ISSN 0965-5425, Computational Mathematics and Mathematical Physics, 2012, Vol. 52, No. 12, pp. 1603–1611. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © A.F. Izmailov, A.S. Kurennoy, 2012, published in Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 2012, Vol. 52, No. 12, pp. 2140–2148.

Multiplier Methods for Optimization Problems with Lipschitzian Derivatives

A. F. Izmailov and A. S. Kurennoy

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119992 Russia
e-mail: izmaf@ccas.ru, alex-kurennoy@yandex.ru

Received May 30, 2012

Abstract—Optimization problems for which the objective function and the constraints have locally Lipschitzian derivatives but are not assumed to be twice differentiable are examined. For such problems, analyses of the local convergence and the convergence rate of the multiplier (or the augmented Lagrangian) method and the linearly constraint Lagrangian method are given.

DOI: 10.1134/S0965542512120081

Keywords: mathematical programming problem, augmented Lagrangian, multiplier method, linearized constraints, Newtonian iterative scheme.

1. INTRODUCTION. THE MULTIPLIER METHOD AND THE LINEARLY CONSTRAINED LAGRANGIAN METHOD

Consider the mathematical programming problem

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \tag{1.1}$$

in which the objective function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ and the mappings $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$ and $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ have locally Lipschitzian first derivatives but are not necessarily twice differentiable. Problems of this smoothness arise in numerous applications such as stochastic programming and optimal control (the so-called extended linear-quadratic problems; see [1–3]), semi-infinite programming and primal decomposition procedures (see [4, 5] and the references therein), smooth and “lifted” reformulations of complementarity constraints (see [6–8]), and so on. Below, we discuss certain methods for solving problem (1.1) and give an analysis of their local convergence, which seems to be the first one under the indicated smoothness requirements.

Let $L: \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \longrightarrow \mathbb{R}$ be the Lagrangian of problem (1.1):

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle.$$

Then the stationary points of this problem and the corresponding Lagrange multipliers are given by the Karush–Kuhn–Tucker (KKT) system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \ge 0, \quad g(x) \le 0, \quad \langle \mu, g(x) \rangle = 0. \tag{1.2}$$

Denote by

$$A(\bar{x}) = \{ i = 1, \ldots, m \,|\, g_i(\bar{x}) = 0 \}$$

the set of indices of active constraints for a feasible point $\bar{x} \in \mathbb{R}^n$ of problem (1.1). Let $\bar{x}$ be a stationary point for this problem. For every Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$, corresponding to $\bar{x}$, the symbols

$$A_+(\bar{x}, \bar{\mu}) = \{ i \in A(\bar{x}) \,|\, \bar{\mu}_i > 0 \}, \quad A_0(\bar{x}, \bar{\mu}) = \{ i \in A(\bar{x}) \,|\, \bar{\mu}_i = 0 \}$$

denote the sets of indices for strongly and weakly active constraints, respectively.

We say that the linear independence constraint qualification is fulfilled at a stationary point of problem (1.1) if the gradients of equality constraints and active inequality constraints are linearly independent at this point. This condition ensures the uniqueness of the Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$, corresponding to $\bar{x}$.

We also define the family of augmented Lagrangians $L_c: \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \longrightarrow \mathbb{R}$ for problem (1.1):

$$L_c(x, \lambda, \mu) = f(x) + \frac{1}{2c}(\|\lambda + ch(x)\|_2^2 + \|\max\{0, \mu + cg(x)\}\|_2^2).$$

Here, $c > 0$ is the penalty parameter and the max operation is performed componentwise.

Based on the current dual approximation $(\lambda^k, \mu^k) \in \mathbb{R}^l \times \mathbb{R}^m$ and the current penalty parameter $c_k > 0$, the multiplier method generates the next primal-dual approximation $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ in the following manner: x^{k+1} is found as a stationary point of the optimization problem

$$L_{c_k}(x, \lambda^k, \mu^k) \longrightarrow \min, \quad x \in \mathbb{R}^n, \tag{1.3}$$

while λ^{k+1} and μ^{k+1} are calculated using the formulas

$$\lambda^{k+1} = \lambda^k + c_k h(x^{k+1}), \quad \mu^{k+1} = \max\{0, \mu^k + c_k g(x^{k+1})\}. \tag{1.4}$$

This method underlies a number of well-known packages such as LANCELOT (see [9]) and ALGENCAN (see [10]).

