Pacific Journal of Optimization
(Vol. 2, No. 3, September 2006)

PRIMAL ERROR BOUNDS BASED ON THE AUGMENTED LAGRANGIAN AND LAGRANGIAN RELAXATION ALGORITHMS*

A. F. Izmailov† and M. V. Solodov‡

ABSTRACT

For a given iterate generated by the augmented Lagrangian or the Lagrangian relaxation based method, we derive estimates for the distance to the primal solution of the underlying optimization problem. The estimates are obtained using some recent contributions to the sensitivity theory, under appropriate first or second order sufficient optimality conditions. The given estimates hold in situations where known (algorithm-independent) error bounds may not apply. Examples are provided which show that the estimates are sharp.

Key words. Error bound, augmented Lagrangian, Lagrangian relaxation, sensitivity.
AMS subject classifications. 90C30, 90C33, 90C55, 65K05

* Research of the first author is supported by the Russian Foundation for Basic Research Grant 04-01-00341 and the RF President's Grant NS-1815.2003.1 for the support of leading scientific schools. The second author is supported in part by CNPq Grants 300734/95-6 and 471780/2003-0, by PRONEX–Optimization, and by FAPERJ.

† Moscow State University, Faculty of Computational Mathematics and Cybernetics, Department of Operations Research, Leninskiye Gori, GSP-2, 119992 Moscow, Russia.
Email: izmaf@ccas.ru

‡ Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: solodov@impa.br

1 Introduction

We consider the optimization problem

$$\begin{array}{ll}\text{minimize} & f(x)\\ \text{subject to} & F(x) = 0,\ G(x) \leq 0,\end{array} \tag{1.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function, $F : \mathbf{R}^n \to \mathbf{R}^l$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings. Specifically, for a local solution $\bar{x}$ of (1.1), we assume that f, F and G are twice differentiable at $\bar{x}$. The stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad F(x) = 0, \quad \mu \geq 0, \quad G(x) \leq 0, \quad \langle \mu,\, G(x)\rangle = 0, \tag{1.2}$$

where

$$L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}, \quad L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, F(x)\rangle + \langle \mu,\, G(x)\rangle,$$

is the standard Lagrangian function of problem (1.1).

Let $\mathcal{M}(\bar{x})$ be the set of Lagrange multipliers associated with $\bar{x}$. In this paper, we do not invoke any specific constraint qualifications, but the main line of our discussion presumes that the multiplier set $\mathcal{M}(\bar{x})$ is nonempty. This setting has been lately receiving much attention in the literature, e.g., [7, 1, 11, 6, 17], with one of the important motivations coming from optimization problems with complementarity constraints. We shall further assume a (first or second order) sufficient condition for optimality, so that $\bar{x}$ is a strict local solution of problem (1.1).

In this setting, a (primal) local error bound is the following property:

$$\|x - \bar{x}\| = O(\delta(x,\, \lambda,\, \mu)), \tag{1.3}$$

where $\delta : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$. The function δ should be (easily) computable and should provide a reasonable upper bound for the distance from x to $\bar{x}$. For example, it should tend to zero when $(x,\, \lambda,\, \mu)$ tends to $\{\bar{x}\} \times \mathcal{M}(\bar{x})$, or to a given point of this set. Moreover, it is desirable that the bound should be sharp, i.e., not improvable under the given assumptions. The point $(x,\, \lambda,\, \mu)$ in (1.3) can either be arbitrary or can be generated by some (primal-dual) algorithm for solving (1.1) (so that there is a certain relation between x and $(\lambda,\, \mu)$, in which case δ may also depend on parameters involved in the algorithm). In the former case, we shall call the error bound *algorithm-independent*, and in the latter case — *algorithm-based*. We refer the reader to [14] for a survey of error bounds and their applications. The setting of an arbitrary $(x,\, \lambda,\, \mu)$ is somewhat more traditional in the study of error bounds. Nevertheless, the case when the error bound holds along trajectories generated by some specific algorithm is also of interest and importance. The significance of an algorithm-based error bound in the context of the algorithm being considered, is essentially the same as that of an algorithm-independent bound. In particular, both provide quantitative information about convergence and, as a consequence, a reliable stopping test for the algorithm (without a valid error bound, a stopping test based on the corresponding residual says nothing about how close we are to a solution of the problem). Furthermore, it can be possible to obtain an algorithm-based error bound under assumptions weaker or different from the ones required for an algorithm-independent bound, as discussed next.

Let $C(\bar{x})$ be the critical cone of problem (1.1) at $\bar{x}$, that is,

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid F'(\bar{x})\xi = 0,\ \langle G_i'(\bar{x}),\, \xi\rangle \leq 0,\ i \in A(\bar{x}),\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\},$$

where $A(\bar{x}) = \{i = 1,\, \ldots,\, m \mid G_i(\bar{x}) = 0\}$ is the set of indices of inequality constraints active at $\bar{x}$. It is known ([8, Lemma 2], [5, Theorem 2]) that an algorithm-independent bound holds in a neighbourhood of $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ satisfying the second-order sufficient condition

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{1.4}$$

In that case, (1.3) holds with δ being the norm of the *natural residual* of the violation of the KKT conditions (1.2). We note that the cited result assumes only the above second-order sufficient condition, and in particular does not require any constraint qualifications. There exist also other error bounds under assumptions which subsume some constraint qualifications, see [12] for a detailed discussion and comparisons.

In the sequel, we shall derive algorithm-based error bounds related to the classical augmented Lagrangian and Lagrangian relaxation algorithms. We shall further provide examples showing that the given estimates are sharp, i.e., that they cannot be improved. Our analysis will assume either the first-order sufficient condition (FOSC):

$$C(\bar{x}) = \{0\}, \tag{1.5}$$

or the second-order sufficient condition (SOSC):

$$\forall\, \xi \in C(\bar{x}) \setminus \{0\}\ \ \exists\, (\lambda,\, \mu) \in \mathcal{M}(\bar{x}) \ \text{ s.t. } \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \lambda,\, \mu)\xi,\, \xi \right\rangle > 0. \tag{1.6}$$

Obviously, (1.6) is a weaker assumption than (1.4), as the latter needs existence of the "universal" multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$, suitable for all $\xi \in C(\bar{x}) \setminus \{0\}$. But perhaps even more importantly, even if the universal multipliers exist, we shall make no assumption of the dual sequence generated by the the given algorithm converging to the set of universal multipliers. In other words, the dual sequence will be allowed to approach the multipliers which do not satisfy (1.4), as long as (1.6) holds. In this situation, the algorithm-independent bound based on (1.4) is not applicable along the trajectory of the algorithm.

