i - \tilde{l}_i}{\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)} = u_i, \quad i = 1, \ldots, m.$$

The second implication is proved by showing that $x \notin \tilde{\mathcal{F}} \Rightarrow x \notin \mathcal{F}$. Recalling that $x \in \mathcal{S}$, we have trivially that if $\tilde{g}_i(x) - \tilde{l}_i < 0$ then $g_i(x) < 0$, $i = 1, \ldots, m$. Moreover, if $\tilde{g}_i(x) > \tilde{u}_i$, we can write the inequality

$$\begin{aligned}&(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x))) \\ &\quad = \alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i) + (\tilde{g}_i - \tilde{l}_i)(\tilde{u}_i - \tilde{g}_i) < \alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i),\end{aligned}$$

and consequently we have

$$g_i(x) > \frac{\tilde{g}_i(x) - \tilde{l}_i}{\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)} > \frac{\tilde{u}_i - \tilde{l}_i}{\alpha_i(\alpha_i + \tilde{u}_i - \tilde{l}_i)} = u_i, \quad i = 1, \ldots, m.$$

Hence we get that $x \in \mathcal{F}$ if and only if $x \in \tilde{\mathcal{F}}$.

Now let us consider the satisfaction of Assumptions A2 and A3. We can write $\nabla g_i(x) = c_i(x)\nabla\tilde{g}_i(x)$ where

$$c_i(x) = \frac{\alpha_i^2 + \alpha_i(\tilde{u}_i - \tilde{l}_i) + (\tilde{g}_i(x) - \tilde{l}_i)^2}{(\alpha_i - (\tilde{g}_i(x) - \tilde{u}_i))^2(\alpha_i - (\tilde{l}_i - \tilde{g}_i(x)))^2} > 0,$$

that is the gradients of $\tilde{g}_i(x)$, $i = 1, \ldots, m$, are modified by positive functions. So that Assumptions A2 and A3 still hold.

A.2. Proof of Proposition 4.3

In order to prove Proposition 4.3, we need the following technical result.

Proposition A.1. *For every $\hat{x} \in \mathcal{F}$ there exist numbers $\epsilon(\hat{x}), \sigma(\hat{x})$, and $\delta(\hat{x})$ such that, for all $\epsilon \in (0, \epsilon(\hat{x})]$ and for all $(x, \lambda) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ such that $\|x - \hat{x}\| \leq \sigma(\hat{x})$ and $\|\nabla_\lambda L_b(x, \lambda; \epsilon)\| \leq \|\gamma(x, \lambda; \epsilon p(\lambda))\|$, it results*

$$\epsilon\|\nabla_x L_b(x, \lambda; \epsilon)\| \geq \delta(\hat{x})\|\gamma(x, \lambda; \epsilon p(\lambda))\|.$$

Proof: From Assumption A3, it follows that matrix $M(\hat{x})$ is positive definite, hence non singular. Let $\mathcal{B}$ be a neighborhood of $\hat{x}$ such that $M(x)$ is nonsingular for all $x \in \mathcal{B}$. If $\|\nabla_\lambda L_b(x, \lambda; \epsilon)\| \leq \|\gamma(x, \lambda; \epsilon p(\lambda))\|$, by (21) we can write

$$\left\| \gamma(x, \lambda; \epsilon p(\lambda)) + \frac{1}{\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2 + 2M(x)\varphi(x, \lambda) \right\| \leq \|\gamma(x, \lambda; \epsilon p(\lambda))\|.$$

Hence we get

$$2\|M(x)\varphi(x, \lambda)\| \leq 2\|\gamma(x, \lambda; \epsilon p(\lambda))\| + \frac{1}{\epsilon}\|\gamma(x, \lambda; \epsilon p(\lambda))\|^2\|\lambda\|,$$

from which, for $x \in \mathcal{B}$, we obtain

$$\|\varphi(x, \lambda)\| \leq \|M(x)^{-1}\| \left\{ 1 + \frac{1}{2\epsilon}\|\lambda\|\|\gamma(x, \lambda; \epsilon p(\lambda))\| \right\} \|\gamma(x, \lambda; \epsilon p(\lambda))\|. \quad (28)$$