There are well-known results concerning the local convergence of the augmented Lagrangian method in the case where the function f and the mappings h and g are twice differentiable. In [11], this issue was treated under the assumption that the linear independence constraint qualification, the strict complementarity condition $A_0(\bar{x}, \bar{\mu}) = \varnothing$, and the second-order sufficient optimality condition are fulfilled (see also [12, Theorem 4.7.4]). It was recently shown that the method converges locally under the sole second-order sufficient condition if one requires additionally that the current and next approximations be always sufficiently close (see [13]). Moreover, the convergence rate is guaranteed to be linear and even superlinear if the penalty parameter c_k tends to infinity as $k \longrightarrow \infty$.

In Section 3, we prove a result on the local convergence and the convergence rate of the multiplier method with no assumption about twice differentiability and admitting the possible violation of the strict complementarity condition. Our analysis uses the linear independence condition and the following version of the strong second-order sufficient condition:

$$\forall H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}), \quad \langle H\xi, \xi\rangle > 0 \quad \forall \xi \in C_+(\bar{x}, \bar{\mu}) \backslash \{0\}. \tag{1.5}$$

Here, the symbol ∂_x denotes the partial Clarke differential with respect to x (see [14]) and

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0, g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0\}.$$

If f, h, and g are twice differentiable, then (1.5) converts into the more conventional condition

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C_+(\bar{x}, \bar{\mu}) \backslash \{0\}.$$

The linearly constrained Lagrangian method (see [15–17]) is traditionally introduced for problem (1.1) in the case where $m = n$ and $g(x) = -x$, $x \in \mathbb{R}^n$, that is, for a problem with equality constraints and the nonnegativity condition for the variables:

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad x \geq 0. \tag{1.6}$$

Based on the current approximation $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and the current penalty parameter $c_k \geq 0$, the method calculates the next primal approximation $x^{k+1} \in \mathbb{R}^n$ as a stationary point of the optimization problem

$$f(x) + \langle \lambda^k, h(x)\rangle + \frac{c_k}{2}\|h(x)\|_2^2 \longrightarrow \min, \quad h(x^k) + h'(x^k)(x - x^k) = 0, \quad x \geq 0, \tag{1.7}$$

while the next point $(\lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^n$ on the dual trajectory is chosen so that the pair $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is a Lagrange multiplier, corresponding to x^{k+1}. It is this method that underlies the very successful package MINOS (see [18]). The objective function of problem (1.7) is the augmented Lagrangian that involves only the equality constraints of problem (1.6).

For twice differentiable f and h, the finest result on the local convergence of this method was obtained in [19]. The local superlinear convergence was proved in [19] under the assumptions that the multiplier is unique and the second-order sufficient condition is fulfilled.

In Section 4, we prove the local convergence of the linearly constrained Lagrangian method under the same assumptions as those for the conventional multiplier method. Moreover, we show that the method converges quadratically.

These results on the local convergence of two methods are obtained as simple implications of a very general fact concerning the abstract Newton's iterative scheme. This fact as applied to strongly regular generalized equations is presented in Section 2. The underlying scheme is of independent interest and has also other applications. At the same time, it is quite remarkable that this abstract scheme covers the multiplier methods, which are traditionally not regarded as Newton-type methods.

2. ABSTRACT NEWTONIAN SCHEME FOR GENERALIZED EQUATIONS

Following [20], we call the problem

$$\Phi(u) + N(u) \ni 0 \tag{2.1}$$

a generalized equation. Here, $\Phi : \mathbb{R}^\nu \longrightarrow \mathbb{R}^\nu$ is a single-valued mapping, while $N(\cdot)$ is a set-valued mapping that acts from $\mathbb{R}^\nu$ to the set of subsets of $\mathbb{R}^\nu$. If $N(\cdot) \equiv \{0\}$, then problem (2.1) is a conventional equation. On the other hand, (2.1) covers many other problem formulations, including KKT systems in mathematical programming (see Section 3).

Let Π be an arbitrary set. Consider the following abstract iterative scheme for generalized equation (2.1): based on the current approximation $u^k \in \mathbb{R}^\nu$ and the current parameter $\pi^k \in \Pi$, the method calculates the next approximation $u^{k+1} \in \mathbb{R}^\nu$ as a solution to the subproblem

$$\mathcal{A}(\pi^k, u^k, u) + N(u) \ni 0, \tag{2.2}$$

where the mapping $\mathcal{A} : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \longrightarrow \mathbb{R}^\nu$ is chosen so that, for every $\pi \in \Pi$ and every $\tilde{u} \in \mathbb{R}^\nu$, the mapping $\mathcal{A}(\pi, \tilde{u}, \cdot)$ approximates Φ in a certain sense in a neighborhood of $\tilde{u}$. The specific properties that we require of this approximation are indicated below in Theorem 1.