Of course, FOSC (1.5) implies both (1.4) and (1.6). But we shall consider FOSC separately, because it allows to obtain a better estimate. Note also that if $\mathcal{M}(\bar{x}) = \emptyset$, then FOSC and SOSC are formally equivalent.

2 Some Sensitivity Results

In this section, we state some recent sensitivity results [10], adapted for our purposes.

Let $\bar{x} \in \mathbf{R}^n$ be a strict local solution of (1.1), and let f, F and G be twice differentiable at $\bar{x}$. Fix some $\varepsilon > 0$ such that $\bar{x}$ is the unique (global) solution of the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0, \quad G(x) \leq 0, \quad x \in B_\varepsilon(\bar{x}), \end{array} \tag{2.1}$$

where $B_\varepsilon(\bar{x}) = \{x \in \mathbf{R}^n \mid \|x - \bar{x}\| \leq \varepsilon\}$.

For each pair $(y,\, z) \in \mathbf{R}^l \times \mathbf{R}^m$, consider the following perturbation of problem (2.1):

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = y, \quad G(x) \leq z, \quad x \in B_\varepsilon(\bar{x}). \end{array} \tag{2.2}$$

Let $\omega(y,\, z)$ and $S(y,\, z)$ stand for the optimal value and the solution set of problem (2.2), respectively.

In Theorem 2.1 below, we assume that for $(y,\, z) \in \mathbf{R}^l \times \mathbf{R}^m$, the following upper bound on the optimal value $\omega(y,\, z)$ holds:

$$\omega(y,\, z) \leq f(\bar{x}) + O(\|(y,\, z)\|). \tag{2.3}$$

If the point $\bar{x}$ satisfies the Mangasarian–Fromovitz constraint qualification, then (2.3) holds for arbitrary perturbations. However, in the absence of constraint qualifications, this property does not hold for arbitrary perturbations. But, as we show below, (2.3) is satisfied in the context of this paper. This is precisely the advantage of algorithm-based setting, which induces rather specific perturbations.

The following theorem is a direct consequence of [10, Theorems 2, 3].

Theorem 2.1 *Assume that* (2.3) *is satisfied.*

Then for $(y,\, z) \in \mathbf{R}^l \times \mathbf{R}^m$, the following assertions hold:

(i) If FOSC (1.5) *holds, then*

$$\sup_{x \in S(y,\, z)} \|x - \bar{x}\| = O(\|(y,\, z)\|),$$

$$\omega(y,\, z) = f(\bar{x}) + O(\|(y,\, z)\|). \tag{2.4}$$

(ii) If SOSC (1.6) *holds, then*

$$\sup_{x \in S(y,\, z)} \|x - \bar{x}\| = O(\|(y,\, z)\|^{1/2}),$$

and (2.4) *holds as well.*

3 Augmented Lagrangian

In this section, all norms are 2-norms. Let $c > 0$. The augmented Lagrangian for problem (1.1) is given by

$$L_c : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}, \quad L_c(x, \lambda, \mu) = f(x) + \langle \lambda, F(x) \rangle + \frac{c}{2}\|F(x)\|^2 + \frac{1}{2c}\sum_{i=1}^{m}((\max\{0, cG_i(x) + \mu_i\})^2 - \mu_i^2).$$

Given some $c > 0$ and $(\lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$, consider the associated subproblem

$$\begin{array}{ll} \text{minimize} & L_c(x, \lambda, \mu) \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{3.1}$$

We next establish the relationship between (3.1) and perturbations of the original problem (1.1).

Proposition 3.1 *For any $(\lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$ and $c > 0$, the point $x_{\lambda, \mu, c}$ which solves* (3.1) *is a solution of*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = F(x_{\lambda, \mu, c}), \\ & G_i(x) \le \max\{-\mu_i/c, G_i(x_{\lambda, \mu, c})\}, \ i = 1, \ldots, m. \end{array} \tag{3.2}$$

Proof. Clearly, $x_{\lambda, \mu, c}$ is feasible in (3.2). Assume that there exists $\hat{x}$ which is also feasible in (3.2) and it holds that $f(\hat{x}) < f(x_{\lambda, \mu, c})$.

We first show that

$$\max\{0, cG_i(\hat{x}) + \mu_i\} \le \max\{0, cG_i(x_{\lambda, \mu, c}) + \mu_i\}, \quad i = 1, \ldots, m. \tag{3.3}$$

The relation obviously holds if $cG_i(\hat{x}) + \mu_i \le 0$. Suppose $cG_i(\hat{x}) + \mu_i > 0$. Then $G_i(\hat{x}) > -\mu_i/c$, and the fact that $\hat{x}$ is feasible in (3.2) implies that $G_i(\hat{x}) \le G_i(x_{\lambda, \mu, c})$. Therefore, (3.3) holds also in that case.

