Mathematical Programming manuscript No.
(will be inserted by the editor)

Dominikus Noll · Jean-Baptiste Thevenet

Augmented Lagrangian methods with smooth penalty functions

Received: date / Revised version: date

Abstract. Since the late 1990s, the interest in augmented Lagrangian methods has been revived, and several models with smooth penalty functions for programs with inequality constraints have been proposed, tested and used in a variety of applications. Global convergence results for some of these methods have been published. Here we present a local convergence analysis for a large class of smooth augmented Lagrangian methods based on spectral penalty functions. Our analysis shows that linear convergence in the neighborhood of a local minimum may be expected. Similar to the case of the Hestenes-Powell-Rockafellar augmented Lagrangian, this may be achieved without driving the penalty parameter to zero.

1. Introduction

We consider optimization programs of the form

$$\begin{aligned} \text{minimize} \quad & f(x), x \in \mathbb{R}^n \\ \text{subject to} \quad & g_i(x) = 0, i \in E \\ & g_i(x) \leq 0, i \in I \end{aligned} \tag{1}$$

D. Noll: Université Paul Sabatier, Institut de Mathématiques, 118, route de Narbonne, 31062 Toulouse, France - Email: `noll@mip.ups-tlse.fr`

J.-B. Thevenet: UPS-IM and ONERA-CERT, Control System Department, 2 av. Edouard Belin, 31055 Toulouse, France - Email : `thevenet@cert.fr`

Mathematics Subject Classification (1991): 20E28, 20G40, 20C20

where the program data are functions of class C^2, and E, I are finite. The Lagrangian associated with (1) is

$$L(x, \lambda) = f(x) + \sum_{i \in E \cup I} \lambda_i g_i(x) \tag{2}$$

where $\lambda = (\lambda_E, \lambda_I) \in \mathbb{R}^E \times \mathbb{R}^I_+$ is the Lagrange multiplier vector.

In order to solve (1), the classical augmented Lagrangian method uses a succession of unconstrained optimization programs, where at each step the augmented Lagrangian function

$$\begin{aligned} F(x, p, \lambda) = f(x) &+ \sum_{i \in E} \left(\lambda_i g_i(x) + \tfrac{1}{2p}\, g_i(x)^2 \right) \\ &+ \frac{p}{2} \sum_{i \in I} \left(\max\left\{0, \lambda_i + \tfrac{g_i(x)}{p}\right\}^2 - \lambda_i^2 \right) \end{aligned} \tag{3}$$

is minimized with respect to x. Let x^+ be the solution, then the penalty parameter is updated ($p^+ \leq p$) and the multipliers are modified according to:

$$\begin{aligned} \lambda_i^+ &= \lambda_i + \tfrac{1}{p} g_i(x^+), i \in E, \\ \lambda_i^+ &= \lambda_i + \tfrac{1}{p} \max\left\{-\lambda_i p, g_i(x^+)\right\} = \max\left\{0, \lambda_i + \tfrac{g_i(x^+)}{p}\right\}, i \in I, \end{aligned} \tag{4}$$

and the process is repeated. Ultimately, the solutions $x^+ = x^+(p, \lambda)$ of the unconstrained program $\min_{x \in \mathbb{R}^n} F(x, p, \lambda)$ are expected to converge to a solution x^* of (1), while the first-order multiplier updates (4) converge to an associated Lagrange multiplier λ^*. The role of local convergence theory is to analyze under what conditions this happens, and moreover, whether convergence may be achieved without driving the penalty parameter p to 0.

Since its creation by Hestenes [11] and Powell [15] in 1969, and its extension to inequality constraints proposed by Rockafellar [16] in 1972, the augmented

Lagrangian method has been one of the most prominent algorithmic tools in constrained programming. An excellent overview covering the period until the early 1990s is presented by Conn *et al.* in [8].

Given the fact that the last term on the right hand side of (3) is not of class C^2, several authors [9,2,5,3,7,17,1,4] have argued that using smooth penalty terms for the inequality constrained part $g_I(x) \leq 0$ might be numerically preferable, and would allow the use of second-order methods in the subproblems.

In this paper we present a local convergence theory for a promising class of smooth aumented Lagrangian, which was originally proposed by Ben-Tal *et al.* in [1] and [17], and which has been intensely studied by several groups since 1997. In [1], the authors already presented numerical tests for a variety of large scale convex problems, while nonconvex studies were added in [6]. The same smooth augmented Lagrangian method was successfully used for truss design problems by Kocvara *et al.* [12], and compared to other NLP solvers such as MINOS, LOQO and KNITRO. Extensions to matrix inequality constraints are presented in [13], and have led to the creation of the PENNON package.

In [4], a systematic comparison of several augmented Lagrangian methods on a collection of 173 problems from the Cute library was organized. While 80.92 % of these problems were solved by the traditional Powell-Hestenes-Rockafellar method (3), the method proposed in [1] based on (6) was successful in 78.03 % of the cases, while the method based on (7), proposed for the first time in [4], worked for 75.72 %. These experiments show the potential of smooth AL methods, in particular for large problems, and that it is time to thoroughly investigate their global and local convergence properties.

Global convergence for smooth augmented Lagrangian methods was already considered in [1] using convexity, and in [4] and [14] for general programs. In the present paper we focus on local convergence, which has not been investigated to date. We observe that the speed of convergence is linear, as in the case of the classical augmented Lagrangian. Moreover, we prove that local linear convergence may be achieved without driving the penalty parameter to zero. This is of the essence for the practical utility of the method, as an exceedingly small p leads to numerical ill-conditioning. The second class of methods (7) analyzed gives even superlinear convergence for inactive multipliers.

The structure of the paper is as follows. In Section 2 we describe the penalty functions F_1, F_2 examined in the sequel, introduce and comment on the algorithm, and prepare a few facts for the local convergence analysis. Section 3 establishes convergence for the algorithm with penalty function (6). In particular, Section 3.2 proves a complexity result which allows to prove that the penalty parameter is frozen, the key property of all succesfull augmented Lagrangian schemes. Penalty function (7) is studied in Section 4, while Section 5 concludes with an example illustrating the differences between the two classes of methods (6) and (7).

2. Preparation

2.1. Nonquadratic augmented Lagrangians for inequality constraints

The following scheme to generate smooth penalty functions is proposed by Ben-Tal *et al.* in [1,17]. Starting out with a function $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ of the form

$$\phi(t) = \begin{cases} t + \frac{1}{2}t^2, & t \geq -\frac{1}{2} \\ -\frac{1}{4}\log(-2t) - \frac{3}{8}, & t \leq -\frac{1}{2} \end{cases} \qquad \text{or} \qquad \phi(t) = \begin{cases} \frac{1}{1-t} - 1, & t < 1 \\ +\infty, & \text{else} \end{cases} \tag{5}$$

the augmented Lagrangian is defined as

$$F_1(x, p, \lambda) = f(x) + \sum_{i \in E} \left(\lambda_i g_i(x) + \tfrac{1}{2p}\, g_i(x)^2 \right) + p \sum_{i \in I} \lambda_i \phi\left(\tfrac{g_i(x)}{p} \right). \tag{6}$$

Other choices of the generator function ϕ are possible. Bertsekas [3, Section 4.2.5] already proposed the exponential method of multipliers $\phi(t) = e^t - 1$ and the modified barrier method based on $\phi(t) = -\log(1-t)$. An even larger variety of choices ϕ is discussed and compared in [4]. A common feature of all of these is that in a neighborhood of 0, they behave roughly like $\phi(t) = t + \frac{1}{2}t^2$. In other words, the right hand branch of ϕ is more or less the same as that of the usual augmented Lagrangian function (3). Smoothing rather concerns the negative branch.

In [4] several augmented Lagrangian models have been compared numerically. An interesting alternative to (6) is for instance

$$F_2(x, p, \lambda) = f(x) + \sum_{i \in E} \left(\lambda_i g_i(x) + \tfrac{1}{2p}\, g_i(x)^2 \right) + p \sum_{i \in I} \lambda_i^2 \phi\left(\tfrac{g_i(x)}{p\lambda_i} \right), \tag{7}$$

which is now non-linear in the multiplier vector λ_I. As we shall see, this has some advantages over (6).

Let us elaborate the first-order multiplier update rules for the smooth augmented Lagrangian functions (6) and (7), replacing the second part of (4). In the case of $F_1(x,p,\lambda)$ we find the rule

$$\lambda_i^+ = \lambda_i \phi' \left(\frac{g_i(x^+)}{p} \right), \quad i \in I, \tag{8}$$

whereas the use of $F_2(x,p,\lambda)$ leads to

$$\lambda_i^+ = \lambda_i \phi' \left(\frac{g_i(x^+)}{p\lambda_i} \right), \quad i \in I. \tag{9}$$

In particular, in the case of (9), the multiplier updates λ_I are maintained strictly positive at all times.

Proof. Let us assume that $\lambda_i > 0$. Then, using (9) and (5), we have that

$$\lambda_i^+ = \begin{cases} \lambda_i + \frac{g_i(x^+)}{p}, \; g_i(x^+) + \frac{p\lambda_i}{2} \geq 0 \\ -\frac{1}{4}\frac{\lambda_i^2}{g_i(x^+)}, \quad g_i(x^+) + \frac{p\lambda_i}{2} \leq 0 \end{cases}$$

In the first case, $\lambda_i^+ = \frac{1}{p}\left(g_i(x^+) + p\lambda_i\right) > \frac{1}{p}\left(g_i(x^+) + \frac{p\lambda_i}{2}\right) \geq 0$. In the second case, we have $g_i(x^+) < 0$. Then $\lambda_i^+ = -\frac{1}{4}\frac{\lambda_i^2}{g_i(x^+)} > 0$. So, as soon as λ_I is initialized with a strictly positive value, it remains so through the subsequent iterations. $\square$

For convenience we will rewrite (8) and (9) as $\lambda_I^+ = \phi'\left(g_I(x^+)/p\right)\lambda_I$, respectively $\lambda_I^+ = \phi'\left(g_I(x^+)/(p\lambda_I)\right)\lambda_I$, where $\phi'\left(g_I(x^+)/p\right)$, resp. $\phi'\left(g_I(x^+)/(p\lambda_I)\right)$, denote the diagonal matrices with entries $\phi'(g_i(x^+)/p)$, resp. $\phi'(g_i(x^+)/(p\lambda_i))$, $i \in I$. Those notations will be used systematically in the sequel.

The last element needed for the augmented Lagrangian scheme is a measure of progress to control the penalty parameter p. For equality constraints, following [2], we leave the penalty unchanged as soon as sufficient progress towards feasibility is made. That is, $\|g(x^+)\| \leq \tau\|g(x)\|$ for some fixed $0 < \tau < 1$. Here the progress measure is $\sigma(x) := \|g(x)\|$. Using Rockafellar's method of slacks, this generalizes to the standard augmented Lagrangian (3), where we leave $p^+ = p$ unchanged as soon as

$$\|g_E(x^+)\| + \|\max\{-\lambda_I p, g_I(x^+)\}\| \leq \tau\left(\|g_E(x)\| + \|\max\{-\lambda_I^- p^-, g_I(x)\}\|\right).$$

The progress measure is now $\sigma(x, p, \lambda) = \|g_E(x)\| + \|\max\{-\lambda_I p, g_I(x)\}\|$; see [2, Ch. 3]. This is in fact a primal-dual measure, as it includes information not only about the iterate x, but the Lagrange multiplier estimates and the penalty parameter p. In the case of (6) we shall use

$$\sigma_1(x, p, \lambda) := \|g_E(x)\| + \|\phi'(g_I(x)/p)\,\lambda_I - \lambda_I\|, \tag{10}$$

for (7) we shall use

$$\sigma_2(x, p, \lambda) = \|g_E(x)\| + \|\phi'(g_I(x)/(p\lambda_I))\,\lambda_I - \lambda_I\|. \tag{11}$$

Let us now recall the general scheme of the augmented Lagrangian algorithm.