For every $\pi \in \Pi$ and $\tilde{u} \in \mathbb{R}^\nu$, define the set

$$U(\pi, \tilde{u}) = \{u \in \mathbb{R}^\nu \mid \mathcal{A}(\pi, \tilde{u}, u) + N(u) \ni 0\}. \tag{2.3}$$

Then $U(\pi^k, u^k)$ is the solution set for subproblem (2.2). Even if u^k is arbitrary close to the desired solution to Eq. (2.1), the set $U(\pi^k, u^k)$ can contain elements that are far from this solution. Therefore, for possible convergence, scheme (2.2) should be modified by complementing it with the so-called localization condition. This condition ensures that the solutions to subproblem (2.2) that are too far from u^k are not accepted as u^{k+1}:

$$u^{k+1} \in U(\pi^k, u^k) \cap B(u^k, \delta), \quad k = 0, 1, \dots . \tag{2.4}$$

Hereinafter, the symbol $B(u, \delta)$ denotes the closed ball of radius $\delta > 0$ centered at the point $u \in \mathbb{R}^\nu$.

The abstract iterative scheme (2.4) is an extension of the scheme developed in [21, Section 6C], which in turn goes back to [22]. The following theorem on the local convergence of scheme (2.4) is close to the result in [21, Exercise 6C.4].

Theorem 1. *Let $\bar{u} \in \mathbb{R}^\nu$ be a solution to the generalized equation* (2.1). *Assume that a mapping $\Phi : \mathbb{R}^\nu \longrightarrow \mathbb{R}^\nu$ is continuous in a neighborhood of $\bar{u}$ and $N(\cdot)$ is a set-valued mapping that acts from $\mathbb{R}^\nu$ to the set of subsets* *of* $\mathbb{R}^\nu$. *Let* Π *be a given set and* $\mathcal{A} : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \longrightarrow \mathbb{R}^\nu$ *be a given mapping. Assume that the following conditions are fulfilled:*

(1) (*strong metric regularity of the solution*) *There exists a scalar* $\rho > 0$ *such that, for every* $r \in \mathbb{R}^\nu$ *sufficiently close to* 0, *the perturbed generalized equation*

$$\Phi(u) + N(u) \ni r \tag{2.5}$$

has a unique solution $u(r)$ *near* $\bar{u}$. *Moreover, the mapping* $u(\cdot)$ *is locally Lipschitzian at the point* 0 *and has* ρ *as a Lipschitz constant.*

(2) (*accuracy of approximation*) *There exists a scalar* $\bar{\varepsilon} > 0$ *such that*

$$\mathcal{A}(\pi, \tilde{u}, \tilde{u}) = \Phi(\tilde{u}) \quad \forall \pi \in \Pi, \quad \forall \tilde{u} \in B(\bar{u}, \bar{\varepsilon}),$$

and there exists a function $\omega : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \times \mathbb{R}^\nu \longrightarrow \mathbb{R}_+$ *such that*

$$q := \rho \sup\{\omega(\pi, \tilde{u}, u^1, u^2) \,|\, \pi \in \Pi, \tilde{u}, u^1, u^2 \in B(\bar{u}, \bar{\varepsilon})\} < \frac{1}{2}, \tag{2.6}$$

$$\begin{gathered} \|(\Phi(u^1) - \mathcal{A}(\pi, \tilde{u}, u^1)) - (\Phi(u^2) - \mathcal{A}(\pi, \tilde{u}, u^2))\| \le \omega(\pi, \tilde{u}, u^1, u^2)\|u^1 - u^2\| \\ \forall \pi \in \Pi, \quad \forall \tilde{u}, u^1, u^2 \in B(\bar{u}, \bar{\varepsilon}). \end{gathered} \tag{2.7}$$

Then there exist scalars $\delta > 0$ *and* $\varepsilon_0 > 0$ *with the following properties: for every initial approximation* $u^0 \in B(\bar{u}, \varepsilon_0)$ *and every sequence* $\{\pi^k\} \subset \Pi$, *there exists a unique sequence* $\{u^k\} \subset \mathbb{R}^\nu$ *satisfying condition* (2.4). *This sequence converges to* $\bar{u}$, *and, for each* k, *we have*

$$\|u^{k+1} - \bar{u}\| \le \frac{\rho\omega(\pi^k, u^k, u^k, u^{k+1})}{1 - \rho\omega(\pi^k, u^k, u^k, u^{k+1})}\|u^k - \bar{u}\| \le \frac{q}{1-q}\|u^k - \bar{u}\|.$$