Using (3.3) and the fact that $F(\hat{x}) = F(x_{\lambda, \mu, c})$, we further obtain

$$\begin{aligned} L_c(\hat{x}, \lambda, \mu) &= f(\hat{x}) + \langle \lambda, F(\hat{x}) \rangle + \frac{c}{2}\|F(\hat{x})\|^2 + \frac{1}{2c}\sum_{i=1}^{m}((\max\{0, cG_i(\hat{x}) + \mu_i\})^2 - \mu_i^2) \\ &< f(x_{\lambda, \mu, c}) + \langle \lambda, F(x_{\lambda, \mu, c}) \rangle + \frac{c}{2}\|F(x_{\lambda, \mu, c})\|^2 \\ &\quad + \frac{1}{2c}\sum_{i=1}^{m}((\max\{0, cG_i(x_{\lambda, \mu, c}) + \mu_i\})^2 - \mu_i^2) \\ &= L_c(x_{\lambda, \mu, c}, \lambda, \mu), \end{aligned}$$

which contradicts the definition of $x_{\lambda, \mu, c}$. ∎

The augmented Lagrangian algorithm (or the method of multipliers), see [9, 15, 16, 2, 4], is the following procedure. Given some $(\lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m_+$ and $c_k > 0$, the primal iterate x^k is generated by solving (3.1) with $(\lambda, \mu) = (\lambda^k, \mu^k)$, $c = c_k$. After this, the multiplier estimates are updated by

$$\lambda^{k+1} = \lambda^k + c_k F(x^k), \quad \mu_i^{k+1} = \max\{0,\, \mu_i^k + c_k G_i(x^k)\},\ i = 1, \ldots, m, \tag{3.4}$$

the parameter c_k is possibly adjusted, and the process is repeated.

Assuming that the iterative process described above generates a primal sequence $\{x^k\}$ converging to the strict local solution $\bar{x}$, and a bounded dual sequence $\{(\lambda^k,\, \mu^k)\}$, we are interested in quantifying the convergence, i.e., obtaining an estimate of the distance from x^k to $\bar{x}$ in terms of a known quantity. Recalling (3.2), for each k define

$$\delta_k = \left(\|F(x^k)\|^2 + \sum_{i=1}^{m} (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 \right)^{1/2}. \tag{3.5}$$

Theorem 3.1 *Let $c_k \geq \bar{c}$ for all k, where $\bar{c} > 0$ is arbitrary. Suppose that the sequence $\{(\lambda^k,\, \mu^k)\}$ generated according to* (3.4) *is bounded. For each k, let x^k be a solution of* (3.1) *with $(\lambda, \mu) = (\lambda^k, \mu^k)$, $c = c_k$. Suppose that the sequence $\{x^k\}$ converges to $\bar{x}$, which is a solution of* (1.1).

Then the following assertions hold:

(i) If FOSC (1.5) *holds, then*

$$\|x^k - \bar{x}\| = O(\delta_k), \tag{3.6}$$

$$f(x^k) = f(\bar{x}) + O(\delta_k). \tag{3.7}$$

(ii) If SOSC (1.6) *holds, then*

$$\|x^k - \bar{x}\| = O(\delta_k^{1/2}), \tag{3.8}$$

and (3.7) *holds as well.*

Proof. Since $\{x^k\}$ converges to $\bar{x}$, we can assume that $x^k \in B_\varepsilon(\bar{x})$ for all k, where ε is the same as in (2.1). We first show that under the given assumptions, $\delta_k \to 0$ as $k \to \infty$. Clearly, $\{F(x^k)\} \to 0$. For $i \in A(\bar{x})$, since $G_i(x^k) \to 0$ while $\mu_i^k/c_k \geq 0$, it holds that $\max\{-\mu_i^k/c_k,\, G_i(x^k)\} \to 0$ as $k \to \infty$. Let $i \notin A(\bar{x})$. Then for all k large enough, say $k \geq \bar{k}$, it holds that $G_i(x^k) \leq -\beta$, where $\beta > 0$. Suppose that there exists an infinite subsequence of indices $\{k_l\}$ such that $\mu_i^{k_l} > 0$ for all l. For $k_l > \bar{k}$, we obtain that

$$0 < \mu_i^{k_l} = \mu_i^{k_l - 1} + c_{k_l - 1} G_i(x^{k_l - 1}) = \mu_i^{\bar{k}} + \sum_{j=\bar{k}}^{k_l - 1} c_j G_i(x^j) \leq \mu_i^{\bar{k}} - \bar{c}\beta(k_l - 1 - \bar{k}),$$

which results in a contradiction for l sufficiently large (so that k_l is sufficiently large). We conclude that $\mu_i^k = 0$ for all $i \notin A(\bar{x})$ and k sufficiently large. Therefore, $\max\{-\mu_i^k/c_k,\, G_i(x^k)\} = 0$ for all $i \notin A(\bar{x})$ and k sufficiently large. This concludes the proof that $\delta_k \to 0$.

Next, note that for any $a, b \in \mathbf{R}$ and $c > 0$, it holds that

$$\frac{1}{2c}\left((\max\{0,\, ca+b\})^2 - b^2\right) = \frac{c}{2}(\max\{-b/c,\, a\})^2 + b\max\{-b/c,\, a\}.$$

Indeed, if $a \le -b/c$, then the right-hand side of the relation above becomes $b^2/(2c) - b^2/c = -b^2/(2c)$, which is the same as the left-hand side. If $a > -b/c$, then the right-hand side is equal to $(c^2a^2 + 2cab)/(2c) = ca^2/2 + ab$, which is the same as the left-hand side.

Using the above relation in the definition of L_{c_k}, we obtain

$$\begin{aligned}
f(x^k) &= L_{c_k}(x^k,\, \lambda^k,\, \mu^k) - \langle \lambda^k,\, F(x^k)\rangle - \frac{c_k}{2}\|F(x^k)\|^2 \\
&\quad -\frac{c_k}{2}\sum_{i=1}^m (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\} \\
&\le L_{c_k}(\bar{x},\, \lambda^k,\, \mu^k) - \langle \lambda^k,\, F(x^k)\rangle - \frac{c_k}{2}\|F(x^k)\|^2 \\
&\quad -\frac{c_k}{2}\sum_{i=1}^m (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\} \qquad (3.9) \\
&\le L_{c_k}(\bar{x},\, \lambda^k,\, \mu^k) - \langle \lambda^k,\, F(x^k)\rangle - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\}. \qquad (3.10)
\end{aligned}$$

Observe that

$$L_{c_k}(\bar{x},\, \lambda^k,\, \mu^k) = f(\bar{x}) + \frac{1}{2c_k}\sum_{i=1}^m \left((\max\{0,\, c_kG_i(\bar{x}) + \mu_i^k\})^2 - (\mu_i^k)^2\right).$$

For $i \in A(\bar{x})$, we have $\max\{0,\, c_kG_i(\bar{x}) + \mu_i^k\}) = \max\{0,\, \mu_i^k\} = \mu_i^k$, because $\mu_i^k \ge 0$.