<u>Augmented Lagrangian Algorithm</u>

1. Fix $0 < \tau < 1$ and $c > 1$. Produce an initial guess x of the solution and an initial Lagrange multiplier estimate $\lambda > 0$, an initial penalty parameter $p > 0$.
2. Given the Lagrange multiplier estimate $\lambda > 0$, the penalty parameter $p > 0$ and the current x, solve

   $(P_{p,\lambda}) \qquad \min_{x \in \mathbb{R}^n} F(x, p, \lambda)$

   possibly using x as the starting point for the inner iteration. Let x^+ be the solution.
3. Update multiplier and penalty as follows

   $\lambda_E^+ = \lambda_E + g_E(x^+)/p,$

   $\lambda_I^+ = \phi'\left(g_I(x^+)/p\right)\lambda_I$ respectively $\lambda_I^+ = \phi'\left(g_I(x^+)/(p\lambda_I)\right)\lambda_I$

   and

   $$p^+ = \begin{cases} p & \text{if } \sigma(x^+, p, \lambda) \leq \tau\sigma(x, p^-, \lambda^-) \\ p/c & \text{else} \end{cases}$$
4. Replace λ^+ by λ, p^+ by p, x^+ by x, and go back to step 2.

The mechanism follows that of the usual augmented Lagrangian algorithm. Driving the penalty parameter p to 0 makes the algorithm behave like a pure penalty method. However, when iterates get close to the neighborhood of attraction of a local minimum, x^*, it may be hoped that the multiplier update λ^+ takes its grip and generates iterates x^+ where $\sigma(x^+, p, \lambda) < \sigma(x, p^-, \lambda^-)$ improves significantly. The test in step 3 of the algorithm will then allow to freeze the penalty parameter at a decent positive value, avoiding ill-conditioning due to exceedingly small p. The role of the local convergence analysis below is to examine under what circumstances such a freezing of p will occur.

2.2. Notation

In an algorithm, x, λ, p denote the current data, x^+, λ^+ and p^+ those at the next sweep, x^-, λ^- and p^- those from the previous sweep.

We let $I_= \subset I$ be the set of active constraints at x^*, $I_< \subset I$ the set of the remaining inactive constraints. For a matrix (possibly a vector) M, the notation $M^=$ (respectively $M^<$) stands for the submatrix of M obtained by extracting the columns of M whose indicies are in $I_=$ (respectively $I_<$). Similarly $M_=$ and $M_<$ are obtained by selecting the corresponding rows.

By strict complementarity we have $\lambda_i^* > 0$ for $i \in I_=$, while of course $\lambda_i^* = 0$ for $i \in I_<$. The critical cone is

$$C(x^*) = \{d \in \mathbb{R}^n : d^T g_i'(x^*) = 0 \text{ for every } i \in I_=\}.$$

2.3. Wedge convergence

The following technical notion will be helpful in our convergence proof.

Definition 1. *The sequence* $(y^k, p_k, \lambda^k) \in \mathbb{R}^I \times \mathbb{R} \times \mathbb{R}^I$ *is said to wedge-converge to* (y, p, λ) *if* $y^k \to y$, $p_k \to p$, $\lambda^k \to \lambda$, *such that* $\frac{y_i^k - y_i}{p_k - p} \to 0$ *and* $\frac{\lambda_i^k - \lambda_i}{p_k - p} \to 0$ *for every* $i \in I$. *We shall use the notation* $(y_k, p_k, \lambda_k) \xrightarrow{w} (y, p, \lambda)$. *Similarly, we* *define wedge convergence* $(x_k, p_k, \lambda^k) \stackrel{w}{\to} (x, p, \lambda) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^I$ *and* $(p_k, \lambda^k) \stackrel{w}{\to} (p, \lambda) \in \mathbb{R} \times \mathbb{R}^I$.

Remark. Suppose $p_k \to 0$, $y^k \to y$, $\lambda^k \to \lambda$ and y, λ are complementary. Then wedge-convergence $(y^k, p_k, \lambda^k) \stackrel{w}{\to} (y, 0, \lambda)$ is equivalent to $\lambda_i^k y_i^k / p_k \to 0$ for every $i \in I$. Indeed, using $\lambda_i y_i = 0$, this follows from the identity

$$\frac{y_i^k \lambda_i^k}{p_k} = \frac{(y_i^k - y_i)\lambda_i^k}{p_k} + \frac{y_i(\lambda_i^k - \lambda_i)}{p_k}.$$

The following concept is also linked to wedge-convergence in those cases where $p_k \to 0$.

Definition 2. *For* $\epsilon > 0$*, define the wedge neighborhood* $\mathcal{W}(\epsilon)$ *of* $(x^*, 0, \lambda^*) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^I$ *as*

$$\mathcal{W}(\epsilon) = \{(x, p, \lambda) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^I : \|x - x^*\| \le p\epsilon, \|\lambda - \lambda^*\| \le p\epsilon, 0 \le p \le \epsilon\}.$$

2.4. Hypotheses

Let us now introduce a list of properties which the penalty functions $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ have to satisfy. We assume that $\mathrm{dom}(\phi) = (-\infty, b)$ for some $b > 0$, including the possibility $b = +\infty$.

(ϕ_1) ϕ is strictly convex, increasing and of class C^2 on $\mathrm{dom}(\phi)$.

(ϕ_2) $\phi(0) = 0$.

(ϕ_3) $\phi'(0) = 1$.

(ϕ_4) $t\phi'(t) = \mathcal{O}(1)$ as $t \to -\infty$.

(ϕ_5) $t^2\phi''(t) = \mathcal{O}(1)$ as $t \to -\infty$.

Remark. Some of these axioms have already been used in [1] and [17, 14, 13] to prove global convergence of the augmented Lagrangian method under convexity assumptions. Notice that $(\phi_1)-(\phi_5)$ are satisfied for several among the examples in [4],[3]. Let us just mention that (ϕ_2) and (ϕ_3) are local properties, which assure that the right hand branch of ϕ behaves roughly like $t + \frac{t^2}{2}$ in a neighborhood of 0. In contrast, (ϕ_4) and (ϕ_5) are global properties, which ensure that on the left hand branch $t < 0$, $p\phi(y/p) \to 0$ sufficiently fast as $p \to 0$.

In addition to axioms (ϕ_1) - (ϕ_5), we will constantly use the following hypotheses:

(H_1) The second-order sufficient optimality condition at (x^*, λ^*).

(H_2) Strict complementarity at (x^*, λ^*).

(H_3) The linear independence constraint qualification (LICQ).

3. Convergence for F_1

In this section we prove convergence of the augmented Lagrangian method for model (6). For simplicity we consider inequality constraints only. As a matter of fact, equality constraints are even easier to handle, see e.g. [2, Prop. 2.2.4].

3.1. First steps

Lemma 1. *Consider (P) with inequality constraints only, and let x^* be a local minimum of (P) such that (x^*, λ^*) is a KKT-pair with associated Lagrange multiplier $\lambda^* \geq 0$. Suppose axioms (ϕ_1) - (ϕ_5) and (H_1) - (H_3) are satisfied. Then* *there exists a wedge neighborhood $\mathcal{W}(\epsilon_1)$ of $(x^*,0,\lambda^*)$ and some $r>0$ such that $F_{xx}(x,p,\lambda) \succeq rI \succ 0$ for every $(x,p,\lambda) \in \mathcal{W}(\epsilon_1)$. Therefore we also have*

$$\|F_{xx}(x,p,\lambda)^{-1}\| \leq K_1 < \infty \tag{12}$$

for some $K_1 > 0$ and every $(x,p,\lambda) \in \mathcal{W}(\epsilon_1)$.

Proof. Suppose on the contrary that there exist sequences $x_k \to x^*$, $\lambda^k \to \lambda^*$ and $p_k \to 0$ such that $\|x_k - x^*\|/p_k \to 0$, $\|\lambda^k - \lambda^*\|/p_k \to 0$, but

$$d_k^T F_{xx}(x_k,p_k,\lambda^k) d_k \leq \delta_k \to 0$$

for certain unit vectors d_k. Passing to a subsequence if required, we may assume that $d_k \to d$ for a unit vector d. Observe that

$$\begin{aligned} d_k^T F_{xx}(x_k,p_k,\lambda^k)d_k = d_k^T L_{xx}\left(x_k, \phi'\left(g(x_k)/p_k\right)\lambda^k\right) d_k + \\ +p_k^{-1}(g'(x_k)^T d_k)^T \Lambda^k \phi''\left(g(x_k)/p_k\right)(g'(x_k)^T d_k), \end{aligned} \tag{13}$$

where Λ^k denotes the diagonal matrix with diagonal entries $\Lambda^k_{ii} = \lambda_i^k$, $\Lambda^k\phi''\left(g(x_k)/p_k\right)$ the diagonal matrix with entries $\lambda_i^k \phi''\left(g_i(x_k)/p_k\right)$.

Now observe that for $i \in I_=$ we have $g_i(x_k)/p_k = \left(g_i(x_k) - g_i(x^*)\right)/p_k \to 0$. This implies $\lambda_i^k\phi'\left(g_i(x_k)/p_k\right) \to \lambda_i^*\phi'(0) = \lambda_i^*$ for $i \in I_=$, using axiom (ϕ_3). Similarly, $\lambda_i^k\phi''\left(g_i(x_k)/p_k\right) \to \lambda_i^k\phi''(0)$ for $i \in I_=$.

On the other hand, for $i \in I_<$, we have $g_i(x_k) \leq -\kappa$ for some $\kappa > 0$, when x_k is sufficiently close to x^*. Hence $g_i(x_k)/p_k \to -\infty$, which proves $\phi'\left(g_i(x_k)/p_k\right) \to 0$ by axiom (ϕ_4), and $\phi''\left(g_i(x_k)/p_k\right) \to 0$ by axiom (ϕ_5). Therefore, as $\lambda_i^k \to \lambda_i^* = 0$ for $i \in I_<$, we have $\lambda_i^k\phi'\left(g_i(x_k)/p_k\right) \to 0$ and $\lambda_i^k\phi''\left(g_i(x_k)/p_k\right) \to 0$.

Altogether, we have proved that $\phi'\left(g(x_k)/p_k\right)\lambda^k \to \phi'(0)\lambda^* = \lambda^*$, and, noting $\Lambda^* = \mathrm{diag}(\lambda^*)$, that $\Lambda^k\phi''\left(g(x_k)/p_k\right) \to \Lambda^*\phi''(0)$. The first of these statements implies in particular that the first term on the right hand side of (13) converges to $d^T L''(x^*,\lambda^*)d$.

We have proved two things. Firstly, the first term on the right hand side of (13) converges as $k \to \infty$. Since the term on the left hand side of (13) converges, we infer that the second term on the right hand side of (13) converges as well. Now this term is of the form $p_k^{-1}\Xi_k$, where $p_k^{-1} \to \infty$. This clearly implies $\Xi_k \to 0$. But Ξ_k converges to $\Xi := \left(g'(x^*)^T d\right)^T \Lambda^*\phi''(0)\left(g'(x^*)^T d\right) = 0$. By the structure of the diagonal matrix Λ^*, $\phi''(0) > 0$ (since ϕ is assumed strictly convex), and strict complementarity, this implies $g_i'(x^*)^T d = 0$ for every $i \in I_=$. Consequently, d is a critical direction: $d \in C(x^*)$.