Now, defining, for $i = 1, \ldots, m$,

$$b_i(x, \lambda; \epsilon) = \begin{cases} \epsilon p(\lambda)(g_i(x) - l_i)\lambda_i + (g_i(x) - u_i)(g_i(x) - l_i) & \text{if } g_i(x) + \epsilon p(\lambda)\lambda_i \geq u_i \\ 0 & \text{if } l_i < g_i(x) + \epsilon p(\lambda)\lambda_i < u_i \\ \epsilon p(\lambda)(g_i(x) - u_i)\lambda_i + (g_i(x) - u_i)(g_i(x) - l_i) & \text{if } g_i(x) + \epsilon p(\lambda)\lambda_i \leq l_i \end{cases}$$

and letting $B(x, \lambda; \epsilon) = \text{diag}\{b(x, \lambda; \epsilon)\}$, we can write

$$(G(x) - U)(G(x) - L)\lambda = \frac{1}{\epsilon p(\lambda)}[B(x, \lambda; \epsilon) - (G(x) - U)(G(x) - L)]\gamma(x, \lambda; \epsilon).$$

Recalling (20) and omitting the arguments to simplify the notation, we can write

$$\begin{aligned} \|\nabla g^\top \nabla_x L_b\| = \Big\| & \nabla g^\top \nabla_x L + \frac{1}{\epsilon p}\nabla g^\top \nabla g \gamma + \nabla g^\top Q \varphi \\ & + (G - U)^2 (G - L)^2 \lambda \\ & - \frac{1}{\epsilon p}[(G - U)(G - L)B - ((G - U)^2 (G - L)^2)]\gamma \Big\|. \end{aligned} \tag{29}$$

Multiplying (29) by ϵp we get

$$\epsilon p \|\nabla g^\top \nabla_x L_b\| = \|\tilde{M}\gamma + \epsilon p(I + \nabla g^\top Q)\varphi\| \tag{30}$$

where

$$\tilde{M}(x, \lambda; \epsilon) = M(x) - (G(x) - U)(G(x) - L)B(x, \lambda; \epsilon).$$

Therefore, employing (30) and (28), we have

$$\epsilon p \|\nabla g^\top \nabla_x L_b\| \geq \|\tilde{M}\gamma\| - \epsilon p \|I + \nabla g^\top Q\| \|M^{-1}\| \left\{1 + \frac{\|\lambda\|\|\gamma\|}{2\epsilon}\right\} \|\gamma\|,$$

from which we obtain

$$\begin{aligned} \epsilon\zeta \|\nabla_x L_b\| \geq \epsilon p \|\nabla g^\top \nabla_x L_b\| \geq & \\ \geq & \left[\sigma_m(\tilde{M}^\top \tilde{M})^{1/2} - \epsilon p \|I + \nabla g^\top Q\| \|M^{-1}\| \left\{1 + \frac{\|\lambda\|\|\gamma\|}{2\epsilon}\right\}\right] \|\gamma\|, \end{aligned} \tag{31}$$

where $\sigma_m(\tilde{M}^\top \tilde{M})$ is the smallest eigenvalue of $\tilde{M}^\top \tilde{M}$ and

$$\zeta = \max_{x \in \mathcal{C}} \|\nabla g(x)\|,$$

with $\mathcal{C}$ defined in Proposition 3.4. Now, due to the fact that by assumption $\hat{x} \in \mathcal{F}$, for every $\lambda \in \mathbb{R}^m$ we have $\tilde{M}(\hat{x}; \lambda; 0) = M(\hat{x})$. Therefore, $\tilde{M}(\hat{x}, \lambda; 0)$ is positive definite. Moreover, the term

$$\epsilon p \left\{ 1 + \frac{\|\lambda\| \|\gamma\|}{2\epsilon} \right\}$$

in (31) vanishes for $\epsilon = 0$ and $\hat{x} \in F$. By Proposition 3.4, points x such that $(x, \lambda) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$, belong to a compact set $\mathcal{C}$ which does not depend on ϵ. This and the expression of $p(\lambda)$ yield both that $p(\lambda)$ and $p(\lambda)\lambda$ are bounded. Moreover, $\gamma(x, \lambda; \epsilon p(\lambda))$ is continuous over $\Omega(x^\circ, \lambda^\circ; \epsilon)$ and it results that