For $i \notin A(\bar{x})$, we have $\max\{0,\, c_kG_i(\bar{x}) + \mu_i^k\} = \max\{0,\, c_kG_i(\bar{x})\} = 0 = \mu_i^k$, which holds for all k sufficiently large, because for such k, as shown above, we have that $\mu_i^k = 0$.

It follows that for all k sufficiently large,

$$L_{c_k}(\bar{x},\, \lambda^k,\, \mu^k) = f(\bar{x}).$$

Hence, from (3.10) we obtain that

$$\begin{aligned}
f(x^k) &\le f(\bar{x}) - \langle \lambda^k,\, F(x^k)\rangle - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\} \\
&= f(\bar{x}) + O(\delta_k),
\end{aligned}$$

where we have taken into account the boundedness of $\{(\lambda^k,\, \mu^k)\}$, and where δ_k is defined by (3.5).

Taking into account Proposition 3.1 and applying Theorem 2.1 to the perturbation (3.2) of problem (1.1), we obtain the announced results. ■

We note that in the given context, a sufficient condition for the dual sequence $\{(\lambda^k,\, \mu^k)\}$ generated according to (3.4) to be bounded is the Mangasarian–Fromovitz constraint qualification at $\bar{x}$. Indeed, by the necessary optimality condition for problem (3.1), we have that

$$\begin{aligned}
0 &= \frac{\partial L_{c_k}}{\partial x}(x^k,\, \lambda^k,\, \mu^k) \\
&= f'(x^k) + (F'(x^k))^{\mathrm{T}}\lambda^k + c_k(F'(x^k))^{\mathrm{T}}F(x^k) + \sum_{i=1}^{m}\max\{0,\, c_kG_i(x^k) + \mu_i^k\}G_i'(x^k) \\
&= f'(x^k) + (F'(x^k))^{\mathrm{T}}\lambda^{k+1} + \sum_{i\in A(\bar{x})}\mu_i^{k+1}G_i'(x^k), \qquad (3.11)
\end{aligned}$$

where we have taken into account that $\mu_i^k = 0$ for all k large enough and $i \notin A(\bar{x})$, as shown above. If $\{x^k\} \to \bar{x}$ while $\{(\lambda^k,\, \mu^k)\}$ is unbounded, dividing the above relation by $\|(\lambda^{k+1},\, \mu^{k+1})\|$ and passing onto the limit as $k \to \infty$ (possibly along an appropriate subsequence), we obtain the existence of some $(\bar{\lambda},\, \bar{\mu}) \in (\mathbf{R}^l \times \mathbf{R}_+^{|A(\bar{x})|}) \setminus \{0\}$ such that

$$0 = (F'(\bar{x}))^{\mathrm{T}}\bar{\lambda} + \sum_{i\in A(\bar{x})}\bar{\mu}_iG_i'(\bar{x}),$$

which contradicts the (dual form of) Mangasarian–Fromovitz constraint qualification at the point $\bar{x}$.

In Theorem 3.1, we do not assume that $\mathcal{M}(\bar{x}) \neq \emptyset$, but in fact, this must be the case under the stated assumptions. Indeed, from (3.11) we obtain that for every limit point $(\bar{\lambda},\, \bar{\mu})$ of $\{(\lambda^k,\, \mu^k)\}$, it holds that

$$f'(\bar{x}) + (F'(\bar{x}))^{\mathrm{T}}\bar{\lambda} + \sum_{i\in A(\bar{x})}\bar{\mu}_iG_i'(\bar{x}) = 0.$$

Moreover, as mentioned above, for all k large enough we have $\mu_i^k = 0$ for all $i \notin A(\bar{x})$, and $\mu^k \geq 0$ according to (3.4). Hence, $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$.

The following observation can also be useful. By the direct computation, from (3.4) it can be seen that

$$\delta_k = \frac{1}{c_k}\|(\lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\|.$$

From this representation of δ_k, it is evident that $\delta_k \to 0$ provided $\{(\lambda^k,\, \mu^k)\}$ is bounded while $c_k \to \infty$, or provided $\{(\lambda^k,\, \mu^k)\}$ is convergent while c_k is bounded away from zero, without *any* assumptions on $\{x^k\}$. Note, however, that the assumptions of Theorem 3.1 above are different, and so $\delta_k \to 0$ had to be established by other considerations.

The next result is more in the spirit of penalty methods. We do not consider any specific rule for updating the dual variables, and in particular, they can even be fixed (the classical penalty method is formally obtained by setting $(\lambda^k,\, \mu^k) = 0$ for all k in the definition of L_{c_k}). In that setting, to ensure convergence it in general must hold that $c_k \to +\infty$ as $k \to \infty$.

Theorem 3.2 *Suppose that the sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded and that $c_k \to +\infty$ as $k \to \infty$. For each k, let x^k be a solution of* (3.1) *with $(\lambda,\, \mu) = (\lambda^k,\, \mu^k)$, $c = c_k$. Suppose that the sequence $\{x^k\}$ converges to $\bar{x}$, which is a solution of* (1.1).

Then both assertions of Theorem 3.1 hold. Moreover, under (3.7) *(in particular, if SOSC* (1.6) *holds), we have that*

$$\delta_k = O(1/c_k). \qquad (3.12)$$

Proof. Again, we can assume that $x^k \in B_\varepsilon(\bar{x})$ for all k. Because $\{\mu^k\}$ is bounded and $c_k \to +\infty$, it holds that $\mu_i^k/c_k \to 0$. Since $\{G(x^k)\} \to G(\bar{x}) \leq 0$, we conclude that $\max\{-\mu_i^k/c_k,\, G_i(x^k)\} \to 0$ for all $i = 1, \ldots, m$. Taking also into account that $\{F(x^k)\} \to 0$, it follows that $\delta_k \to 0$ as $k \to \infty$ (alternatively, the argument preceding the theorem could be used here).