In particular, $\{u^k\}$ *converges at a linear rate. Moreover, the convergence rate is superlinear if* $\omega(\pi^k, u^k, u^k, u^{k+1}) \longrightarrow 0$ *as* $k \longrightarrow \infty$. *The convergence rate is quadratic if* $\omega(\pi^k, u^k, u^k, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

Proof. In view of (2.3), for any $\pi \in \Pi$ and any $\tilde{u} \in \mathbb{R}^\nu$, every point $u \in U(\pi, \tilde{u})$ satisfies the generalized equation (2.5) for

$$r = r(\pi, \tilde{u}, u) = \Phi(u) - \mathcal{A}(\pi, \tilde{u}, u).$$

Moreover, assumption (2) implies that the following relations hold for all $\tilde{u}$, u^1, $u^2 \in B(\bar{u}, \bar{\varepsilon})$:

$$\begin{gathered} r(\pi, \tilde{u}, \tilde{u}) = 0, \\ \|r(\pi, \tilde{u}, u^1) - r(\pi, \tilde{u}, u^2)\| \le \omega(\pi, \tilde{u}, u^1, u^2)\|u^1 - u^2\| \le \frac{1}{2\rho}\|u^1 - u^2\|. \end{gathered} \tag{2.8}$$

In particular, we have

$$\|r(\pi, \tilde{u}, u)\| \le \omega(\pi, \tilde{u}, \tilde{u}, u)\|u - \tilde{u}\| \le \frac{1}{2\rho}\|u - \tilde{u}\|, \tag{2.9}$$

if $\tilde{u}$, $u \in B(\bar{u}, \bar{\varepsilon})$. Using assumption (1) and reducing $\bar{\varepsilon}$ if necessary, we can ensure that the following inequalities hold for all $\pi \in \Pi$, $\tilde{u} \in B(\bar{u}, \bar{\varepsilon})$, and $u \in B(\bar{u}, \bar{\varepsilon}) \cap U(\pi, \tilde{u})$:

$$\|u - \bar{u}\| \le \rho\|r(\pi, \tilde{u}, u)\| \le \rho\omega(\pi, \tilde{u}, \tilde{u}, u)\|u - \tilde{u}\| \le \rho\omega(\pi, \tilde{u}, \tilde{u}, u)(\|u - \bar{u}\| + \|\tilde{u} - \bar{u}\|).$$

In view of (2.6), these inequalities imply that

$$\|u - \bar{u}\| \le \frac{\rho\omega(\pi, \tilde{u}, \tilde{u}, u)}{1 - \rho\omega(\pi, \tilde{u}, \tilde{u}, u)}\|\tilde{u} - \bar{u}\| \le \frac{q}{1-q}\|\tilde{u} - \bar{u}\|. \tag{2.10}$$

Note that, according to (2.8), the function $r(\pi, \tilde{u}, \cdot)$ is Lipschitzian in the ball $B(\bar{u}, \bar{\varepsilon})$ and has the Lipschitzian constant $1/(2\rho)$ for all $\pi \in \Pi$ and $\tilde{u} \in B(\bar{u}, \bar{\varepsilon})$. Moreover, by virtue of (2.9), we have $r(\pi, \tilde{u}, \bar{u}) \longrightarrow 0$ as $\tilde{u} \longrightarrow \bar{u}$ uniformly with respect to $\pi \in \Pi$. Then, applying Theorem 1.4 in [23], we can easily establish the existence of $\varepsilon \in (0, \bar{\varepsilon}/3]$ and $\tilde{\varepsilon} \in (0, \varepsilon]$ such that the set $U(\pi, \tilde{u}) \cap B(\bar{u}, \varepsilon)$ contains exactly one element for every $\pi \in \Pi$ and every $\tilde{u} \in B(\bar{u}, \tilde{\varepsilon})$.