$$\gamma(\hat{x}, \lambda; 0) = 0$$

whenever $\hat{x} \in \mathcal{F}$. Hence, the term

$$\frac{\|\lambda\| \|\gamma(x, \lambda; \epsilon p(\lambda))\|}{2}$$

tends to zero when $x \to \hat{x}$ and $\epsilon \to 0$. Therefore, it is always possible to find numbers $\epsilon(\hat{x}) > 0, \sigma(\hat{x}) > 0$ and $\delta(\hat{x}) > 0$ such that, for all $\epsilon \in (0, \epsilon(\hat{x})]$ and for all $(x, \lambda) \in \Omega(x^\circ, \lambda^\circ; \epsilon)$ satisfying $\|x - \hat{x}\| \leq \sigma(\hat{x})$ and $\|\nabla_\lambda L_b(x, \lambda; \epsilon)\| \leq \|\gamma(x, \lambda; \epsilon p(\lambda))\|$, it results:

$$\frac{1}{\eta} \left[\sigma_m (\tilde{M}^\top \tilde{M})^{1/2} - \epsilon p \|I + \nabla g^\top Q\| \|M^{-1}\| \left\{ 1 + \frac{\|\lambda\| \|\gamma\|}{2\epsilon} \right\} \right] \geq \delta(\hat{x}) > 0. \tag{32}$$

Now, the assertion is proved by considering that (31) and (32) yields

$$\epsilon \|\nabla_x L_b(x, \lambda; \epsilon)\| \geq \delta(\hat{x}) \|\gamma(x, \lambda; \epsilon p(\lambda))\|.$$ □

Proof of Proposition 4.3: The proof is by contradiction. Let us suppose that the assertion be false. In this case, subsequences $\{\epsilon^k\}$ and $\{(x^k, \lambda^k)\}$ would exist such that:

$$\epsilon^k \to 0, \tag{33}$$

$$(x^k, \lambda^k) \in \Omega(x^\circ, \lambda^\circ; \epsilon^k), \tag{34}$$

$$x^k \to \tilde{x} \in \mathcal{C}, \tag{35}$$

$$\|\nabla L_b(x^k, \lambda^k; \epsilon^k)\| < \|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\|. \tag{36}$$

From (36) we get the two following relations:

$$\|\nabla_\lambda L_b(x^k, \lambda^k; \epsilon^k)\| < \|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\|, \tag{37}$$

$$\epsilon^k p(\lambda^k) \|\nabla_x L_b(x^k, \lambda^k; \epsilon^k)\| < \epsilon^k p(\lambda^k) \|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\|. \tag{38}$$

Now, employing Proposition 3.4 and recalling the expressions of $p(\lambda)$ and $\gamma(x, \lambda; \epsilon p(\lambda))$ we have that

$$\lim_{k\to\infty} \epsilon^k p(\lambda^k)\|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\| = 0,$$

that, along with Eq. (38), implies

$$\lim_{k\to\infty} \epsilon^k p(\lambda^k)\|\nabla_x L_b(x^k, \lambda^k; \epsilon^k)\| = 0. \tag{39}$$

Again, by the expression of $p(\lambda)$, Proposition 3.4 and the continuity assumption, it results that the sequences $\{p(\lambda^k)\nabla g(x^k)\lambda^k\}$ and $\{p(\lambda^k)Q(x^k, \lambda^k)\varphi(x^k, \lambda^k)\}$ are bounded. Then, recalling (20) and taking the limit for $k \to \infty$, we obtain

$$0 = \lim_{k\to\infty} \epsilon^k p(\lambda^k)\nabla_x L_b(x^k, \lambda^k; \epsilon^k) = \nabla g(\tilde{x})\rho(\tilde{x}),$$

where $\rho(x)$ are given by (4). By Assumption A2(b), setting $r(\tilde{x}) = \rho(\tilde{x})$, we obtain $\tilde{x} \in \mathcal{F}$. On the other hand, if Assumption A2(a) holds true, namely if we assume $x^\circ \in \mathcal{F}$, by Proposition 3.1(b) and (c), we have, for every k:

$$\begin{aligned} f(x^k) - \frac{\epsilon^k}{2} + \frac{\|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\|^2}{2\epsilon^k} &\le L_b(x^k, \lambda^k; \epsilon^k) \\ &\le L_b(x^\circ, \lambda^\circ; \epsilon^k) \le f(x^\circ) + \eta(x^\circ, \lambda^\circ). \end{aligned}$$