Going back with this information to (13), we infer that the first term on the right hand side of (13) converges to $d^T L_{xx}(x^*,\lambda^*)d$, which is strictly positive by the second order sufficient optimality condition, given that $d \in C(x^*)$ and $d \neq 0$. Since the second term on the right hand side of (13) is nonnegative by $\phi''(0) > 0$, this contradicts $\delta_k \to 0$. Altogether this proves indeed the existence of $\mathcal{W}(\epsilon_1)$ as claimed. $\square$

The following preparatory result uses similar arguments.

Lemma 2. *Under the same assumptions as in Lemma 1, there exists a constant $K_2 > 0$ and a wedge neighborhood $\mathcal{W}(\epsilon_2)$ of $(x^*, 0, \lambda^*)$ such that*

$$\|F_{x\lambda}(x,p,\lambda)\| \leq K_2$$

for every $(x, p, \lambda) \in \mathcal{W}(\epsilon_2)$.

Proof. Observe that $F_{x\lambda}(x, p, \lambda) = g'(x)\phi'(g(x)/p)$. By assumption $g_j(x)/p \to 0$ for $j \in I_=$, giving $\phi'(g_j(x)/p) \to 1$. On the other hand, $\phi'(g_j(x)/p) \to 0$ for $j \in I_<$ by axiom (ϕ_4). This shows $F_{x\lambda}(x, p, \lambda)$ is bounded on a wedge neighborhood $\mathcal{W}(\epsilon_2)$. □

Let us introduce the matrix

$$H(x^*, \lambda^*, 0, \lambda^*) = \begin{bmatrix} L_{xx}(x^*, \lambda^*) & [g'(x^*)]^= \\ \left[\Lambda^* \phi''(0) g'(x^*)^T\right]_= & 0 \end{bmatrix}, \tag{14}$$

where $\Lambda^* = \operatorname{diag}(\lambda^*)$. Then we have the following

Lemma 3. *Under the assumptions* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3), *the matrix* $H(x^*, \lambda^*, 0, \lambda^*)$ *is invertible.*

Proof. Let $[d, \mu]$ be a test vector in $\mathbb{R}^n \times \mathbb{R}^{I_=}$ such that

$$H(x^*, \lambda^*, 0, \lambda^*) \begin{bmatrix} d \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We need to show $[d, \mu] = [0, 0]$. First observe that the second line in (14) gives $[\Lambda^* \phi''(0) g'(x^*)^T]_= \, d = 0$, which by strict complementarity and $\phi''(0) > 0$ is the same as $[[g'(x^*)]^=]^T d = 0$. Altogether we have shown that either $d = 0$ or if $d \neq 0$, then necessarily $d \in C(x^*)$.

Clearly $d = 0$, when substituted in the first line of (14), implies $[g'(x^*)]^= \mu = 0$, which in explicit form reads $\sum_{i \in I_=} \mu_i g_i'(x^*) = 0$. Linear independence of the active constraint gradients implies $\mu = 0$. In this case we are done. So let us now consider the case where $d \neq 0$, but $d \in C(x^*)$. Then the first line of (14) reads

$L_{xx}(x^*, \lambda^*)d + [g'(x^*)]^= \mu = 0$. Multiplying from the left with $d \in C(x^*)$ gives $d^T L_{xx}(x^*, \lambda^*)d + d^T[g'(x^*)]^= \mu = d^T L_{xx}(x^*, \lambda^*)d = 0$, a contradiction because as a nonzero critical direction, d ought to satisfy $d^T L_{xx}(x^*, \lambda^*)d > 0$ by second order optimality. This proves invertibility of $H(x^*, \lambda^*, 0, \lambda^*)$. □

Let us now consider matrices of the form

$$H(x, \lambda^+, p, \lambda) = \begin{bmatrix} L_{xx}(x, \lambda^+) & [g'(x)]^= \\ \left[\Lambda\phi''\left(g(x)/p\right) g'(x)^T\right]_= & -pI \end{bmatrix} \tag{15}$$

for $x \in \mathbb{R}^n$, $\lambda^+ \in \mathbb{R}^I$, $p > 0$ and $\lambda \in \mathbb{R}^I$, where $\Lambda = \mathrm{diag}(\lambda)$. Then:

Lemma 4. *Assuming* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3)*, there exists a wedge neighborhood* $\mathcal{W}(\epsilon_3)$ *of* $(x^*, 0, \lambda^*)$*, a neighborhood* U_0 *of* λ^**, and a constant* $K_3 > 0$ *such that* $\|H(x, \lambda^+, p, \lambda)^{-1}\| \leq K_3 < \infty$ *whenever* $(x, p, \lambda) \in \mathcal{W}(\epsilon_3)$ *and* $\lambda^+ \in U_0$*.*

Proof. It suffices to show that if $\|\lambda^k - \lambda^*\|/p_k \to 0$, $\|x_k - x^*\|/p_k \to 0$, $p_k \to 0$ and $\lambda_k^+ \to \lambda^*$, then $H(x_k, \lambda_k^+, p_k, \lambda^k) \to H(x^*, \lambda^*, 0, \lambda^*)$, because the limiting matrix $H(x^*, \lambda^*, 0, \lambda^*)$ is invertible by Lemma 3. This, however, is clear because, as we have seen in the proof of Lemma 1 already, $\Lambda^k\phi''\left(g(x_k)/p_k\right) \to \Lambda^*\phi''(0)$.
□

Putting $\epsilon = \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$, we get a wedge neighborhood $\mathcal{W}(\epsilon)$ which guarantees the properties of Lemma 1, 2 and 4. Let us now consider the system of nonlinear equations

$$F_x(x, p, \lambda) = 0,$$

then (x^*, p, λ^*) is solution for every $p > 0$. By Lemma 1 we have $\|F_{xx}(x, p, \lambda)^{-1}\| \leq K_1 < \infty$ for all $(x, p, \lambda) \in \mathcal{W}(\epsilon)$, which invites using the implicit function theorem.

Let us fix $0 < p_1 < p_2 \leq \epsilon$ and consider the compact interval $\mathcal{I} = [p_1, p_2]$. We apply the implicit function theorem under the specific form Lemma 13 given in Appendix: There exists an open neighborhood $\mathcal{N}_{p_1,p_2}$ of the set $[p_1, p_2] \times \{\lambda^*\}$ in $\mathbb{R} \times \mathbb{R}^I$, an open neighborhood $\mathcal{M}_{p_1,p_2}$ of $\{x^*\} \times [p_1, p_2] \times \{\lambda^*\}$ with $\mathcal{M}_{p_1,p_2} \subset \mathcal{W}(\epsilon)$, together with a function $x^+(\cdot, \cdot) : \mathcal{N}_{p_1,p_2} \to \mathbb{R}^n$ of class C^1 on $\mathcal{N}_{p_1,p_2}$ such that

$$F_x\left(x^+(p, \lambda), p, \lambda\right) = 0 \tag{16}$$

for all $(p, \lambda) \in \mathcal{N}$ and such that $x^+(p, \lambda^*) = x^*$ for all $p \in \mathcal{I} = [p_1, p_2]$. Moreover, this function is unique in the sense that $(x^+, p, \lambda) \in \mathcal{M}_{p_1,p_2}$ and $F_x(x^+, p, \lambda) = 0$ if and only if $x^+ = x^+(p, \lambda)$ and $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$. In other words,

$$\{(x^+, p, \lambda) \in \mathcal{M}_{p_1,p_2} : F_x(x^+, p, \lambda) = 0\} = \{\left(x^+(p, \lambda), p, \lambda\right) : (p, \lambda) \in \mathcal{N}_{p_1,p_2}\}.$$

In fact, Lemma 13 is applied to $\Psi = F_x$, the compact set K^* in question is $[p_1, p_2] \times \{\lambda^*\}$, and $y = (p, \lambda)$, while x is x^+. During the following, with a slight abuse of notation, we shall use the expression $x^+(p, \lambda)$ for the ensemble of these implicit functions. We shall say that a particular $x^+(p, \lambda)$ is associated with the choice of an interval $\mathcal{I} = [p_1, p_2]$.

Notice that we may arrange that $\mathcal{M}_{p_1,p_2} = \mathcal{U}_{p_1,p_2} \times \mathcal{N}_{p_1,p_2}$ for an open neighborhood $\mathcal{U}_{p_1,p_2}$ of x^*. In the same vein, we may have $\mathcal{N}_{p_1,p_2} = I_{p_1,p_2} \times N_{p_1,p_2}$

where N_{p_1,p_2} is an open neighborhood of λ^*, I_{p_1,p_2} an open interval containing $\mathcal{I} = [p_1, p_2]$. Finally, as part of the implicit function theorem, we have the following formulae for the derivatives of $x^+(p,\lambda)$ with respect to λ and p:

$$(x^+)_\lambda(p,\lambda) = -F_{xx}\left(x^+(p,\lambda), p, \lambda\right)^{-1} F_{x\lambda}\left(x^+(p,\lambda), p, \lambda\right), \tag{17}$$

$$(x^+)_p(p,\lambda) = -F_{xx}\left(x^+(p,\lambda), p, \lambda\right)^{-1} F_{xp}\left(x^+(p,\lambda), p, \lambda\right). \tag{18}$$

Let us now fix $p_2 \le \epsilon$, and consider $0 < p = p_1 < p_2$ as mobile. Then we may may without loss assume that $\mathcal{U}_{p,p_2} \subset \mathcal{U}_{p',p_2}$ and $N_{p,p_2} \subset N_{p',p_2}$ for $0 < p < p' < p_2$. On the other hand, I_{p,p_2} will typically shrink at the right end p_2, but of course grow at the left end p, so there is no monotonicity. We have the following very helpful

Lemma 5. *Assume* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3). *Then we have the following unicity results:*

(1) Let $0 < p_1 < p_2$. *Suppose* $(p,\lambda) \in \mathcal{N}_{p_1,p_2}$. *Then* $x^+(p,\lambda)$ *is the unique local minimum (even the unique critical point) of program* $(P_{p,\lambda})$ *in the neighborhood* $\mathcal{U}_{p_1,p_2}$ *of* x^*.

(2) Suppose $0 < p' < p_2$ *and* $0 < p'' < p_2$. *Let* $(p,\lambda) \in \mathcal{N}_{p',p_2} \cap \mathcal{N}_{p'',p_2}$. *Then the two implicit functions* $x^+(\cdot,\cdot)$ *associated with* $[p', p_2]$ *and* $[p'', p_2]$ *give the same value* $x^+(p,\lambda)$.

Proof. Both results follow from the unicity part of the implicit function theorem Lemma 13. Indeed, in item (1) it suffices to recall that if $F_x(x^+, p, \lambda) = 0$ and $(x^+, p, \lambda) \in \mathcal{M}_{p_1,p_2}$, then $x^+ = x^+(p,\lambda)$ for the implicit function associated with $[p_1, p_2]$. Now if $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$ and $x^+ \in \mathcal{U}_{p_1,p_2}$, then we have $(x^+, p, \lambda) \in \mathcal{M}_{p_1,p_2} = \mathcal{U}_{p_1,p_2} \times \mathcal{N}_{p_1,p_2}$. Therefore, $x^+ = x^+(p, \lambda)$ as claimed.

The proof of (2) is based on the same argument. □

Remark. This result justifies our notation $x^+(\cdot, \cdot)$, where we suppress dependency on the interval $\mathcal{I}_{p,p_2}$.

Remark. By construction of the neighborhoods $\mathcal{M}_{p_1,p_2} \subset \mathcal{W}(\epsilon_1)$ and $\mathcal{N}_{p_1,p_2}$, we observe that whenever $\left(x^+(p_k, \lambda^k), p_k, \lambda^k\right)$ is defined for some implicit function and (p_k, λ^k) wedge converges to $(0, \lambda^*)$, then $\left(x^+(p_k, \lambda^k), p_k, \lambda^k\right)$ wedge converges to $(x^*, 0, \lambda^*)$. Due to unicity in Lemma 5, this holds regardless of the choice of the implicit function.