As in the proof of Theorem 3.1, we have the relation (3.10). However, it now holds that

$$\begin{aligned} L_{c_k}(\bar{x},\, \lambda^k,\, \mu^k) &= f(\bar{x}) + \frac{1}{2c_k}\sum_{i=1}^m \Big((\max\{0,\, c_k G_i(\bar{x}) + \mu_i^k\})^2 - (\mu_i^k)^2\Big) \\ &= f(\bar{x}) - \frac{1}{2c_k}\sum_{i \notin A(\bar{x})} (\mu_i^k)^2 \\ &\leq f(\bar{x}), \end{aligned} \tag{3.13}$$

where the second equality follows from the fact that $c_k G_i(\bar{x}) + \mu_i^k < 0$ for all $i \notin A(\bar{x})$ and all k large enough (recall that $c_k \to +\infty$ while $\{\mu^k\}$ is bounded).

Combining this relation with (3.10), we again obtain that

$$\begin{aligned} f(x^k) &\leq f(\bar{x}) - \langle \lambda^k,\, F(x^k)\rangle - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\} \\ &= f(\bar{x}) + O(\delta_k), \end{aligned}$$

and the assertions of Theorem 3.1 follow.

Suppose now that (3.7) holds. Then combining (3.9) and (3.13), we obtain that

$$\begin{aligned} & f(\bar{x}) + O(\delta_k) = f(x^k) \\ \leq\ & f(\bar{x}) - \frac{1}{2c_k}\sum_{i \notin A(\bar{x})} (\mu_i^k)^2 - \langle \lambda^k,\, F(x^k)\rangle - \frac{c_k}{2}\|F(x^k)\|^2 \\ & -\frac{c_k}{2}\sum_{i=1}^m (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 - \sum_{i=1}^m \mu_i^k \max\{-\mu_i^k/c_k,\, G_i(x^k)\} \\ =\ & f(\bar{x}) - \frac{1}{2c_k}\sum_{i \notin A(\bar{x})} (\mu_i^k)^2 - \frac{c_k}{2}\|F(x^k)\|^2 - \frac{c_k}{2}\sum_{i=1}^m (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 + O(\delta_k). \end{aligned}$$

Hence,

$$\frac{1}{2c_k}\sum_{i \notin A(\bar{x})} (\mu_i^k)^2 + \frac{c_k}{2}\|F(x^k)\|^2 + \frac{c_k}{2}\sum_{i=1}^m (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2 = O(\delta_k). \tag{3.14}$$

Evidently, for all $i \notin A(\bar{x})$ and each k large enough, it holds that

$$\max\{-\mu_i^k/c_k,\, G_i(x^k)\} = -\mu_i^k/c_k.$$

Hence,

$$\frac{1}{2c_k}\sum_{i \notin A(\bar{x})} (\mu_i^k)^2 = \frac{c_k}{2}\sum_{i \notin A(\bar{x})} (\max\{-\mu_i^k/c_k,\, G_i(x^k)\})^2,$$

and recalling the definition of δ_k, it now follows from (3.14) that

$$c_k \delta_k^2 = O(\delta_k).$$

Therefore, (3.12) holds. ■

Note that while $\mathcal{M}(\bar{x}) \neq \emptyset$ is not assumed in Theorem 3.2, this is again the case in the setting of (3.7) (the second part of Theorem 3.2). Indeed, defining the auxiliary sequences

$$\tilde{\lambda}^k = \lambda^k + c_k F(x^k), \quad \tilde{\mu}_i^k = \max\{0,\, \mu_i^k + c_k G_i(x^k)\},\ i = 1,\, \ldots,\, m, \tag{3.15}$$

it can be seen by direct computation that

$$\delta_k = \frac{1}{c_k} \|(\tilde{\lambda}^k - \lambda^k,\, \tilde{\mu}^k - \mu^k)\|. \tag{3.16}$$

The last assertion of Theorem 3.2 implies that the sequence $\{c_k \delta_k\}$ is bounded. Hence, (3.16) combined with boundedness of $\{(\lambda^k,\, \mu^k)\}$ implies boundedness of $\{(\tilde{\lambda}^k,\, \tilde{\mu}^k)\}$. The fact that every limit point of the latter sequence belongs to $\mathcal{M}(\bar{x})$ can be established the same way as its counterpart for the method of multipliers; see the discussion following the proof of Theorem 3.1.

The following observation is also worth mentioning. Suppose that under the assumptions of Theorem 3.2, $\mathcal{M}(\bar{x})$ is a singleton, i.e., $\mathcal{M}(\bar{x}) = \{(\bar{\lambda},\, \bar{\mu})\}$. Then $\{\tilde{\lambda}^k,\, \tilde{\mu}^k\}$ must converge to this $(\bar{\lambda},\, \bar{\mu})$, whatever one takes as $\{(\lambda^k,\, \mu^k)\}$. Moreover, if good multiplier approximations are used, i.e., $\{(\lambda^k,\, \mu^k)\} \to (\bar{\lambda},\, \bar{\mu})$, then estimate (3.12) can be sharpened. Specifically, from (3.16) it follows that

$$\delta_k = o(1/c_k). \tag{3.17}$$

We complete this section with two examples demonstrating that the estimates obtained above are sharp. The first example below satisfies FOSC, while the second satisfies SOSC. It is interesting to note that the second example does not satisfy constraint qualifications (the multiplier set is unbounded).

Example 3.1 Let $n = l = 1$, $m = 0$, $f(x) = x + x^2/2$, $F(x) = x$. The only feasible point (and hence, the only solution) of (1.1) is $\bar{x} = 0$. Furthermore, $\mathcal{M}(\bar{x}) = \{-1\}$, and $C(\bar{x}) = \ker F'(\bar{x}) = \{0\}$, that is, FOSC (1.5) holds.

It can be easily seen that for each $\lambda \in \mathbf{R}$ and $c > 0$, the only solution of (3.1) is given by $x_{\lambda,\, c} = -(\lambda + 1)/(c + 1)$.