By taking the limit for $k \to \infty$, and by the continuity assumption, we obtain

$$f(\tilde{x}) + \limsup_{k\to\infty} \frac{\|\gamma(x^k, \lambda^k; \epsilon^k p(\lambda^k))\|^2}{2\epsilon^k} \le f(x^\circ) + \eta(x^\circ, \lambda^\circ),$$

which, considering Proposition 3.4, implies $\rho(\tilde{x}) = 0$, so that again we have $\tilde{x} \in \mathcal{F}$. In conclusion, if Assumption A2 holds true, the sequence $\{x^k\}$ converges to a point $\tilde{x}$ which is feasible. This fact along with (33), (37) and Proposition A.1 imply that, for sufficiently large values of k, we get a contradiction with (36). □

A.3. Proof of Proposition 5.1

Let $(\bar{x}, \bar{\lambda})$ be a KKT pair of Problem (P). We consider a point (x, λ) in a neighborhood $\mathcal{B}$ of $(\bar{x}, \bar{\lambda})$ and a subsequence $\{(x^k, \lambda^k)\}$ converging to (x, λ) and such that the Hessian of L_b is defined for every (x^k, λ^k).

We recall that the generalized Hessian $\partial^2 L_b(x, \lambda; \epsilon)$ in Clarke's sense is the set of matrices given by:

$$\partial^2 L_b(x, \lambda; \epsilon) = \mathrm{co}\{\partial_B^2 L_b(x, \lambda; \epsilon)\},$$

where $\partial_B^2 L_b(x, \lambda; \epsilon) = \{W \in \mathbb{R}^{(n+m)\times(n+m)} : \exists\{(x^k, \lambda^k)\} \to (x, \lambda)$ with ∇L_b differentiable at (x^k, λ^k) and $\{\nabla^2 L_b(x^k, \lambda^k; \epsilon)\} \to W\}$.

We prove separately the two inclusions

$$\partial_B^2 L_b(x,\lambda;\epsilon) \subseteq \{H(x,\lambda;\epsilon,A)+K(x,\lambda;\epsilon,A) : A\in\mathcal{A}\},$$
$$\partial_B^2 L_b(x,\lambda;\epsilon) \supseteq \{H(x,\lambda;\epsilon,A)+K(x,\lambda;\epsilon,A) : A\in\mathcal{A}\}.$$

We note that ∇L_b is differentiable at (x^k,λ^k) whenever it occurs that:

1. for every index $i : g_i(x^k)-u_i \neq -\epsilon p(\lambda^k)\lambda_i^k$ and $g_i(x^k)-l_i \neq -\epsilon p(\lambda^k)\lambda_i^k$, or
2. for every index i such that $g_i(x^k)-u_i = -\epsilon p(\lambda^k)\lambda_i^k$ or $g_i(x^k)-l_i = -\epsilon p(\lambda^k)\lambda_i^k$ it results:

$$\nabla g_i(x^k) = 0 = -\epsilon\nabla_{\lambda_i}\big(p(\lambda^k)\lambda_i^k\big) = p(\lambda^k)\big(1-2\big(\lambda_i^k\big)^2\big).$$

Let us define the set of indices:

$$A^k = \big\{i : \big(g_i(x^k)-u_i+\epsilon p(\lambda^k)\lambda_i^k\big)\big(g_i(x^k)-l_i+\epsilon p(\lambda^k)\lambda_i^k\big)\geq 0\big\},$$
$$N^k = \{1,\ldots,m\}\setminus A^k.$$

From now on we omit the arguments to simplify the notation. By partitioning the vectors g, λ and γ according to A^k and N^k, we can rewrite ∇L_b as:

$$\nabla_x L_b = \nabla_x L + \frac{1}{\epsilon p}\nabla g_{A^k}\gamma_{A^k} - \nabla g_{N^k}\lambda_{N^k}^k + Q\varphi, \tag{40}$$

$$\nabla_{\lambda_{A^k}} L_b = \frac{1}{\epsilon}\big[\|\gamma_{A^k}\|^2 + (\epsilon p)^2\big\|\lambda_{N^k}^k\big\|^2\big]\lambda_{A^k}^k + \gamma_{A^k} + 2\big[\nabla g_{A^k}^\top\nabla g + \big((G-U)_{A^k}^2(G-L)_{A^k}^2 \,\vdots\, 0_{A^kN^k}\big)\big]\varphi \tag{41}$$

$$\nabla_{\lambda_{N^k}} L_b = \frac{1}{\epsilon}\big[\|\gamma_{A^k}\|^2 + (\epsilon p)^2\big\|\lambda_{N^k}^k\big\|^2\big]\lambda_{N^k}^k - \epsilon p\lambda_{N^k}^k + 2\big[\nabla g_{N^k}^\top\nabla g + \big(0_{N^kA^k} \,\vdots\, (G-U)_{N^k}^2(G-L)_{N^k}^2\big)\big]\varphi. \tag{42}$$