Lemma 6. *Under the assumptions* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3)*, there exists a constant* $K_4 > 0$ *such that for fixed* $p_2 \leq \epsilon$*, and for every interval* $\mathcal{I} = [p_1, p_2]$ *and implicit function* $x^+(\cdot, \cdot)$ *associated with* $\mathcal{I}$*, we have*

$$\|(x^+)_\lambda(p, \lambda)\| \leq K_4 \tag{19}$$

for every $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$*. In other words, the constant* K_4 *is independent of the choice* $\mathcal{I} = [p_1, p_2]$ *as long as* $p_2 \leq \epsilon$*.*

Proof. Observe that $\|F_{xx}^{-1}\|$ is bounded by K_1 on $\mathcal{W}(\epsilon)$ by Lemma 1, while $\|F_{x\lambda}\|$ is bounded by K_2 on $\mathcal{W}(\epsilon)$ by Lemma 2.

Now observe that by construction, we have $(x^+(p, \lambda), p, \lambda) \in \mathcal{W}(\epsilon)$ whenever $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$. This is because we arranged that $\mathcal{M}_{p_1,p_2} \subset \mathcal{W}(\epsilon)$ for the open sets arising in the implicit function theorem. Formula (17) then proves that $\|(x^+)_\lambda\|$ is bounded by $K_1 K_2 =: K_4$. □

Remark. Let us now take a time and indicate the principal idea of our local convergence proof. At some stage we will have to fix $\mathcal{I} = [p_1, p_2]$ by fixing the lower index p at some value p_1. This will obviously fix the implicit function $x^+(\lambda, p)$, and the neighborhoods $\mathcal{U}_{p_1,p_2}$ and N_{p_1,p_2}. However, before doing this, we will have to collect information which depends only on the wedge neighborhood $\mathcal{W}(\epsilon)$. For instance, constant K_4 in Lemma 4 is of this type. It is universal for *all* the implicit functions $x^+(p, \lambda)$, regardless of the choice of $\mathcal{I}_{p_1,p_2}$, as long as $p_2 \leq \epsilon$.

The rationale of this will become clear. We have to provide those constants beforehand, because our choice of the lower $p = p_1$ will depend on them. This mechanism will become clear in Lemma 10. □

Let us introduce the multiplier update function:

$$\lambda^+(p, \lambda) := \phi' \left(g \left(x^+(p, \lambda)/p \right) \right) \lambda, \tag{20}$$

defined on $\mathcal{N}_{p_1,p_2}$. As λ^+ is defined with the help of $x^+(p, \lambda)$, its domain $\mathcal{N}_{p_1,p_2}$ depends on the choice of $0 < p_1 < p_2 \leq \epsilon$, just as in the case of $x^+(p, \lambda)$. But we have the following

Lemma 7. *Under the assumptions* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3)*, the Lipschitz constant of* $x^+ := x^+(p, \lambda)$ *with respect to the second variable* λ *is independent of the choice of* $\mathcal{I} = [p_1, p_2]$*, as long as* $p_2 \leq \epsilon$*. The same is true for the Lipschitz constant of* $\lambda^+(p, \lambda)$ *with respect to* λ*.*

Proof. 1) Indeed, in view of (19), the Lipschitz constant of $x^+ := x^+(p, \lambda)$ is the same for every $\mathcal{N}_{p_1,p_2}$, as long as $p_2 \leq \epsilon$ is respected.

2) Let us now prove the statement for $\lambda_<^+$, the inactive part of λ^+. We first differentiate (20) with respect to λ:

$$(\lambda^+)_\lambda = \phi'\left(g(x^+)/p\right) + p^{-1}\Lambda\,\phi''\left(g(x^+)/p\right) g'(x^+)^T\,(x^+)_\lambda. \tag{21}$$

Now for $j \in I_<$ we have $g_j(x^+)/p \to -\infty$, so the first term on the right hand side of (21) is bounded by axiom (ϕ_4). As for the second term, again for inactive j, we have from axiom (ϕ_5) that $\phi''\left(g_j(x^+)/p\right) \to 0$, while $\lambda_j/p = (\lambda_j - \lambda_j^*)/p \to 0$ by wedge convergence. Since the remaining terms $g'(x^+)$ and $(x^+)_\lambda$ are bounded, the latter by (19), we conclude that $(\lambda_<^+)_\lambda$ is bounded.

3) To prove the result for the active part, $\lambda_=^+$, we argue as follows. Observe the following formula, whose classical analogue is well-known:

$$F_x(x^+, p, \lambda) = L_x(x^+, \lambda^+),$$

hence $L_x\left(x^+, \lambda^+\right) = 0$ by (16). Differentiating this relation with respect to λ gives

$$L_{xx}\left(x^+, \lambda^+\right)(x^+)_\lambda + g'(x^+)(\lambda^+)_\lambda = 0.$$

Splitting the second term in its $I_=$ and $I_<$ parts gives:

$$\left[g'(x^+)\right]^= (\lambda_=^+)_\lambda = -L_{xx}\left(x^+, \lambda^+\right)(x^+)_\lambda - \left[g'(x^+)\right]^< (\lambda_<^+)_\lambda.$$

From what we have seen in part 2) of this proof, the last term on the right hand side is bounded. As for the first term on the right hand side, given that x^+ and λ^+ are bounded on a wedge neighborhood and using (19), this term is bounded as well. This proves that $[g'(x^+)]^= (\lambda_=^+)_\lambda$ is bounded.

What we need, though, is boundedness of $(\lambda_{=}^{+})_\lambda$, and this follows as soon as we show that the operator $[g'(x^+)]^=$ is injective. This is now where the LICQ (hypothesis (H_3)) comes in. Indeed, it says that $[g'(x^*)]^=$ is of maximal rank. Since $x^+ \to x^*$ under wedge convergence, we are done, because $[g'(x^+)]^=$ is then of maximal rank in the neighborhood of $(0, \lambda^*)$. $\square$

Let us now define the matrix $H(p, \lambda)$ as

$$H(p, \lambda) = H\left(x^+(p, \lambda), \lambda^+(p, \lambda), p, \lambda\right), \tag{22}$$

where the right hand side is defined in (15). Again the definition depends on a choice of $\mathcal{I} = [p_1, p_2]$, because the implicit functions $x^+(p, \lambda)$ and $\lambda^+(p, \lambda)$ are used.

Lemma 8. *Under the standing assumptions* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3)*, there exists a constant* $K_5 > 0$ *such that*

$$\|H(p, \lambda)^{-1}\| \leq K_5$$

for every $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$ *and for all implicit functions* $x^+(p, \lambda)$*,* $\lambda^+(p, \lambda)$ *associated with any* $\mathcal{I}_{p_1} = [p_1, p_2]$*, provided that* $p_2 \leq \epsilon$*.*

Proof. This follows from Lemma 4, as soon as we show that $(x^+(p, \lambda), p, \lambda) \in \mathcal{W}(\epsilon)$ and $\lambda^+(p, \lambda) \in U_0$, where U_0 appears in Lemma 4. Now according to the choice of $\mathcal{M}_{p_1,p_2}$ in the implicit function theorem, the first of these properties is guaranteed. As for the second property, notice that by Lemma 7, λ^+ is Lipschitz with the same Lipschitz constant on each $\mathcal{N}_{p_1,p_2}$ with $p_2 \leq \epsilon$. So $\lambda^+\left(\mathcal{N}_{p_1,p_2}\right) \subset U_0$ may be arranged for all p_1 simultaneously. What matters is $p_2 \leq \epsilon$. $\square$

We are now ready to prove the following major step towards the local convergence proof.

Lemma 9. *There exists $\epsilon > 0$ and a constant $K > 0$ such that for every interval $\mathcal{I} = [p_1, p_2]$ with $p_2 \leq \epsilon$ and implicit functions $x^+(p, \lambda)$ and $\lambda^+(p, \lambda)$ defined on the neighborhood $\mathcal{N}_{p_1,p_2}$ of $[p_1, p_2] \times \{\lambda^*\}$, the following estimates are satisfied:*

$$\|x^+(p, \lambda) - x^*\| \leq Kp\|\lambda - \lambda^*\|, \qquad \|\lambda^+(p, \lambda) - \lambda^*\| \leq Kp\|\lambda - \lambda^*\|. \tag{23}$$

Proof. For $0 < p_1 < p_2 \leq \epsilon$ let us consider the following system of $n + |I_=|$ nonlinear equations for $n + |I| + 1 + |I|$ variables $x^+, \lambda^+, p, \lambda$:

$$\begin{cases} f'(x^+) + \sum_{i \in I} \lambda_i^+ g_i'(x^+) = 0 \\ \\ p\, \lambda_i \phi'\left(g_i(x^+)/p\right) - p\, \lambda_i^+ = 0, \quad i \in I_= \end{cases} \tag{24}$$

Let us write $\Psi(x^+, \lambda^+, p, \lambda) = 0$ in abridged form. By construction, setting $(x^+, \lambda^+) := (x^+(p, \lambda), \lambda^+(p, \lambda))$ gives a solution of (24) for every $(p, \lambda) \in \mathcal{N}_{p_1,p_2}$. That is,

$$\Psi\left(x^+(p, \lambda), \lambda^+(p, \lambda), p, \lambda\right) = 0.$$

Let us differentiate (24), respectively $\Psi = 0$, with respect to $\lambda \in \mathbb{R}^I$. We obtain

$$\begin{bmatrix} L_{xx}(x^+, \lambda^+) & [g'(x^+)]^= \\ \left[\Lambda \phi''\left(g(x^+)/p\right) g'(x^+)^T\right]_= & -p\, I \end{bmatrix} \begin{bmatrix} (x^+)_\lambda \\ (\lambda_=^+)_\lambda \end{bmatrix} = \begin{bmatrix} -[g'(x^+)]^< \left(\lambda_<^+\right)_\lambda \\ -p\left[\phi'\left(g(x^+)/p\right)\right]_= \end{bmatrix} \tag{25}$$

where $\lambda_=^+ = [\lambda^+(p, \lambda)]_= = \lambda_=^+(p, \lambda)$, $\lambda_<^+ = [\lambda^+(p, \lambda)]_< = \lambda_<^+(p, \lambda)$, and similarly for $(\lambda_=^+)_\lambda$. The matrix on the left hand side of (25) is just $H(p, \lambda)$ introduced in (22). We claim that the right hand side in (25) is of the form $\begin{bmatrix} -p\,\Xi \\ -p\,\Theta \end{bmatrix}$ for bounded terms Ξ and Θ. Indeed, we have

$$\Theta = [\phi'\left(g\left(x^+\right)/p\right)]_=,$$

which is bounded on each $\mathcal{N}_{p_1,p_2}$, because $(x^+(p,\lambda),p,\lambda)$ belongs to the wedge neighborhood $\mathcal{W}(\epsilon)$ for every $(p,\lambda)_i n\mathcal{N}_{p_1,p_2}$. For the first coordinate on the right hand side of (25), equation (21) gives

$$\begin{aligned} -[g'(x^+)]^< \, (\lambda^+_<)_\lambda \; &= \; -[g'(x^+)]^< \, [\phi'\left(g(x^+)/p\right)]_< \\ &\quad - \; p^{-1}[g'(x^+)]^< \, [\Lambda\phi''\left(g(x^+)/p\right)]_< \, g'(x^+)^T \, (x^+)_\lambda, \end{aligned}$$

which means

$$\begin{aligned} \Xi \; &= \; p^{-1}[g'(x^+)]^< \, [\phi'\left(g(x^+)/p\right)]_< \\ &\quad + \; p^{-2}[g'(x^+)]^< \, [\Lambda\,\phi''\left(g(x^+)/p\right)]_< \, g'(x^+)^T \, (x^+)_\lambda. \end{aligned}$$

Here the first term is bounded by axiom (ϕ_4). As for the second term, notice that $(x^+)_\lambda$ is bounded by Lemma 7. With axiom (ϕ_5) we deduce that the second term in Ξ is bounded.