We first consider the case when $\{\lambda^k\} \subset \mathbf{R}$ is generated according to the first equality in (3.4), while $c > 0$ is fixed. Let $x^k = x_{\lambda^k,\, c}$ for each k. Then

$$\begin{aligned} \lambda^{k+1} + 1 &= \lambda^k + c x^k + 1 \\ &= \lambda^k - c\frac{\lambda^k + 1}{c + 1} + 1 \\ &= \frac{\lambda^k + 1}{c + 1}, \end{aligned}$$

and since $c > 0$, it is now evident that $\{\lambda^k\} \to \bar{\lambda} = -1$, which is the only multiplier associated with $\bar{x}$. Hence, $\{x^k\} \to \bar{x}$, $\delta_k = |x^k| \to 0$, $f(x^k) = x^k + o(|x^k|)$, and the estimates (3.6), (3.7) are exact (for the latter, note that for each k, $x^k = \delta_k$ if $\lambda^0 < -1$, and $x^k = -\delta_k$ if $\lambda^0 > -1$).

We next consider the case when $\lambda \in \mathbf{R}$ is fixed, while $c_k \to +\infty$. Let $x^k = x_{\lambda,\, c_k}$ for each k. All the conclusions for the previous case remain valid (note that for each k, $x^k = \delta_k$ if $\lambda < -1$, and $x^k = -\delta_k$ if $\lambda > -1$). Moreover, for each k,

$$\tilde{\lambda}^k = \frac{\lambda + 1}{c_k + 1},$$

where $\tilde{\lambda}^k$ is defined according to the first equality in (3.15). Clearly, $\{\tilde{\lambda}^k\} \to \bar{\lambda}$. Finally, estimate (3.12) obviously holds and is in general sharp (it can be improved if we let λ tend to $\bar{\lambda}$, in which case (3.17) is valid).

Example 3.2 Let $n = l = 2$, $m = 0$, $f(x) = x_1 + x_1^2/2 + x_2^4/2$, $F(x) = (x_1, x_2^2)$. The only feasible point (and hence, the only solution) of (1.1) is $\bar{x} = 0$. Furthermore, $\mathcal{M}(\bar{x}) = \{-1\} \times \mathbf{R}$, and it can be easily verified that SOSC (1.6) holds.

It can be easily seen that for each $\lambda \in \mathbf{R}^2$ and $c > 0$, the only solution of (3.1) is given by

$$x_{\lambda,\, c} = \begin{cases} \left(-\frac{\lambda_1+1}{c+1},\, 0\right) & \text{if } \lambda_2 \geq 0, \\ \left(-\frac{\lambda_1+1}{c+1},\, \pm\left(\frac{-\lambda_2}{c+1}\right)^{1/2}\right) & \text{if } \lambda_2 < 0. \end{cases}$$

Consider the case when $\{\lambda^k\} \subset \mathbf{R}^2$ is generated according to the first equality in (3.4), while $c > 0$ is fixed. Let $x^k = x_{\lambda^k,\, c}$ for each k. Let $\bar{\lambda} = (-1,\, 0) \in \mathcal{M}(\bar{x})$. Then, assuming that $\lambda_2^k < 0$, we obtain

$$\begin{aligned} \lambda^{k+1} - \bar{\lambda} &= (\lambda_1^k + c x_1^k,\, \lambda_2^k + c(x_2^k)^2) - \bar{\lambda} \\ &= \left(\lambda_1^k - c\frac{\lambda_1^k + 1}{c+1},\, \lambda_2^k - c\frac{\lambda_2^k}{c+1}\right) - \bar{\lambda} \\ &= \left(\frac{\lambda_1^k + 1}{c+1},\, \frac{\lambda_2^k}{c+1}\right) \\ &= \frac{\lambda^k - \bar{\lambda}}{c+1}. \end{aligned}$$

In particular, if $\lambda_2^0 < 0$, then λ_2^k remains negative for each k, and since $c > 0$, it is evident that $\{\lambda^k\} \to \bar{\lambda}$. Hence, $\{x^k\} \to \bar{x}$ and

$$\delta_k = \frac{((\lambda_1^k + 1)^2 + (\lambda_2^k)^2)^{1/2}}{c+1} \to 0.$$

Fix an arbitrary $\theta > 0$ and take $\lambda^0 = (-1 \pm \theta,\, -\theta)$. Then

$$\lambda^k = \left(-1 + \frac{\pm\theta}{(c+1)^k},\, -\frac{\theta}{(c+1)^k}\right),$$

$$x^k = \left(\frac{\mp\theta}{(c+1)^{k+1}},\, \left(\frac{\theta}{(c+1)^{k+1}} \right)^{1/2} \right),$$

$$f(x^k) = \frac{\mp\theta}{(c+1)^{k+1}} + o\left(\frac{1}{(c+1)^{k+1}} \right),$$

$$\delta_k = \frac{\sqrt{2}\theta}{(c+1)^{k+1}},$$

and the estimates (3.7), (3.8) are exact.

Consider now the case when $\lambda \in \mathbf{R}^2$ is fixed, while $c_k \to +\infty$. Fix an arbitrary $\theta > 0$ and take $\lambda = (-1 \pm \theta,\, -\theta)$. Let $x^k = x_{\lambda,\, c_k}$ for each k. All the conclusions for the previous case remain valid. Moreover, for each k,

$$\tilde{\lambda}^k = \left(-1 + \frac{\pm\theta}{c_k + 1},\, -\frac{\theta}{c_k + 1} \right),$$

where $\tilde{\lambda}^k$ is defined according to the first equality in (3.15). Clearly, $\{\tilde{\lambda}^k\} \to \bar{\lambda}$. Finally, estimate (3.12) obviously holds and is in general sharp (it can be improved if we let λ tend to $\bar{\lambda}$, that is, let θ to tend to 0, in which case (3.17) is valid).