Setting $\delta = \varepsilon + \tilde{\varepsilon}$, we see that the relations

$$\|u - \tilde{u}\| \le \|u - \bar{u}\| + \|\tilde{u} - \bar{u}\| \le \varepsilon + \tilde{\varepsilon} = \delta$$

hold for all $\tilde{u} \in B(\bar{u}, \tilde{\varepsilon})$ and $u \in B(\bar{u}, \varepsilon)$. Consequently,

$$U(\pi, \tilde{u}) \cap B(\tilde{u}, \delta) \neq \emptyset \quad \forall \pi \in \Pi, \quad \forall \tilde{u} \in B(\bar{u}, \tilde{\varepsilon}). \tag{2.11}$$

Moreover, for all $\tilde{u} \in B(\bar{u}, \tilde{\varepsilon})$ and $u \in B(\tilde{u}, \delta)$, we have

$$\|u - \bar{u}\| \le \|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\| \le \delta + \tilde{\varepsilon} = \varepsilon + 2\tilde{\varepsilon} \le \bar{\varepsilon}.$$

It follows that (2.10) is fulfilled for all $\pi \in \Pi$, $\tilde{u} \in B(\bar{u}, \tilde{\varepsilon})$, and $u \in U(\pi, \tilde{u}) \cap B(\tilde{u}, \delta)$. Moreover, (2.6) implies that $q/(1-q) < 1$. In particular, $u \in B(\bar{u}, \tilde{\varepsilon}) \subset B(\bar{u}, \varepsilon)$. Hence, the set in (2.11) contains exactly one element. If we set $\varepsilon_0 = \tilde{\varepsilon}$, then this fact, combined with (2.10), immediately yields the required assertion.

Theorem 1 shows that the superlinear convergence rate is attained if $\Phi(u^{k+1})$ is increasingly accurately approximated by $\mathcal{A}(\pi^k, u^k, u^{k+1})$ as k grows. This improved accuracy of approximation can be achieved by two different methods, namely, $\omega(\pi^k, u^k, u^k, u^{k+1})$ can be reduced in a natural way as u^k and u^{k+1} come closer to $\bar{u}$ or in an artificial way through controlling the parameters π^k.

For instance, if a mapping Φ is differentiable near $\bar{u}$ and its derivative is continuous at this point, then Φ can be approximated by its linearization $\mathcal{A}(\tilde{u}, u) = \Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u})$ without any parameters. In this case, the mean-value theorem implies that assumption (2) in Theorem 1 is fulfilled for $\omega(\tilde{u}, u^1, u^2) = \sup_{t \in [0,1]}\|\Phi'(tu^1 + (1-t)u^2) - \Phi'(\tilde{u})\|$. This value tends to zero in a natural way as $\tilde{u}$, u^1, and u^2 approach $\bar{u}$. Iterative scheme (2.4) with this choice of $\mathcal{A}$ corresponds to the well-known Josephy–Newton method, and Theorem 1 entails the results on the local superlinear convergence of this method obtained in [24, 25].

The application of Theorem 1 to the multiplier method discussed in Section 3 involves an adequate control of the penalty parameter. On the other hand, in the case of the linearly constrained Lagrangian method, the quality of approximation is attained in a natural way.

3. LOCAL CONVERGENCE OF THE MULTIPLIER METHOD

It is well known that the KKT system (1.2) for problem (1.1) can be restated in the form of generalized equation (2.1) if one sets $\nu = n + l + m$, defines the mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \longrightarrow \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ by the equality

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), -g(x)\right),$$

where $u = (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, and sets

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+. \tag{3.1}$$

Here,

$$N_Q(u) = \begin{cases} \{ v \in \mathbb{R}^\nu \,|\, \langle v, w - u\rangle \le 0 \ \forall w \in Q\}, & \text{if} \quad u \in Q, \\ \emptyset, & \text{if} \quad u \notin Q, \end{cases}$$

is the normal cone of the convex closed set $Q \subset \mathbb{R}^\nu$ at a point $u \in \mathbb{R}^\nu$.

In view of relations (1.3) and (1.4), an iteration step of the multiplier method can be written as

$$0 = \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \lambda^k, \mu^k) = f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^k + c_k h(x^{k+1})) + \sum_{i=1}^{m} \max\{0, c_k g_i(x^{k+1}) + \mu_i^k\} g_i'(x^{k+1})$$

$$= f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}}\lambda^{k+1} + (g'(x^{k+1}))^{\mathrm{T}}\mu^{k+1} = \frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}),$$

$$0 = h(x^{k+1}) - \frac{1}{c_k}(\lambda^{k+1} - \lambda^k),$$

$$0 = \max\{-\mu^{k+1}, c_k g(x^{k+1}) - (\mu^{k+1} - \mu^k)\} = -\min\{\mu^{k+1}, -c_k g(x^{k+1}) + (\mu^{k+1} - \mu^k)\}.$$