By differentiating (40)–(42), we obtain the Hessian of L_b at (x^k,λ^k), that can be written as

$$\nabla^2 L_b(x^k,\lambda^k;\epsilon) = H(x^k,\lambda^k;\epsilon,A^k) + K(x^k,\lambda^k;\epsilon,A^k),$$

where $H(x^k,\lambda^k;\epsilon,A^k)$ is given by (25)–(27) and $K(x^k,\lambda^k;\epsilon,A^k)$ is given by the summation of all the matrices whose elements contain as a factor either a component of γ_{A^k} or a component of $\lambda_{N^k}^k$ or a component of $\nabla_x L^k$. Since, for sufficiently large values of k (see, e.g., [10]), it results

$$A_+(\bar{x},\bar{\lambda}) \subseteq A^k \subseteq A_0(\bar{x}),$$

we have that $g_{A^k}(x^k) \to 0$ and $\lambda^k_{N^k} \to 0$, provided that $(x^k, \lambda^k) \to (\bar{x}, \bar{\lambda})$ and that, for sufficiently large values of k, $A^k \in \mathcal{A}$. These considerations imply that

$$\partial^2_B L_b(x, \lambda; \epsilon) \subseteq \{H(x, \lambda; \epsilon, A) + K(x, \lambda; \epsilon, A) : A \in \mathcal{A}\}.$$

Now we have to prove that also the opposite inclusion holds, that is, we have to show that for every choice of $A \in \mathcal{A}$ it is possible to find a sequence $\{(x^k, \lambda^k)\}$ converging toward $(\bar{x}, \bar{\lambda})$ such that:

$$A^k = A, \quad N^k = \{1, \ldots, m\} \setminus A,$$

and

$$\nabla^2 L_b(x^k, \lambda^k; \epsilon) = H(x^k, \lambda^k; \epsilon, A) + K(x^k, \lambda^k; \epsilon, A).$$

We denote by N the set $\{1, \ldots, m\} \setminus A$ and

$$\begin{aligned} A_1 &= A \cap A_+, & A_2 &= A \cap (A_0 \setminus A_+), \\ N_1 &= \{i \in N : l_i < g_i(\bar{x}) < u_i, \bar{\lambda}_i = 0\}, & N_2 &= N \cap (A_0 \setminus A_+); \end{aligned}$$

recalling the definition of A we have that:

$$A = A_1 \cup A_2, \qquad N = N_1 \cup N_2.$$

For every pair (x^k, λ^k) sufficiently close to $(\bar{x}, \bar{\lambda})$ it results that:

$$A^k \supseteq A_1, \qquad N^k \supseteq N_1.$$

To conclude the proof, we show that it is possible to further refine the choice of points (x^k, λ^k) in such a way that also the two inclusions:

$$A^k \supseteq A_2, \tag{43}$$

$$N^k \supseteq N_2 \tag{44}$$

are satisfied.

To this aim, let $\tilde{\delta} > 0$ be a number such that $|\bar{\lambda}_i| \leq \tilde{\delta}$, $i = 1, \ldots, m$. Then, we consider $\delta = \min_{i=1,\ldots,m}\{\frac{u_i - l_i}{2\epsilon}, \tilde{\delta}\}$. Since for every index $i \in A_2 \cup N_2$ we have either $g_i(\bar{x}) = u_i$ or $g_i(\bar{x}) = l_i$, we choose the subsequence $\{x^k\}$ in such a way to satisfy the following requirements:

$$\frac{2|g_i(x^k) - u_i|}{\epsilon} \leq \frac{\delta}{1 + m\tilde{\delta}^2} \qquad \forall i \in A_2 \cup N_2, g_i(\bar{x}) = u_i, \tag{45}$$

$$\frac{2|g_i(x^k) - l_i|}{\epsilon} \leq \frac{\delta}{1 + m\tilde{\delta}^2} \qquad \forall i \in A_2 \cup N_2, g_i(\bar{x}) = l_i. \tag{46}$$