Finally, consider the case when $\lambda = (-1,\, -1)$. For each k, set $x^k = x_{\lambda,\, c_k}$, and note that $x_1^k = 0$ while $x_2^k \neq 0$. At $(x^k,\, \lambda)$, the residual of the optimality system (1.2) is of order $(x_2^k)^2$. Hence, the algorithm-independent error bound of [8] does not hold along the sequence $\{(x^k,\, \lambda)\}$. At the same time, $\delta_k = (x_2^k)^2$, and the estimate (3.8) holds and is exact. The same conclusions apply if λ^k is not fixed at $\lambda = (-1,\, -1)$, but the generated dual sequence tends to this λ.

4 Lagrangian Relaxation

The approach of Lagrangian relaxation [3, Chapter 7] is a useful tool for solving various classes of optimization problems [13]. It consists of solving (1.1) via solving its dual

$$\begin{array}{ll} \text{maximize} & \varphi(\lambda,\, \mu) \\ \text{subject to} & (\lambda,\, \mu) \in \Delta, \end{array} \tag{4.1}$$

where

$$\Delta = \{(\lambda,\, \mu) \in \mathbf{R}^l \times \mathbf{R}_+^m \mid \varphi(\lambda,\, \mu) > -\infty\},$$

$$\varphi : \mathbf{R}^l \times \mathbf{R}_+^m \to \mathbf{R}, \quad \varphi(\lambda,\, \mu) = \inf_{x \in \mathbf{R}^n} L(x,\, \lambda,\, \mu).$$

The dual problem (4.1) is a concave maximization problem, in general nonsmooth, which is solved by appropriate subgradient or bundle methods [3, Chapter 7]. Bundle methods are in general more reliable and practical, although the subgradient methods are also sometimes useful, thanks to their simplicity of implementation. Below we shall consider explicitly the subgradient method only, because introducing the more sophisticated bundle methods would have required an extensive discussion which is secondary to the subject of this paper. However, we believe that bundle methods can also be treated within our framework.

We shall assume that the problem

$$\begin{array}{ll} \text{minimize} & L(x,\,\lambda,\,\mu) \\ \text{subject to} & x \in \mathbf{R}^n \end{array} \tag{4.2}$$

has a solution $x_{\lambda,\,\mu}$ for every $(\lambda,\,\mu) \in \mathbf{R}^l \times \mathbf{R}^m_+$ of interest (artificial bounds on x are often introduced to guarantee this, so that the dual function φ is finite everywhere).

The following relation between (4.2) and perturbations of the original problem (1.1) is well known (we include its short proof for completeness).

Proposition 4.1 *[3, Theorem 10.1] For any $(\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$, a point $x_{\lambda,\,\mu}$ which solves (4.2), is a solution of*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = F(x_{\lambda,\,\mu}), \\ & G_i(x) \le \left\{ \begin{array}{ll} G_i(x_{\lambda,\,\mu}), & \text{if } \mu_i > 0, \\ \max\{0,\,G_i(x_{\lambda,\,\mu})\}, & \text{if } \mu_i = 0, \end{array} \right. \\ & i = 1,\,\ldots,\,m. \end{array} \tag{4.3}$$

Proof. For any $x \in \mathbf{R}^n$ which is feasible in (4.3), we obtain that

$$\begin{array}{rcl} f(x_{\lambda,\,\mu}) - f(x) & \le & f(x_{\lambda,\,\mu}) - f(x) + \sum\limits_{\mu_i > 0} \mu_i (G_i(x_{\lambda,\,\mu}) - G_i(x)) \\ & = & f(x_{\lambda,\,\mu}) - f(x) + \langle \lambda,\,F(x_{\lambda,\,\mu}) - F(x) \rangle + \langle \mu,\,G(x_{\lambda,\,\mu}) - G(x) \rangle \\ & = & L(x_{\lambda,\,\mu},\,\lambda,\,\mu) - L(x,\,\lambda,\,\mu). \end{array}$$

Since $x_{\lambda,\,\mu}$ is feasible in (4.3), the assertion follows. ∎

As is well known [3, Chapter 7] and easy to see, solving (4.2) provides not only the value of φ at the point $(\lambda,\,\mu)$, but also (at no additional cost) of one of its subgradients. In particular, it holds that

$$(F(x_{\lambda,\,\mu}),\,G(x_{\lambda,\,\mu})) \in \partial\varphi(\lambda,\,\mu). \tag{4.4}$$

The subgradient Lagragian relaxation based method is the following iterative procedure. Given some $(\lambda^k,\,\mu^k) \in \mathbf{R}^l \times \mathbf{R}^m_+$, the first part of the iteration consists of solving the minimization problem (4.2) with $(\lambda,\,\mu) = (\lambda^k,\,\mu^k)$. This generates a primal point, which we shall denote x^k. After this, the dual variables are updated by the projected subgradient step

$$(\lambda^{k+1},\,\mu^{k+1}) = P_{\mathbf{R}^l \times \mathbf{R}^m_+}\left((\lambda^k,\,\mu^k) + \alpha_k g^k\right), \quad g^k \in \partial\varphi(\lambda^k,\,\mu^k),\ \alpha_k > 0,$$

where $P_{\mathbf{R}^l \times \mathbf{R}^m_+}$ denotes the orthogonal projection onto the set $\mathbf{R}^l \times \mathbf{R}^m_+$, and α_k is the stepsize. In particular, the implementation based on (4.4) gives

$$\lambda^{k+1} = \lambda^k + \alpha_k F(x^k), \quad \mu_i^{k+1} = \max\{0,\,\mu_i^k + \alpha_k G_i(x^k)\},\ i = 1,\,\ldots,\,m. \tag{4.5}$$

Theoretically, for convergence of the subgradient projection method, the stepsize sequence $\{\alpha_k\}$ must satisfy

$$\sum_{k=0}^{\infty} \alpha_k \|g^k\| = +\infty \tag{4.6}$$

and

$$\sum_{k=0}^{\infty} \alpha_k^2 \|g^k\|^2 < +\infty. \tag{4.7}$$

More precisely, if the dual problem (4.1) is solvable, and conditions (4.6) and (4.7) hold, then the method generates the trajectory which converges to a solution of (4.1). If, in addition, the solution set of problem (4.1) is bounded, then trajectory converges to this set even if (4.7) is replaced by a weaker condition

$$\lim_{k\to\infty} \alpha_k \|g^k\| = 0.$$

In our setting, we assume that the sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded and the corresponding sequence $\{x^k\}$ converges to $\bar{x}$, the strict local solution of (1.1) under consideration. The latter is not automatic, but can be expected to happen in some situations (for example, if convexity and constraint qualifications are assumed). Under these assumptions, the sequence $\{g^k\}$ is automatically bounded, and (4.6) implies

$$\sum_{k=0}^{\infty} \alpha_k = +\infty, \tag{4.8}$$

which is the condition to be used below.