Now we have the identities

$$\begin{aligned}
\begin{bmatrix} x^+(p,\lambda) - x^+(p,\lambda^*) \\ \lambda_=^{\pm}(p,\lambda) - \lambda_=^{\pm}(p,\lambda^*) \end{bmatrix} &= \int_0^1 \frac{d}{d\eta} \begin{bmatrix} x^+\left(p,\lambda^* + \eta(\lambda-\lambda^*)\right) \\ \lambda_=^{\pm}\left(p,\lambda^* + \eta(\lambda-\lambda^*)\right) \end{bmatrix} d\eta \\
&= \int_0^1 \begin{bmatrix} (x^+)_\lambda(p,\lambda_\eta)\cdot(\lambda-\lambda^*) \\ (\lambda_=^{\pm})_\lambda(p,\lambda_\eta)\cdot(\lambda-\lambda^*) \end{bmatrix} d\eta \\
&= \int_0^1 \begin{bmatrix} (x^+)_\lambda(p,\lambda_\eta) \\ (\lambda_=^{\pm})_\lambda(p,\lambda_\eta) \end{bmatrix} (\lambda-\lambda^*)\, d\eta \\
&= \int_0^1 H(p,\lambda_\eta)^{-1} \begin{bmatrix} -p\,\Xi_\eta \\ -p\,\Theta_\eta \end{bmatrix} (\lambda-\lambda^*)\, d\eta \qquad (26)
\end{aligned}$$

where $\lambda_\eta := \lambda^* + \eta(\lambda-\lambda^*)$, and where Ξ_η and Θ_η refer to the elements Ξ, Θ at the intermediate points $(x^+)_\eta = x^+(p,\lambda_\eta)$.

From Lemma 8 we know that $\|H(p,\lambda)^{-1}\| \leq K_5$ on any set $\mathcal{N}_{p_1,p_2}$, where K_5 depends only on ϵ, as long as $0 < p_1 < p_2 \leq \epsilon$.

From what we have seen above, Ξ_η and Θ_η are bounded on a wedge neighborhood. Therefore, the right hand term under the integral above is bounded by $K_6\, p\, \|\lambda - \lambda^*\|$ for some $K_6 > 0$. Then, regarding that $x^+(p,\lambda^*) = x^*$ and $\lambda^+(p,\lambda^*) = \lambda^*$ for $p \in I_{p_1,p_2} = [p_1,p_2]$, we have

$$\left(\|x^+(p,\lambda) - x^*\|^2 + \|[\lambda_=^{\pm}(p,\lambda) - \lambda_=^*\|^2\right)^{1/2} \;\leq\; K_7 p \|\lambda - \lambda^*\| \qquad (27)$$

for $K_7 = K_5 K_6$ depending on ϵ, as long as $0 < p_1 < p_2 \leq \epsilon$ and $p \in I_{p_1,p_2}$. Here we use Lemma 8.

In order to prove (23), it remains to estimate $\|\lambda_<^{\pm}(p,\lambda) - \lambda_<^*\|$ against $\|\lambda_< - \lambda_<^*\|$. But notice that by complementarity $\lambda_<^* = 0$, hence we have to compare

$\|\lambda_<^+\|$ to $\|\lambda_<\|$, which leads us back to the update formula. Its $I_<$-part reads:

$$\lambda_<^+(p,\lambda) = [\phi'\left(g(x^+(p,\lambda)/p\right)\ \lambda]_<.$$

Let $i \in I_<$, then $g_i\left(x^+(p,\lambda)\right)/p \to -\infty$. Therefore, by axiom (ϕ_4), we have $\phi'\left(g_i\left(x^+(p,\lambda)\right)/p\right) \le K_8 p$ for some constant $K_8 > 0$. This proves $\lambda_i^+(p,\lambda) \le K_8\, p\lambda_i$ for $i \in I_<$. Combining with (27) proves (23) with $K = \max\{K_7, K_8\}$. □

Remark. Notice that the constant $\theta := K\epsilon$ in (23) may be chosen arbitrarily small. In particular, we will assure that $\theta = K\epsilon < \frac{1}{2}$, so that $Kp < \frac{1}{2}$ for each of the p involved in the estimates (23). We then have to choose the neighborhoods $\mathcal{U}$, $\mathcal{N}$, which requires fixing $0 < p_1 < p_2 \le \epsilon$. Here we will first fix p_2, the only request being $p_2 \le \epsilon$. Then, when fixing p_1, we will make sure that $p_1 < p_2/c$, where $c > 1$ is the constant used in step 3 of the augmented Lagrangian algorithm. Namely, the fact that $p_{k+1} \in \{p_k, p_k/c\}$ will then guarantee that as the p_k get smaller, some p_k will fall within the range $[p_1, p_2]$, where the conclusions of Lemma 9 are valid. □

3.2. Complexity for F_1

In order to prove convergence of the augmented Lagrangian method based on F_1, we need one last element. We know that the penalty parameter p may be frozen when λ is sufficiently close to λ^*. However, what is needed is a mechanism in the algorithm which does this automatically. This last element is assured by the progress measure $\sigma_1(x,p,\lambda)$ from (10), whose role is to fix p as soon as iterates make sufficient progress towards feasibility.

Lemma 10. *Under hypotheses* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3), *there exists* $\epsilon > 0$, $\overline{p} < \epsilon$ *and neighborhoods* N *of* λ^*, $\mathcal{U}$ *of* x^* *such that the sequences* x_k, p_k *and* λ^k *generated by the augmented Lagrangian algorithm for (6) with* $\overline{p} \leq p_1 \leq \epsilon$ *and* $\lambda_1 \in N$ *have the following properties: The sequence* x_k *stays in* $\mathcal{U}$, *the sequence* λ^k *stays in* N, *and the sequence* p_k *stays bounded away from 0. Moreover,* x_{k+1} *is the unique local minimum (even the unique critical point) of program* (P_{p_k,λ^k}) *in* $\mathcal{U}$.

Proof. 1) Let K, ϵ be the constants in (23). Choose $\overline{p} \leq \epsilon$ such that $\theta := K\overline{p} < \frac{1}{2}$. Then $\frac{\theta}{1-\theta} \in (0,1)$. Without loss we may even arrange that $\frac{\theta}{1-\theta} = \tau$, where $\tau \in (0,1)$ is the parameter used in the algorithm.

Choose a neighborhood N_0 of $\lambda^* \neq 0$ such that $0 < m \leq \|\lambda\| \leq M < +\infty$ for $\lambda \in N_0$. Then choose neighborhoods $\mathcal{U}_0$ of $x^* \in \mathbb{R}^n$ and W_0 of 0 in $\mathbb{R}$ such that g has Lipschitz constant $\mathrm{Lip}(g,x^*) < +\infty$ on $\mathcal{U}_0$, ϕ' has Lipschitz constant $\mathrm{Lip}(\phi',0) < +\infty$ on W_0. Put $K' = K\,\mathrm{Lip}(g,x^*)\,\mathrm{Lip}(\phi',0)$. We may assume without loss that $K'M > 1$.

Let us now next select $0 < \underline{p} < \overline{p}$ such that $\underline{p} < \overline{p}/c^\beta$, where

$$\beta = \left\lceil \frac{\log\left(\theta + (1-\theta)MK'\right)}{\log c} \right\rceil + 2.$$

We now consider the interval $\mathcal{I} = [\underline{p}, \overline{p}]$ fixed, and also the implicit functions $x^+ = x^+(p,\lambda)$ and $\lambda^+ = \lambda^+(p,\lambda)$ associated with it.

Let the neighborhoods $\mathcal{U} = \mathcal{U}_{\underline{p},\overline{p}}$ of x^* and $N = N_{\underline{p},\overline{p}}$ of λ^* be fixed as well. Assume without loss of generality that $N \subset N_0, \mathcal{U} \subset \mathcal{U}_0$ and $g(\mathcal{U}) \subset W_0$, so that the Lipschitz constant of g on $\mathcal{U}$ is $\leq \mathrm{Lip}(g, x^*)$, the Lipschitz constant of ϕ' on $g(\mathcal{U})$ is $\leq \mathrm{Lip}(\phi', 0)$.

We will ultimately show that the sequence p_k stays in the interval $[\underline{p}, \overline{p}]$. We may assume that there exists a first index k_0 such that $p_{k_0} \leq \overline{p}$, otherwise there is nothing to prove. Notice that $p_{k_0} > \overline{p}/c$. Also, by the choice of $\overline{p}$ we have $p_{k-2} < K^{-1}$ for $k \geq k_0 + 2$, where K is the constant in (23).

2) We will now obtain an estimate $\sigma_1(x_{k+1}, p_k, \lambda^k) \leq \tau_k \sigma_1(x_k, p_{k-1}, \lambda^{k-1})$ for certain constants τ_k approaching $\tau = \theta/(1-\theta)$.

Let us first consider indicies $i \in I_=$. Then we have

$$\begin{aligned}
\left|\lambda_i^k \left(1 - \phi'\left(g_i(x_{k+1})/p_k\right)\right)\right| &\leq |\lambda_i^k| \,\mathrm{Lip}(\phi', 0)\, \left|\left(g_i(x_{k+1}) - g_i(x^*)\right)/p_k\right| \\
&\leq |\lambda_i^k| \,\mathrm{Lip}(\phi', 0)\, \mathrm{Lip}(g, x^*)\, \|x_{k+1} - x^*\|/p_k \\
&\leq |\lambda_i^k| \,\mathrm{Lip}(\phi', 0)\, \mathrm{Lip}(g, x^*)\, K p_k \|\lambda^k - \lambda^*\|/p_k \\
&= |\lambda_i^k| K' \|\lambda^k - \lambda^*\|
\end{aligned}$$

where the first estimate uses $g_i(x^*) = 0$, $g_i(x_k)/p_k \to 0$ and $\phi'(0) = 1$, while the last estimate is (23).

Using the definition of the first-order update rule (8) at stage $k-1$, we have:

$$\begin{aligned}
\|\lambda^k - \lambda^*\| &\leq K p_{k-1} \|\lambda^{k-1} - \lambda^*\| \\
&\leq K p_{k-1} \left(\|\lambda^{k-1} - \lambda^k\| + \|\lambda^k - \lambda^*\|\right)
\end{aligned}$$

and therefore

$$\|\lambda^k - \lambda^*\| \leq \left(K^{-1} p_{k-1}^{-1} - 1\right)^{-1} \|\lambda^{k-1} - \lambda^k\|,$$

where $K^{-1} p_{k-1}^{-1} > 1$ by assumption.

Now consider $i \in I_<$. Recall that $\theta = K\overline{p}$, and that $\lambda_i^* = 0$ by complementarity. Hence by (23) in Lemma 9 we have $\lambda_i^{k+1} = \theta_i^k \lambda_i^k$ and $\lambda_i^k = \theta_i^{k-1}\lambda_i^{k-1}$ for certain $0 \leq \theta_i^k, \theta_i^{k-1} \leq \theta$, $k \geq k_0 + 1$. Then

$$\lambda_i^k - \lambda_i^{k+1} = (1-\theta_i^k)\,\lambda_i^k = (1-\theta_i^k)\,\theta_i^{k-1}\lambda_i^{k-1}$$

and

$$\lambda_i^{k-1} - \lambda_i^k = (1-\theta_i^{k-1})\,\lambda_i^{k-1}.$$

This gives

$$\lambda_i^k - \lambda_i^{k+1} = \frac{(1-\theta_i^k)\theta_i^{k-1}}{1-\theta_i^{k-1}}(\lambda_i^{k-1} - \lambda_i^k) \leq \frac{\theta}{1-\theta}\left(\lambda_i^{k-1} - \lambda_i^k\right)$$

for every $i \in I_<$.