Now we consider a sequence $\{\lambda^k\}$ converging to $\bar\lambda$ and such that:

$$\left|\lambda_i^k\right| \leq \tilde\delta \qquad i \in A_1 \cup N_1, \tag{47}$$

$$\lambda_i^k = \max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} > 0 \qquad i \in A_2,\, g_i(\bar x) = u_i, \tag{48}$$

$$\lambda_i^k = -\max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} < 0 \qquad i \in N_2,\, g_i(\bar x) = u_i, \tag{49}$$

$$\lambda_i^k = -\max\left\{\frac{2|g_i(x^k) - l_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} < 0 \qquad i \in A_2,\, g_i(\bar x) = l_i, \tag{50}$$

$$\lambda_i^k = \max\left\{\frac{2|g_i(x^k) - l_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} > 0 \qquad i \in N_2,\, g_i(\bar x) = l_i. \tag{51}$$

Employing (45), (48), the definition of $p(\lambda)$ and the fact that $|\lambda_i^k| \leq \tilde\delta$ for all $i = 1, \ldots, m$, we get that for all $i \in A_2$ and such that $g_i(\bar x) = u_i$:

$$\begin{aligned}-\lambda_i^k &= -\max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\}\\ &\leq -\max\left\{\frac{2|g_i(x^k) - u_i|(1 + \|\lambda^k\|^2)}{\epsilon}, \frac{\delta}{k}\right\} < -\frac{|g_i(x^k) - u_i|}{\epsilon p(\lambda^k)} \leq \frac{g_i(x^k) - u_i}{\epsilon p(\lambda^k)};\end{aligned}$$

whereas, for all $i \in A_2$ and such that $g_i(\bar x) = l_i$:

$$\begin{aligned}-\lambda_i^k &= \max\left\{\frac{2|g_i(x^k) - l_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} \geq \max\left\{\frac{2|g_i(x^k) - l_i|(1 + \|\lambda^k\|^2)}{\epsilon}, \frac{\delta}{k}\right\}\\ &> \frac{|g_i(x^k) - l_i|}{\epsilon p(\lambda^k)} \geq \frac{g_i(x^k) - l_i}{\epsilon p(\lambda^k)};\end{aligned}$$

which proves (43). Now we prove (44). By employing analogous reasoning, if we consider an index $i \in N_2$ and such that $g_i(\bar x) = u_i$ we get:

$$\begin{aligned}-\lambda_i^k &= \max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\} \geq \max\left\{\frac{2|g_i(x^k) - u_i|(1 + \|\lambda^k\|^2)}{\epsilon}, \frac{\delta}{k}\right\}\\ &> \frac{|g_i(x^k) - u_i|}{\epsilon p(\lambda^k)} \geq \frac{g_i(x^k) - u_i}{\epsilon p(\lambda^k)};\end{aligned}$$

hence

$$g_i(x^k) < u_i - \epsilon p(\lambda^k)\lambda_i^k.$$

By (49) we obtain

$$\begin{aligned}\lambda_i^k &= -\max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde\delta^2)}{\epsilon}, \frac{\delta}{k}\right\}\\ &\leq -\max\left\{\frac{2|g_i(x^k) - u_i|(1 + \|\lambda^k\|^2)}{\epsilon}, \frac{\delta}{k}\right\} < -\frac{|g_i(x^k) - u_i|}{\epsilon p(\lambda^k)} \leq \frac{g_i(x^k) - u_i}{\epsilon p(\lambda^k)};\end{aligned}$$

hence

$$u_i + \epsilon p(\lambda^k)\lambda_i^k < g_i(x^k) < u_i - \epsilon p(\lambda^k)\lambda_i^k.$$

Now we show that $l_i - \epsilon p(\lambda^k)\lambda_i^k \le u_i + \epsilon p(\lambda^k)\lambda_i^k$. By (49) and from the definition of δ we have that

$$0 < -\lambda_i^k = \max\left\{\frac{2|g_i(x^k) - u_i|(1 + m\tilde{\delta}^2)}{\epsilon}, \frac{\delta}{k}\right\} \le \delta \le \frac{u_i - l_i}{2\epsilon}.$$