In this setting, we are again interested in obtaining an estimate of the distance from x^k to $\bar{x}$. Recalling (4.3), define

$$\delta_k = \left(\|F(x^k)\|^2 + \sum_{\mu_i^k > 0} (G_i(x^k))^2 + \sum_{\mu_i^k = 0} (\max\{0,\, G_i(x^k)\})^2 \right)^{1/2}. \tag{4.9}$$

Theorem 4.1 *Suppose that the solution set of the dual problem* (4.1) *is nonempty and that the sequence* $\{(\lambda^k,\, \mu^k)\}$ *is generated according to* (4.5) *and* (4.8)*, where for each* k*,* x^k *is a solution of* (4.2) *for* $(\lambda, \mu) = (\lambda^k, \mu^k)$*. Suppose that the sequence* $\{x^k\}$ *converges to* $\bar{x}$*, which is a solution of* (1.1)*.*

Then all the assertions of Theorem 3.1 hold, with δ_k *defined by* (4.9)*.*

Proof. As before, we can assume that $x^k \in B_\varepsilon(\bar{x})$ for all k. To prove that $\delta_k \to 0$ as $k \to \infty$, we first show that $\mu_i^k = 0$ for all $i \notin A(\bar{x})$ and all k sufficiently large. Let $i \notin A(\bar{x})$ and suppose the opposite, i.e., that there exists an infinite subsequence of indices $\{k_l\}$ such that $\mu_i^{k_l} > 0$ for all l. For all k large enough, say $k \geq \bar{k}$, it holds that $G_i(x^k) \leq -\beta$, where $\beta > 0$. Then for $k_l > \bar{k}$, we obtain that

$$0 < \mu_i^{k_l} = \mu_i^{k_l - 1} + \alpha_{k_l - 1} G_i(x^{k_l - 1}) = \mu_i^{\bar{k}} + \sum_{j=\bar{k}}^{k_l - 1} \alpha_j G_i(x^j) \leq \mu_i^{\bar{k}} - \beta \sum_{j=\bar{k}}^{k_l - 1} \alpha_j,$$

which results in a contradiction when $l \to \infty$ (and then $k_l \to \infty$), because of the first condition in (4.8). It follows that if $\mu_i^k > 0$ for some k large enough, then $i \in A(\bar{x})$, so that $G_i(x^k) \to 0$. It now easily follows that $\delta_k \to 0$.

We have that

$$\begin{aligned} f(x^k) + \langle \lambda^k, F(x^k) \rangle + \langle \mu^k, G(x^k) \rangle &= L(x^k, \lambda^k, \mu^k) \\ &\leq L(\bar{x}, \lambda^k, \mu^k) \\ &= f(\bar{x}) + \langle \lambda^k, F(\bar{x}) \rangle + \langle \mu^k, G(\bar{x}) \rangle \\ &\leq f(\bar{x}), \end{aligned}$$

where the first inequality is by the definition of x^k, and the last follows from $F(\bar{x}) = 0$, $G(\bar{x}) \leq 0$ and $\mu^k \geq 0$. Hence,

$$f(x^k) \leq f(\bar{x}) - \langle \lambda^k, F(x^k) \rangle - \langle \mu^k, G(x^k) \rangle.$$

In particular, taking into account the boundedness of $\{(\lambda^k, \mu^k)\}$ (this sequence converges, by our assumptions and the properties of the subgradient projection method) and the definition (4.9) of δ_k, we conclude that

$$f(x^k) \leq f(\bar{x}) + O(\delta_k).$$

The assertions follow from Proposition 4.1 and Theorem 2.1. ■

It can be verified that Examples 3.1 and 3.2 also show that the corresponding estimates obtained for the subgradient Lagrangian relaxation based method are sharp. We shall not provide any specifics, as conceptually the situation is very similar to the case of the augmented Lagrangian, considered in full detail in Section 3.

References

[1] M. Anitescu. Nonlinear programs with unbounded Lagrange multiplier sets, 2002. Preprint ANL/MCS-P796-0200, Mathematics and Computer Science Devision, Argonne National Laboratory.

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

[3] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects.* Springer–Verlag, Berlin, Germany, 2003.

[4] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal of Numerical Analysis*, 28:545–572, 1991.

[5] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming*, 94:91–124, 2002.

[6] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints, 2002. Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee.

[7] W.W. Hager. Stabilized sequential quadratic programming. *Computational Optimization and Applications*, 12:253–273, 1999.

[8] W.W. Hager and M.S. Gowda. Stability in the presence of degeneracy and error estimation. *Mathematical Programming*, 85:181–192, 1999.

[9] M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–330, 1969.

[10] A.F. Izmailov. Mathematical programs with complementarity constraints: regularity, optimality comditions, and sensitivity. *Computational Mathematics and Mathematical Physics*, 44:1145–1164, 2004.

[11] A.F. Izmailov and M.V. Solodov. Newton-type methods for optimization problems without constraint qualifications. *SIAM Journal on Optimization*, 15:210–228, 2004.

[12] A.F. Izmailov and M.V. Solodov. Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. *Mathematical Programming*, 95:631–650, 2003.

[13] C. Lemaréchal. Lagrangian relaxation. In *Computational Combinatorial Optimization,* Lecture Notes in Computer Science, 2241, pages 112–156. Springer, Berlin, 2001.

[14] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[15] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, London, 1969.

[16] R.T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[17] S.J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. Optimization Technical Report 03-02, Computer Sciences Department, University of Wisconsin-Madison, October 2003.