Suppose now without loss of generality that the norm used to define σ_1 is the maximum norm. Then piecing together the two ends in $I_=$ and the estimate in $I_<$ just found, we obtain

$$\sigma_1\left(x_{k+1}, p_k, \lambda^k\right) \leq \tau_k \sigma_1\left(x_k, p_{k-1}, \lambda^{k-1}\right),$$

where

$$\tau_k := \max\left\{\frac{\theta}{1-\theta}, \frac{\|\lambda^k\|K'}{K^{-1}p_{k-1}^{-1} - 1}\right\}.$$

This proves what was claimed in part 2).

3) Observe that the second term in the definition of τ_k tends to zero as soon as $p_k \to 0$. This is because $\|\lambda^k\| \leq M$ as long as the multiplier estimates stay in the neighborhood N of λ^*, which is the case as long as estimate (23) in Lemma 9 applies. Therefore, if p_j is reduced sufficiently often, we will eventually have

$\tau_k = \theta/(1-\theta) = \tau < 1$. Could the term on the right hand side in τ_k also get smaller because $\|\lambda^k\| \to 0$? The answer is no, because $\|\lambda^k\| \geq m > 0$ as long as $\lambda^k \in N$. Therefore, if eventually $\tau = \tau_k$, this is due to the fact that p_k becomes too small.

Let us estimate the index k where $\tau_k = \tau$ for the first time. This index satisfies

$$\frac{\theta}{1-\theta} < \frac{\|\lambda^{k-1}\|K'}{K^{-1}p_{k-2}^{-1} - 1}, \qquad \frac{\theta}{1-\theta} \geq \frac{\|\lambda^{k}\|K'}{K^{-1}p_{k-1}^{-1} - 1}.$$

The left hand estimate implies

$$p_{k-2}^{-1} < K\left(1 + \frac{(1-\theta)K'\|\lambda^{k-1}\|}{\theta}\right).$$

Now observe that

$$p_{k-2} = \frac{\overline{p}}{c^\alpha},$$

where $\alpha < k - k_0 - 1$. Indeed, at each step, either $p_{j+1} = p_j$, or $p_{j+1} = p_j/c$ and $c^\alpha = \overline{p}/p_{k-2} < cp_{k_0}/p_{k-2} \leq c^{k-k_0-1}$. This implies

$$\alpha \leq \frac{\log\left(K\overline{p}\left(1 + \frac{(1-\theta)\|\lambda^{k-1}\|K'}{\theta}\right)\right)}{\log c} = \frac{\log\left(\theta + (1-\theta)\|\lambda^{k-1}\|K'\right)}{\log c} \leq \beta - 2$$

by the definition of β and $\underline{p}$. Here we use again that $\lambda^{k-1} \in N$, hence $\|\lambda^{k-1}\| \leq M$. That means $p_k > \underline{p}$. But, according to its definition, k is the moment where the algorithm definitely stops reducing p (if it did not do so before). This is because $\tau_k = \tau$ means the test in step 3 of the algorithm accepts p, and keeps doing so because the τ_k decrease monotonically. That means, the iterates $p_{k+1}, p_{k+2}, \ldots$ do *not* decrease any further. Since $\underline{p}$ is not yet reached at stage k, it is never reached. This proves that the p_k are trapped in $[\underline{p}, \overline{p}]$.

Since the iterates now satisfy $\lambda^k \in N$ and $p_k \in [\underline{p}, \overline{p}]$ at each k, the hypotheses of Lemma 9 are all met, and the conclusions are therefore valid for every $k \geq k_0$. This completes the proof. $\square$

As a consequence we have the following:

Theorem 1. *Let x^* be a local minimum of (1) with inequality constraints only. Let λ^* be an associated Lagrange multiplier such that (x^*, λ^*) is a KKT-pair where hypotheses (H_1) - (H_3) are satisfied. Suppose further that the penalty function ϕ satisfies axioms (ϕ_1) - (ϕ_5), and that the augmented Lagrangian function (6) with progress measure (10) is used. Then there exists a neighborhood N of λ^*, a neighborhood $\mathcal{U}$ of x^* and $\epsilon > 0$ such that for every $\lambda^1 \in N$, and every p_1 with $\epsilon/c < p_1 \leq \epsilon$, the sequences x_k, λ^k and p_k generated by the augmented Lagrangian algorithm have the following properties:*

1. *For every k, program $\min\limits_{x \in \mathbb{R}^n} F_1\left(x, p_k, \lambda^k\right)$ has a unique strict local minimum $x_{k+1} \in \mathcal{U}$.*
2. *The sequence x_k converges R-linearly to x^*.*
3. *The sequence λ^k with $\lambda^{k+1} = \phi'\left(g(x_{k+1})/p_k\right) \lambda^k$ stays in N and converges Q-linearly to λ^*.*
4. *The sequence p_k stays bounded away from 0.*

Proof. As a consequence of the initial condition, Lemma 10 tells us that the sequence p_k stays in the interval $\mathcal{I} = [\underline{p}, \overline{p}]$. Since the neighborhoods $\mathcal{U}$ and N are chosen with respect to $\mathcal{I}$, we have $x_{k+1} = x^+(p, \lambda^k)$ and $\lambda^{k+1} = \lambda^+(p, \lambda^k)$ from some index k_0 onwards, where $p = p_k \in \mathcal{I}$ for $k \geq k_0$. As a consequence of estimate (23) with $K\overline{p} < \frac{1}{2}$, we immediately deduce Q-linear convergence of λ^k, while the left hand part of (23) gives R-linear convergence of x_k. That completes the proof. □

4. Convergence for F_2

In this section we prove the analogue of Theorem 1 for the augmented Lagrangian (7) with progress measure (11). Again, we only consider the case of inequality constraints.

4.1. First steps

What we plan to do is follow the proof of Theorem 1 and indicate the necessary changes in Lemmas 1 - 9 and Lemma 10 as we go. We make the same assumptions (H_1) - (H_3) and (ϕ_1) - (ϕ_5).

Lemma 11. *Under axioms $(\phi_1) - (\phi_5)$ and (H_1) - (H_3), the conclusions of Lemma 9 remain valid when model (7) with associated progress measure (11) is used. In addition to what is claimed in Lemma 9, however, the sequence $\lambda^k_<$ of multipliers associated with inactive inequality constraints now converges even Q-superlinearly to $\lambda^*_< = 0$.*

Proof. We follow the steps in Lemmas 1 - 9, and indicate what changes have to be made.

1) In Lemma 1 it is shown that $F_{xx} \succeq rI \succ 0$ on a wedge neighborhood of $(x^*, 0, \lambda^*)$. We now need to replace (13) by the equation

$$\begin{aligned} d_k^T F_{xx}\left(x_k, p_k, \lambda^k\right) d_k &= d_k^T L_{xx}\left(x_k, \phi'\left(g(x_k)/(p_k\lambda^k)\right)\lambda^k\right) d_k + \\ &+ p_k^{-1}\left(g'(x_k)^T d_k\right)^T \Lambda^k \phi''\left(g(x_k)/(p_k\lambda^k)\right)\left(g'(x_k)^T d_k\right). \end{aligned}$$

Strict complementarity implies $g_i(x_k)/(p_k\lambda_i^k) \to 0$ for $i \in I_=$, and $g_i(x_k)/(p_k\lambda_i^k) \to -\infty$ for $i \in I_<$. Therefore the same conclusions $\phi'\left(g(x_k)/(p_k\lambda^k)\right)\lambda^k \to \phi'(0)\lambda^* = \lambda^*$ and $\Lambda^k\phi''\left(g(x_k)/(p_k\lambda^k)\right) \to \Lambda^*\phi''(0)$ are obtained, using axioms (ϕ_4), (ϕ_5). The argument remains essentially the same and proves $F_{xx} \succeq rI \succ 0$.

Similarly, the prior estimate in Lemma 2 carries over to the new objective F_2.

2) In Lemma 3 the matrix $H(x^*, \lambda^*, 0, \lambda^*)$ is introduced. Its definition is now as follows:

$$H(x^*, \lambda^*, 0, \lambda^*) = \begin{bmatrix} L_{xx}(x^*, \lambda^*) & [g'(x^*)]^= \\ \left[\phi''(0) g'(x^*)^T\right]_= & 0 \end{bmatrix}.$$

Showing that $H(x^*, \lambda^*_=, 0, \lambda^*)$ is invertible follows the same line as in the proof of Lemma 3, and we omit the details.

3) Following the lead of Lemma 4, we now define $H(x, \lambda^+, p, \lambda)$ as

$$H(x, \lambda^+, p, \lambda) = \begin{bmatrix} L_{xx}(x, \lambda^+) & [g'(x)]^= \\ \left[\phi''\left(g(x)/(p\lambda)\right) g'(x)^T\right]_= & -pI \end{bmatrix},$$

where $\phi''\left(g(x)/(p\lambda)\right)$ stands for the diagonal matrix whose diagonal entries are $\phi''\left(g_i(x)/(p\lambda_i)\right)$. As in Lemma 4, we plan to prove

$$H(x, \lambda^+, p, \lambda) \xrightarrow{w} H(x^*, \lambda^*_=, 0, \lambda^*).$$

In order to do this, notice first that $\phi''\left(g_i(x)/(p\lambda_i)\right) \to \phi''(0)$ for $i \in I_=$. Indeed, for $i \in I_=$, strict complementarity gives $g_i(x)/(p\lambda_i) \to 0$, because $\lambda_i \to \lambda_i^* > 0$ and $g_i(x)/p = \left(g_i(x) - g_i(x^*)\right)/p \to 0$. On the other hand, for $i \in I_<$ we have $g_i(x)/(p\lambda_i) \to -\infty$. By axiom (ϕ_5) we deduce $\phi''\left(g_i(x)/(p\lambda_i)\right) \to 0$. Altogether we have shown that wedge convergence $(x,p,\lambda) \xrightarrow{w} (x^*,0,\lambda^*)$ together with $\lambda^+ \to \lambda^*$ implies $H(x,\lambda^+,p,\lambda) \to H(x^*,\lambda^*,0,\lambda^*)$. Therefore there exists a wedge neighborhood $\mathcal{W}(\epsilon_2)$ of $(x^*,0,\lambda^*)$ and a neighborhood U_0 of λ^* such that $H(x,\lambda^+,p,\lambda)$ is invertible and its inverse is bounded for all $(x,p,\lambda) \in \mathcal{W}(\epsilon_2)$ and $\lambda^+ \in U_0$.

4) Right after Lemma 4, the implicit function theorem is applied to a full family of reference intervals $\mathcal{I} = [p_1, p_2]$, where p_2 is considered fixed but sufficiently small, while the lower end p_1 is kept mobile. The procedure is now the same. In particular, the unicity result Lemma 5 is obtained in precisely the same fashion.

What changes is the expression for the derivative $x_\lambda^+ = -F_{xx}^{-1}F_{x\lambda}$. We have

$$F_{x\lambda}\left(x,p,\lambda\right) = g'(x)\phi'\left(g(x)/(p\lambda)\right) - p^{-1}g'(x)\Lambda^{-1}\phi''\left(g(x)/(p\lambda)\right)G(x), \quad (28)$$

where $\Lambda^{-1}\phi''\left(g(x)/(p\lambda)\right)G(x)$ denotes the diagonal matrix with diagonal entries $\lambda_i^{-1}\phi''\left(g_i(x)/(p\lambda_i)\right)g_i(x)$. This notation will be kept in the sequel. In order to prove the analogue of Lemma 6, we have to show that $F_{x\lambda}$ is bounded on a wedge neighborhood of $(x^*,0,\lambda^*)$.