These relations allow us to conclude that the multiplier method is a particular case of scheme (2.2) in which $\mathcal{A} : (\mathbb{R}_+\backslash\{0\}) \times \mathbb{R}^{\nu} \times \mathbb{R}^{\nu} \longrightarrow \mathbb{R}^{\nu}$ is defined as

$$\mathcal{A}(c, \tilde{u}, u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x) - \frac{1}{c}(\lambda - \tilde{\lambda}), -g(x) + \frac{1}{c}(\mu - \tilde{\mu})\right),$$

where $\tilde{u} = (\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ and $u = (x, \lambda, \mu)$, and $N(\cdot)$ is defined by (3.1). It is obvious that $\mathcal{A}(c, \tilde{u}, \tilde{u}) = \Phi(\tilde{u})$ and

$$\Phi(u) - \mathcal{A}(c, \tilde{u}, u) = \left(0, \frac{1}{c}(\lambda - \tilde{\lambda}), -\frac{1}{c}(\mu - \tilde{\mu})\right).$$

It follows that

$$\|(\Phi(u^1) - \mathcal{A}(c, \tilde{u}, u^1)) - (\Phi(u^2) - \mathcal{A}(c, \tilde{u}, u^2))\| = \frac{1}{c}\|(\lambda^1 - \lambda^2, \mu^1 - \mu^2)\|,$$

which implies that condition (2.7) is fulfilled if $\omega(c, \tilde{u}, u^1, u^2) = 1/c$ and $u^1 = (x^1, \lambda^1, \mu^1)$, $u^2 = (x^2, \lambda^2, \mu^2)$ for any $c > 0$ and any $x^1, x^2 \in \mathbb{R}^n$, $\lambda^1, \lambda^2 \in \mathbb{R}^l$, and $\mu^1, \mu^2 \in \mathbb{R}^m$. Thus, the local convergence and the convergence rate of the multiplier method are immediately given by Theorem 1 if the solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ to the generalized equation corresponding to KKT system (1.2) is strongly metrically regular.

Suppose that a mapping Φ is locally Lipschitzian at the solution $\bar{u}$ to generalized equation (2.1). By the main result of [26], the strong metric regularity of this solution is implied by the property (called CD-regularity in [26]) that can be stated as follows: for every matrix $J \in \partial\Phi(\bar{u})$, the solution $\bar{u}$ to the generalized equation

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni 0$$

is strongly metrically regular or, which is the same, is strongly regular in the classical sense (see [20]). Here, ∂ denotes the Clark differential.

Assume that f, h, and g are locally Lipschitzian at $\bar{x}$. Then, according to [27, Proposition 2.3, Remark 2.1], the linear independence constraint qualification and the strong second-order sufficient condition (1.5) imply that $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ is a CD-regular solution to generalized equation (2.1) corresponding to KKT system (1.2).

Summing up what was said above, we arrive at the following result.

Theorem 2. *Assume that* $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$, *and* $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *are differentiable in a neighborhood of a point* $\bar{x} \in \mathbb{R}^n$ *and their derivatives are locally Lipschitzian at this point. Let* $\bar{x}$ *be a stationary point of problem* (1.1) *satisfying the linear independence constraint qualification. Assume that the unique Lagrange multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *corresponding to* $\bar{x}$ *satisfies the strong second-order sufficient condition* (1.5).

Then there exist scalars $\bar{c} > 0$ *and* $\delta > 0$ *with the following properties: for every initial approximation* $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *sufficiently close to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *and every sequence* $\{c_k\} \subset [\bar{c}, +\infty)$, *there exists a unique sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *such that, for* $k = 0, 1, \ldots, x^{k+1}$ *is a stationary point of problem* (1.3), *the pair* $(\lambda^{k+1}, \mu^{k+1})$ *satisfies relations* (1.4), *and*

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \le \delta. \tag{3.2}$$

This sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *and the convergence rate is linear. Moreover, the convergence rate is superlinear if* $c_k \longrightarrow +\infty$ *and quadratic if* $1/c_k = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|)$.

4. LOCAL CONVERGENCE OF THE LINEARLY CONSTRAINED LAGRANGIAN METHOD

In the analysis of the local convergence of the linearly constrained Lagrangian method, the natural assumption is that the penalty parameter in (1.7) is constant, that is, $c_k = c$ for all k (see the discussion of this issue in [17]; in the original paper [15], the choice was $c_k = 0$ for all k).