This equation and the fact that $p(\lambda^k) \le 1$ for every k, implies $l_i - \epsilon p(\lambda^k)\lambda_i^k \le u_i + \epsilon p(\lambda^k)\lambda_i^k$ and, hence, $i \in N^k$. Now we show that

$$l_i - \epsilon p(\lambda^k)\lambda_i^k < g_i(x^k) < l_i + \epsilon p(\lambda^k)\lambda_i^k$$

for every index $i \in N_2$ and such that $g_i(\bar{x}) = l_i$. To this aim, we show that $l_i + \epsilon p(\lambda^k)\lambda_i^k \le u_i - \epsilon p(\lambda^k)\lambda_i^k$. In fact, from (51) and by the expression of δ we obtain:

$$0 < \lambda_i^k = \max\left\{\frac{2|g_i(x^k) - l_i|(1 + m\tilde{\delta}^2)}{\epsilon}, \frac{\delta}{k}\right\} \le \delta \le \frac{u_i - l_i}{2\epsilon}.$$

The preceding formula and the fact that $p(\lambda^k) \le 1$ for every k, imply $l_i + \epsilon p(\lambda^k)\lambda_i^k \le u_i - \epsilon p(\lambda^k)\lambda_i^k$ and, hence, that $i \in N^k$.

To summarize, we have shown that the sequence $\{(x^k, \lambda^k)\}$ converging to $(\bar{x}, \bar{\lambda})$ is such that:

$$A^k \supseteq A_1 \cup A_2 = A, \qquad N^k \supseteq N_1 \cup N_2 = N.$$

By noting that $\{A, N\}$ is a partition of the set of indices $\{1, \ldots, m\}$, we have

$$A^k = A, \qquad N^k = N,$$

which concludes the proof. □

References

1. D.P. Bertsekas, Constrained Optimization and Lagrange Multipliers Methods, Academic Press: New York, 1982.
2. F.H. Clarke, Optimization and Nonsmooth Analysis, John Wiley and Sons: New York, 1983.
3. G. Di Pillo and L. Grippo, "An augmented Lagrangian for inequality constraints in nonlinear programming problems," J. Optim. Theory and Appl., vol. 36, pp. 495–519, 1982.
4. G. Di Pillo and L. Grippo, "An exact penalty method with global convergence properties," Math Programming, vol. 36, pp. 1–18, 1986.
5. G. Di Pillo and L. Grippo, "Exact penalty functions in constrained optimization," SIAM J. Control and Optimization, vol. 27, pp. 1333–1360, 1989.

6. G. Di Pillo and S. Lucidi, “On exact augmented Lagrangian functions in nonlinear programming,” in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi (Eds.), Plenum Press: New York, pp. 85–100, 1996.
7. G. Di Pillo and S. Lucidi, “An augmented Lagrangian function with improved exactness properties,” SIAM J. Optimization, vol. 12, pp. 376–406, 2001.
8. G. Di Pillo, S. Lucidi, and L. Palagi, “Convergence to 2nd order stationary points of a primal-dual algorithm model for nonlinear programming,” TR 10-01, Department of Computer and Systems Science, University of Rome “La Sapienza”, Rome, Italy, 2001.
9. F. Facchinei, “Minimization of SC^1 functions and the Maratos effect,” Operations Research Letters, vol. 17, pp. 131–137, 1995.
10. F. Facchinei and S. Lucidi, “Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems,” J. Optim. Theory and Appl., vol. 85, pp. 265–289, 1995.
11. A.V. Fiacco and G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons: New York, 1969.
12. T. Glad and E. Polak, A multiplier method with automatic limitation of penalty growth, Math Programming, vol. 17, pp. 140–155, 1979.
13. D. Klatte and K. Tammer, “On second-order sufficient optimality conditions for $C^{1,1}$-optimization problems,” Optimization, vol. 19, pp. 169–179, 1988.
14. S. Lucidi, “New results on a class of exact augmented Lagrangians,” J. Optim. Theory and Appl., vol. 58, pp. 259–282, 1988.
15. S. Lucidi, “New results on a continuously differentiable exact penalty function,” SIAM J. Optimization, vol. 2, pp. 558–574, 1992.
16. O.L. Mangasarian and S. Fromowitz, “The Fritz-John necessary optimality conditions in the presence of equality constraints,” J. Math. Analysis and Appl., vol. 17, pp. 34–47, 1967.
17. E. Polak, “On the global stabilization of locally convergent algorithms,” Automatica, vol. 12, pp. 337–342, 1972.
18. R. Rockafellar, “The multiplier method of Hestenes and Powell applied to convex programming,” J. Optim. Theory and Appl., vol. 12, pp. 555–562, 1973.