Let us start with the first term on the right hand side of (28). Assume $(x,p,\lambda) \xrightarrow{w} (x^*,0,\lambda^*)$. For $i \in I_=$ we have $g_i(x)/(p\lambda_i) \to 0$ by strict complementarity, so $\phi'\left(g_i(x)/(p\lambda_i)\right) \to \phi'(0) = 1$. On the other hand, for $i \in I_<$ we have $g_i(x)/(p\lambda_i) \to -\infty$. By (ϕ_4), the first term of $F_{x\lambda}$ is indeed bounded on a wedge neighborhood of $(x^*, 0, \lambda^*)$. Let us now consider the second term on the right hand side of (28). For $i \in I_=$, $(g_i(x)/p)\,\phi''\left(g_i(x)/(p\lambda_i)\right) \to 0$ by wedge convergence and strict complementarity. On the other hand, for $i \in I_<$, $(g_i(x)/(p\lambda_i))\,\phi''\left(g_i(x)/(p\lambda_i)\right)$ converges to zero because $t\phi''(t) = \mathcal{O}(1)$ as $t \to -\infty$ by axiom (ϕ_5). This completes the proof of the boundedness of $F_{x\lambda}$, and therefore proves the analogue of Lemma 7.

5) Let us now examine what changes need to be made in the proof of Lemma 8. Here we define $\lambda^+(p,\lambda)$ differently, namely as

$$\lambda_i^+(p,\lambda) = \lambda_i \phi'\left(g_i(x^+(p,\lambda))/(p\lambda_i)\right).$$

Then we introduce $H(p,\lambda)$ in the same way as in (22). By the new version of Lemma 7, obtained in step 4) above, the conclusion is that $H(p,\lambda)^{-1}$ is bounded on $\mathcal{N}_{p_1,p_2}$ of $p_2 \leq \epsilon$. This proves the analogue of Lemma 8.

6) In the proof of Lemma 9, a system of nonlinear equations is considered. It is now the following

$$\begin{cases} f'(x^+) + \displaystyle\sum_{i\in I} \lambda_i^+ g_i'(x^+) = 0 \\ p\,\lambda_i\,\phi'\left(g_i(x^+)/(p\lambda_i)\right) - p\lambda_i^+ = 0, \quad i \in I_= \end{cases}$$

The vector $(x^+(p,\lambda), \lambda^+(p,\lambda), p, \lambda)$ is solution for $(p,\lambda) \in \mathcal{N}$, just as in the proof of Lemma 9. Differentiation with respect to λ leads to

$$\begin{bmatrix} L_{xx}(x^+,\lambda^+) & [g'(x^+)]^= \\ \left[\phi''\left(g(x^+)/(p\lambda)\right) g'(x^+)^T\right]_= & -pI \end{bmatrix} \begin{bmatrix} (x^+)_\lambda \\ (\lambda_=^+)_\lambda \end{bmatrix} = \begin{bmatrix} -p\,\Xi \\ -p\,\Theta \end{bmatrix},$$

where $\Xi \;=\; p^{-1}[g'(x^+)]^< \;(\lambda^+_<)_\lambda$, or, explicitly:

$$\begin{aligned}\Xi \;=\; & p^{-1}[g'(x^+)]^< \;[\phi'\left(g(x^+)/(p\lambda)\right)]_< \\ & -\; p^{-2}[g'(x^+)]^< \;[\Lambda^{-1}\phi''\left(g(x^+)/(p\lambda)\right) G(x^+)]_< \\ & +\; p^{-2}[g'(x^+)]^< \;[\phi''\left(g(x^+)/(p\lambda)\right)]_< \, g'(x^+)^T \;(x^+)_\lambda, \quad (29)\end{aligned}$$

and $\quad \Theta \;=\; [\phi'\left(g(x^+)/(p\lambda)\right)]_= \;-\; p^{-1}[\Lambda^{-1}\phi''\left(g(x^+)/(p\lambda)\right) G(x^+)]_=$.

The first term on the right-hand side in (29) is bounded by (ϕ_4), whereas the two remaining terms are bounded by (ϕ_5). On the other hand, boundedness of Θ is clear since $g_i(x^+)/(p\lambda_i) \to 0$ for $i \in I_=$.

Now we get litterally the same estimate (26) as in the proof of Lemma 9, where Ξ_η, Θ_η refer to the expressions Ξ, Θ at the intermediate points $(x^+)_\eta$, but of course have to be actualized as in 6) above. As in Lemma 9 it follows that Θ_η, Ξ_η are bounded. The conclusion are then the same as in the proof of Lemma 9, i.e., we get the estimate (27).

It remains to prove the second part of estimate (23) for the $i \in I_<$. This is done in much the same way as in the proof of Lemma 9, where we consider the $I_<$ part of the update formula:

$$\lambda_i^+(p,\lambda) = \lambda_i \phi'\left(g_i(x^+(p,\lambda))/(p\lambda_i)\right).$$

Recall that $g_i\left(x^+(p,\lambda)\right) < 0$ for $x^+(p,\lambda)$ sufficiently close to x^*, which may be arranged to hold as soon as $(p,\lambda) \in \mathcal{N}_{p_1,p_2}$. In other words, $\phi'\left(g_i\left(x^+(p,\lambda)\right)/(p\lambda_i)\right) \to 0$. This proves even a superlinear rate of convergence $\lambda_i^+(p,\lambda) = \mathrm{o}(\lambda_i)$ for $i \in I_<$, which completes the proof of Lemma (11). $\square$

4.2. Complexity for F2

Let us next examine the proof of Lemma 10 for the new augmented Lagrangian (7) with associated progress measure (11).

Lemma 12. *Under the assumptions* (ϕ_1) - (ϕ_5) *and* (H_1) - (H_3)*, the conclusions of Lemma 10 remain valid if the augmented Lagrangian method with objective (7) and progress measure (11) are used.*

Proof. We follow the arguments in Lemma 10. Let $\tau = \theta/(1-\theta)$ with $K\overline{p} = \theta < \frac{1}{2}$ and $\overline{p} \leq \epsilon$. Now define

$$\beta = \left\lceil \frac{\log\left(\theta + (1-\theta)K'\right)}{\log c} \right\rceil + 2,$$

where $K' = K\mathrm{Lip}(g, x^*)\mathrm{Lip}(\phi', 0)$. Let $\underline{p} < \overline{p}/c^\beta$. We will show that if k_0 is the smallest index with $p_{k_0} \in [\underline{p}, \overline{p}]$, then the p_k $(k \geq k_0)$ never leave this interval. As in Lemma 10 we establish this by showing that $\sigma_2(x_{k+1}, p_k, \lambda^k) \leq \tau_k \sigma_2(x_k, p_{k-1}, \lambda^{k-1})$ with $\tau_k \to \tau$.

Let us first consider indicies $i \in I_=$. Then we have

$$\begin{aligned}
\left|\lambda_i^k \left(1 - \phi'\left(g_i(x_{k+1})/(p_k\lambda_i^k)\right)\right)\right| &\leq |\lambda_i^k|\,\mathrm{Lip}(\phi', 0)\,|\left(g_i(x_{k+1}) - g_i(x^*)\right)/(p_k\lambda_i^k)| \\
&\leq |\lambda_i^k|\,\mathrm{Lip}(\phi', 0)\,\mathrm{Lip}(g_i, x^*)\,\|x_{k+1} - x^*\|/(p_k\lambda_i^k) \\
&\leq \mathrm{Lip}(\phi', 0)\,\mathrm{Lip}(g_i, x^*)\,Kp_k\|\lambda^k - \lambda^*\|/p_k \\
&\leq K'\|\lambda^k - \lambda^*\|
\end{aligned}$$

where $K' := \max_{i \in I_=}\{\mathrm{Lip}(\phi', 0)\,\mathrm{Lip}(g_i, x^*)\,K\}$.

As in the proof of Lemma 10 we have the estimate

$$\|\lambda^k - \lambda^*\| \leq \left(K^{-1}p_{k-1}^{-1} - 1\right)^{-1}\|\lambda^{k-1} - \lambda^k\|.$$

For constraints $i \in I_<$ the argument is literally the same as in the proof of Lemma 10. Altogether we get again that

$$\sigma_2\left(x_{k+1}, p_k, \lambda^k\right) \leq \tau_k \sigma_2\left(x_k, p_{k-1}, \lambda^{k-1}\right)$$

now for

$$\tau_k = \max\left\{\frac{\theta}{1-\theta}, \frac{K'}{K^{-1}p_{k-1}^{-1} - 1}\right\}.$$

As in Lemma 10 we estimate the index $k \geq k_0$ where for the first time $\tau_k = \tau$. Letting $p_k = \overline{p}/c^\alpha$, we have $\alpha < k - k_0 - 1$, and we find that

$$\alpha \leq \frac{\log\left(\theta + (1-\theta)K'\right)}{\log c} \leq \beta - 2.$$

Hence $p_k > \underline{p}$. Since the test in step 3 of the algorithms leaves p_k unchanged as soon as $\tau_k = \tau$, we see that the p_k $(k \geq k_0)$ never leave $[\underline{p}, \overline{p}]$, as claimed. $\square$

Theorem 2. *Under the assumptions of Theorem 1, the sequences of iterates x_k, λ^k and p_k generated by the augmented Lagrangian algorithm based on model (7) and (11) have the same properties 1. - 3. as in that theorem. In addition, either the sequence x_k converges Q-linearly, or x_k (and then also λ^k) converges R-superlinearly.*

Proof. The proof is covered by the previous Lemmas and follows the overall scheme of the previous section.

The major difference with F_1 is that now x_k converges locally Q-linearly. To prove this we consider the situation of Lemmata 11, 12. We may assume that the sequence $p_k = p$ is already constant. Then $x_{k+1} = x^+(p, \lambda^k)$. Using $x^+(p, \lambda^*) = x^*$, we have the Taylor expansion

$$x^+(p, \lambda) - x^* = (x^+)_\lambda(p, \lambda^*)(\lambda - \lambda^*) + \mathrm{o}(\|\lambda - \lambda^*\|). \tag{30}$$

Now observe that $(x^+)_\lambda = -F_{xx}^{-1} F_{x\lambda}$ by the differentiation rule for the implicit function (Lemma 5 and its alter ego), where $F_{x\lambda}$ is given in (28). We now consider separately the terms of the right-hand side in (28), when approaching (x^*, p, λ^*).

Let us start with the second term in (28). Obviously entries associated with active constraints tend to 0. The same follows for entries related to inactive constraints, because of axiom (ϕ_5). Thus the second term in (28) ultimately vanishes. As for the first term in (28), we can assume without loss of generality that $g'(x) = [g'_=(x),\, g'_<(x)]$. Combining axioms (ϕ_3) and (ϕ_4), we have the following:

$$F_{x\lambda}(x^*, p, \lambda^*) = [[g'(x^*)]^=,\, 0],$$

which means that

$$F_{x\lambda}(x^*, p, \lambda^*)(\lambda - \lambda^*) = [g'(x^*)]^= [\lambda - \lambda^*]_=.$$

Altogether, substituting this back into formula (30) gives

$$F_{xx}(x^*, p, \lambda^*) \left(x^+(p, \lambda) - x^*\right) \;=\; -[g'(x^*)]^= \;(\lambda_= - \lambda^*_=) \;+\; \mathrm{o}(\|\lambda - \lambda^*\|).$$

Now we observe a difference between F_1 and F_2. With F_2 inactive multipliers converge Q-superlinearly to 0, because $\lambda_i^+/\lambda_i = \phi'\left(g_i(x^+)/(p\lambda_i)\right) = \mathcal{O}(\lambda_i)$ by axiom (ϕ_4), and $\lambda_i \to \lambda_i^* = 0$ for $i \in I_<$. On the other hand, for active multipliers $i \in I_=$ we have only estimate (23), which assures Q-linear convergence. There are now two possibilities. Case 1 is when $\lambda_< \to \lambda_<^* = 0$ faster than $\|\lambda_= - \lambda_=^*\| \to 0$. This is the situation we expect. Here we will establish Q-linear convergence $x \to x^*$. On the other hand, we cannot apriori exclude the possibility that $\|\lambda_= - \lambda_=^*\| \to 0$ with the same speed as $\lambda_< \to \lambda_<^* = 0$. This is case 2. Here, $\lambda \to \lambda^*$ R-superlinearly (and Q-linearly), so in that case estimate (23) shows that $x_k \to x^*$ converges even R-superlinearly.