The KKT system for problem (1.7) has the form

$$f'(x) + (h'(x))^{\mathrm{T}}\lambda^k + (h'(x^k))^{\mathrm{T}}\lambda - \mu + c(h'(x))^{\mathrm{T}}h(x) = 0,$$
$$h(x^k) + h'(x^k)(x - x^k) = 0, \quad \mu \ge 0, \quad x \ge 0, \quad \langle \mu, x\rangle = 0.$$

This system and the rule defining λ^{k+1} imply that the linearly constrained Lagrangian method is a particular case of scheme (2.2) in which $\nu = n + l + n$, $\mathcal{A} : \mathbb{R}^\nu \times \mathbb{R}^\nu \longrightarrow \mathbb{R}^\nu$ is defined as

$$\mathcal{A}(\tilde{u}, u) = (f'(x) + (h'(x))^{\mathrm{T}}\lambda - \mu + c(h'(x))^{\mathrm{T}}h(x) - (h'(x) - h'(\tilde{x}))^{\mathrm{T}}(\lambda - \tilde{\lambda}), h(\tilde{x}) + h'(\tilde{x})(x - \tilde{x}), x),$$

and $N(\cdot)$ is defined by (3.1).

In order to apply Theorem 1 in this case, the mapping Φ should be redefined as follows:

$$\Phi(u) = (f'(x) + (h'(x))^{\mathrm{T}}\lambda - \mu + c(h'(x))^{\mathrm{T}}h(x), h(x), x). \tag{4.1}$$

For every fixed c, the KKT system for problem (1.6) remains equivalent to generalized equation (2.1) with $N(\cdot)$ defined by (3.1). This reflects the well-known fact that, in the optimality conditions, the conventional Lagrangian function can be replaced by the augmented one. On the other hand, this generalized equation is identical to the generalized equation corresponding to the KKT system for problem

$$f(x) + \frac{c}{2}\|h(x)\|_2^2 \longrightarrow \min, \quad h(x) = 0, \quad x \ge 0. \tag{4.2}$$

Suppose that the triple $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ is a solution to the KKT system for problem (1.6) satisfying the linear independence condition and the second-order sufficient condition (1.5). It is easy to see that the same properties hold with respect to problem (4.2). As was explained in the preceding section, this ensures the strong metric regularity of the solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ to generalized equation (2.1) in which Φ is defined by (4.1) and N is defined by (3.1).

Furthermore, it is easy to see that $\mathcal{A}(\tilde{u}, \tilde{u}) = \Phi(\tilde{u})$. By the mean-value theorem, we have

$$\|(\Phi(u^1) - \mathcal{A}(\tilde{u}, u^1)) - (\Phi(u^2) - \mathcal{A}(\tilde{u}, u^2))\| \le \omega(\tilde{u}, u^1, u^2)\|u^1 - u^2\|,$$

where

$$\omega(\tilde{u}, u^1, u^2) = O\Big(\sup_{t \in [0, 1]}\|h'(tx^1 + (1 - t)x^2) - h'(\tilde{x})\| + \|\lambda^2 - \tilde{\lambda}\|\Big). \tag{4.3}$$

To complete the verification of assumption (2) of Theorem 1, it remains to note that $\omega(\tilde{u}, u^1, u^2) \longrightarrow 0$ as $\tilde{u}, u^1, u^2 \longrightarrow \bar{u}$ because the mapping h' is continuous at the point $\bar{x}$.

Moreover, equality (4.3) and the fact that h' is locally Lipschitzian at $\bar{x}$ imply that

$$\omega(\tilde{u}, \tilde{u}, u) = O(\|u - \tilde{u}\|). \tag{4.4}$$

If the sequence $\{u^k\} \subset \mathbb{R}^\nu$ converges (super)linearly to $\bar{u}$, then $\|u^{k+1} - u^k\| = O(\|u^k - \bar{u}\|)$, and relation (4.4) entails the estimate $\omega(u^k, u^k, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

Theorem 3. *Assume that $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$ are differentiable in a neighborhood of a point $\bar{x} \in \mathbb{R}^n$ and their derivatives are locally Lipschitzian at this point. Let $\bar{x}$ be a stationary point of problem* (1.6) *satisfying the linear independence constraint qualification. Assume that the unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$ corresponding to $\bar{x}$ satisfies the strong second-order sufficient condition* (1.5).

Then, for every $c > 0$, there exists a scalar $\delta > 0$ with the following properties: for every initial approximation $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ sufficiently close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a unique sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ such that, for $k = 0, 1, \ldots, x^{k+1}$ is a stationary point of problem (1.7), *the pair $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is the corresponding Lagrange multiplier, and inequality* (3.2) *is fulfilled. This sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the convergence rate is quadratic.*

In closing, we emphasize that the superlinear convergence of the multiplier method is attained at the expense of the unlimited growth of the penalty parameter, which leads to the deterioration of the conditioning of subproblems (1.3) in this method. On the other hand, the linearly constrained Lagrangian method converges quadratically for a constant penalty parameter; however, its subproblems (1.7) are substantially more complex than unconstrained subproblems (1.3).