Let us concentrate on the more likely case 1, where $\lambda_= \to \lambda_=^*$ with Q-linear speed, while $\lambda_< \to \lambda_<^* = 0$ Q-superlinearly. This means $\|\lambda - \lambda^*\| \le K' \|\lambda_= - \lambda_=^*\|$, for some $K' > 0$, as soon as λ is close to λ^*. As a consequence we obtain

$$x^+(p,\lambda) - x^* = F_{xx}(x^*,p,\lambda^*)^{-1}\,[g'(x^*)]^=\,(\lambda_= - \lambda_=^*) + \mathrm{o}(\|\lambda_= - \lambda_=^*\|). \tag{31}$$

If follows that, sufficiently close to λ^*,

$$\|x^+(p,\lambda) - x^*\| \le 2\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|\,\|\lambda_= - \lambda_=^*\|$$

By assumption $[g'(x^*)]^=$ is of maximal rank $|I_=|$. From (31) we then also have

$$\|\lambda_= - \lambda_=^*\| \le 2\,\|[g'(x^*)]^{=\dagger}\|\,\|F_{xx}\|\,\|x^+(p,\lambda) - x^*\|.$$

We apply these to λ and to $\lambda^+ = \lambda^+(p,\lambda)$ and then combine with (23). The result is

$$\begin{aligned}
\|x^+(p,\lambda^+) - x^*\| &\le 2\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|\,\|\lambda_=^+ - \lambda_=^*\| \\
&\le 2\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|\,\|\lambda^+ - \lambda^*\| \\
&\le 2\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|\,Kp\|\lambda - \lambda^*\| \\
&\le 2\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|\,K'Kp\|\lambda_= - \lambda_=^*\| \\
&\le 4\,\|F_{xx}^{-1}\|\,\|[g'(x^*)]^=\|K'Kp\|[g'(x^*)]^{=\dagger}\|\,\|F_{xx}\|\,\|x^+(p,\lambda) - x^*\|.
\end{aligned}$$

If p is chosen so small that $4\|F_{xx}^{-1}\|\|[g'(x^*)]^{=}\|KK'p\|[g'(x^*)]^{=\dagger}\|\|F_{xx}\| < 1$, Q-linear convergence of $x_{k+1} = x^+(p,\lambda^k)$ follows readily. □

Remark. Notice that the argument above applies as soon as $\|\lambda - \lambda^*\| = \mathcal{O}(\|\lambda_= - \lambda_=^*\|)$. In particular, with this extra assumption, the proof would also work for model F_1.

5. An example

The reader will have noticed a difference between (6) and (7). We can see that multipliers for inactive constraints $i \in I_<$ converge faster to 0 when (9) is used. Indeed, while (8) gives linear speed for active and inactive multipliers, (9) gives superlinear speed $\lambda_i \to 0$ for $i \in I_<$. This should be an advantage, and indeed, (7) seems to perform somewhat better in a neighborhood of attraction of a local minimum with strict complementarity. To demonstrate this we consider the following example

$$\begin{aligned} &\text{minimize} \quad f(x) = \tfrac{1}{2}(x_1^2 + x_2^2) \\ &\text{subject to } g_1(x) = x_1 - 2 \leq 0 \\ &\qquad\qquad\quad g_2(x) = 1 - x_2 \leq 0 \end{aligned}$$

The optimal solution is at $x^* = (0,1)$, constraint g_1 is inactive, constraint g_2 is active. The Lagrange multiplier is $\lambda^* = (0,1)$. We use the left hand ϕ in (5) and compare both augmented Lagrangian models (6) and (7).

To begin with consider F_1, which is

$$F_1(x,p,\lambda) = \frac{1}{2}(x_1^2 + x_2^2) + p\lambda_1 \left(-\frac{1}{4} \left(\log(4-2x_1) - \log p \right) - \frac{3}{8} \right) \\ + \lambda_2(1-x_2) + \frac{\lambda_2}{2p}(1-x_2)^2.$$

The optimality conditions $F_1' = 0$ imply

$$x_1 = 1 - \sqrt{1 + \frac{p\lambda_1}{4}} \qquad x_2 = \frac{p\lambda_2 + \lambda_2}{p + \lambda_2}$$

The update rules are

$$\lambda_2^+ = \lambda_2 \left(1 + \frac{1-\lambda_2}{p+\lambda_2} \right) \qquad \lambda_1^+ = \lambda_1 \frac{p}{8-4x_1} = \lambda_1 \frac{p}{4 + 4\sqrt{1 + \frac{p\lambda_1}{4}}}$$

Since λ_2 is supposed to converge to 1, we consider the quotient

$$\frac{\lambda_2^+ - 1}{\lambda_2 - 1} = \frac{p}{p + \lambda_2} \to \frac{p}{p+1} \in [0, 1[.$$

Here we have convergence with linear rate $p/(p+1)$.

For the inactive multiplier, which is supposed to converge to 0 we have

$$\frac{\lambda_1^+}{\lambda_1} = \frac{p}{4 + 4\sqrt{1 + \frac{p\lambda_1}{4}}} \to \frac{p}{8}$$

which is again a linear rate for $p < 8$.

Next consider F_2, which is

$$F_2(x,p,\lambda) = \frac{1}{2}(x_1^2 + x_2^2) + p\lambda_1^2 \left(-\frac{1}{4} \left(\log(4-2x_1) - \log(p\lambda_1) \right) \right) \\ + \lambda_2 \left(1-x_2\right) + \frac{1}{2p}(1-x_2)^2.$$

Taking derivatives $F_2' = 0$ gives

$$x_2 = \frac{p\lambda_2 + 1}{p+1} \qquad x_1 = 1 - \sqrt{1 + \frac{p\lambda_1^2}{4}}.$$

The first-order update rule is

$$\lambda_2^+ = \lambda_2 \left(1 + \frac{1 - x_2}{p\lambda_2}\right) \qquad \lambda_1^+ = \frac{\lambda_1^2 p}{4(2 - x_1)}$$

Now remember that λ_2 converges to 1, while λ_1 converges to 0. Then

$$\frac{\lambda_2^+ - 1}{\lambda_2 - 1} = \frac{p}{1+p},$$

which is a linear rate of convergence. Concerning the inactive constraint (g_1) :

$$\frac{\lambda_1^+}{\lambda_1} = \frac{p\lambda_1}{4(1 + \sqrt{1 + \frac{p\lambda_1^2}{4}})} \sim \frac{p\lambda_1}{8} \to 0.$$

So the inactive multiplier converges to 0 with quadratic speed. □

Appendix: Implicit function theorem

Below we state the following implicit function theorem from Hestenes [10, Theorem 7.2].

Lemma 13. *Let Ω be an open subset of $\mathbb{R}^n \times \mathbb{R}^m$ and let $\Psi : \Omega \to \mathbb{R}^n$ be of class $C^k(\Omega)$ for some $k \geq 1$. Let K^* be a compact subset of $\mathbb{R}^m$ and suppose there exists a vector $x^* \in \mathbb{R}^n$ with $\{x^*\} \times K^* \subset \Omega$ such that $\Psi(x^*, y) = 0$ for every $y \in K^*$. Suppose $\Psi_x(x^*, y)$ is invertible for every $y \in K^*$. Then there exists a neighborhood W of $\{x^*\} \times K^*$, a neighborhood V of K^*, and a function $x(\cdot) : V \to \mathbb{R}^n$ of class C^k such that $\Psi(x(y), y) = 0$ for every $y \in V$ and $x(y) = x^*$ for every $y \in K^*$. The function is unique in the sense that for every $(x, y) \in W$, $\Psi(x, y) = 0$ if and only if $y \in V$ and $x = x(y)$. Moreover,*

$$x'(y) = -\left[\Psi_x\left(x(y), y\right)\right]^{-1} \Psi_y\left(x(y), y\right).$$

□

This coincides with the usual implicit function theorem when the set $K^* = \{y^*\}$ is a singleton set.

References

1. A. Ben-Tal and M. Zibulevsky, *Penalty/barrier multiplier methods for convex programming problems*. SIAM J. Optimization, 1997, vol. 7, pp. 347 – 366.

2. D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, 1982.

3. D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1995.

4. E.G. Birgin and R.A. Castillo and J.M. Martínez, *Numerical comparison of augmented Lagrangian algorithms for nonconvex problems,* to appear in Computational Optimization and Applications, 2004.

5. P.T. Boggs and J.W. Tolle, *Augmented Lagrangians which are quadratic in the multiplier*, J. Opt. Th. and Appl., 1980, vol. 73, pp. 17 - 26.

6. M.G. Breitfeld and D.F. Shanno, *Computational experience with penalty-barrier methods for nonlinear programming*, in *Rutcor Research Report*, RRR 17-93, August 1993, Revised March 1994, Rutgers Center for Operations Research, Rutgers University, New Brunswick, New Jersey 08903, March 1994.

7. M.G. Breitfeld and D.F. Shanno, *A globally convergent penalty-barrier algorithm for nonlinear programming and its computational performance.*, in *Rutcor Research Report*, RRR 12-94, April 1994, Rutgers Center for Operations Research, Rutgers University, New Brunswick, New Jersey 08903, March 1994.

8. A.R. Conn and N.I.M. Gould and Ph. Toint, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds,* SIAM J. Numer. Anal., 1991, vol. 28, no. 2, pp. 545 – 572.

9. G. DiPillo and L. Grippo, *A new class of augmented Lagrangians in nonlinear programming*, SIAM J. on Control and Optim., 1979, vol. 17, pp. 618 - 628.

10. M.R. Hestenes, *Optimization Theory. The finite dimensional case.* Wiley & Sons, 1975.

11. M.R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 1969, vol. 4, pp. 303 – 320.

12. M. Kocvara and J. Outrata, *Effective Reformulations of the Truss Topology Design Problem*, in *IMA Preprint 1917*, University of Minnesota, Minneapolis, March 2003. Submitted to Optimization and Engineering.

13. M. Kocvara and M. Stingl, *A code for convex nonlinear and semidefinite programming*, Optimization Methods and Software, 2003, vol. 18, no. 3, pp. 317 – 333.

14. L. Mosheyev and M. Zibulevsky, *Penalty/barrier multiplier algorithm for semidefinite programming*, Optimization Methods and Software, 2000, vol. 13, no. 4, pp. 235 – 261.

15. M.J.D. Powell, *A method for nonlinear constraints in minimization problems*, in *Optimization*, R. Fletcher (ed.), Academic Press, New York, 1969, pp. 283 – 298.

16. R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming* SIAM J. Control, 1974, vol. 12, pp. 268 – 285.

17. M. Zibulevsky, *Penalty/barrier methods for large-scale nonlinear and semidefinite programming*, PhD Thesis, Technion, Israel Institute of Technology, 1996.