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 10-01-00251, 12-01-31025, and 12-01-33023.

REFERENCES

1. R. T. Rockafellar, "Computational Schemes for Solving Large-Scale Problems in Extended Linear-Quadratic Programming," Math. Program. **48**, 447–474 (1990).
2. R. T. Rockafellar and J. B. Wets, "Generalized Linear-Quadratic Problems of Deterministic and Stochastic Optimal Control in Discrete Time," SIAM J. Control Optim. **28**, 810–922 (1990).
3. L. Qi, "Superlinearly Convergent Approximate Newton Methods for LC1 Optimization Problems," Math. Program. **64**, 277–294 (1994).
4. D. Klatte and K. Tammer, "On the Second Order Sufficient Conditions to Perturbed $C^{1,1}$ Optimization Problems," Optimization **19**, 169–180 (1988).
5. L. Qi, *LC^1 Functions and LC^1 Optimization Problems*, Technical Report AMR 91/21 (School of Mathematics, University of New South Wales, Sydney, 1991).
6. A. F. Izmailov and M. V. Solodov, "The Theory of 2-Regularity for Mappings with Lipschitzian Derivatives and Its Applications to Optimality Conditions," Math. Operat. Res. **27**, 614–635 (2002).
7. O. Stein, "Lifting Mathematical Programs with Complementarity Constraints," Math. Program. **131** (1–2), 71–94 (2012).
8. A. F. Izmailov, A. L. Pogosyan, and M. V. Solodov, "Semismooth Newton Method for the Lifted Reformulation of Mathematical Programs with Complementarity Constraints," Comput. Optim. Appl. **51**, 199–221 (2012).
9. LANCELOT, http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml.
10. ALGENCAN, http://www.ime.usp.br/~egbirgin/tango/.
11. D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic, New York, 1982; Radio i Svyaz', Moscow, 1987).

12. A. F. Izmailov and V. M. Solodov, *Numerical Optimization Methods*, 2nd. ed. (Fizmatlit, Moscow, 2008) [in Russian].
13. D. Fernández and M. V. Solodov, "Local Convergence of Exact and Inexact Augmented Lagrangian Methods under the Second-Order Sufficient Optimality Condition," SIAM J. Optim. **22**, 384–407 (2012).
14. F. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983; Nauka, Moscow, 1988).
15. S. M. Robinson, "A Quadratically Convergent Algorithm for General Nonlinear Programming Problems," Math. Program. **3**, 145–156 (1972).
16. B. A. Murtagh and M. A. Saunders, "A Projected Lagrangian Algorithm and Its Implementation for Sparse Nonlinear Constraints," Math. Program. Study **16**, 84–117 (1982).
17. M. P. Friedlander and M. A. Saunders, "A Globally Convergent Linearly Constrained Lagrangian Method for Nonlinear Optimization," SIAM J. Optim. **15**, 863– 897 (2005).
18. B. A. Murtagh and M. A. Saunders, *MINOS 5.0 User's Guide*, Technical Report SOL 83.20 (Stanford Univ., 1983).
19. A. F. Izmailov and M. V. Solodov, "Inexact Josephy–Newton Framework for Generalized Equations and Its Applications to Local Analysis of Newtonian Methods for Constrained Optimization," Comput. Optim. Appl. **46**, 347–368 (2010).
20. S. M. Robinson, "Strongly Regular Generalized Equations," Math. Oper. Res. **5**, 43–62 (1980).
21. A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings* (Springer-Verlag, New York, 2009).
22. S. M. Robinson, "Newton's Method for a Class of Nonsmooth Functions," Set-Valued Anal. **2**, 291–305 (1994).
23. A. L. Dontchev and R. T. Rockafellar, "Newton's Method for Generalized Equations: A Sequential Implicit Function Theorem," Math. Program. **123**, 139–159 (2010).
24. N. H. Josephy, *Newton's Method for Generalized Equations*, Technical Summary Report, No. 1965 (Mathematics Research Center, Univ. of Wisconsin, Madison, 1979).
25. N. H. Josephy, *Quasi-Newton Methods for Generalized Equations*, Technical Summary Report, No. 1966 (Mathematics Research Center, Univ. of Wisconsin, Madison, 1979).
26. A. F. Izmailov, Preprint Ser. D 096/2012 (IMPA, Rio de Janeiro, 2012).
27. A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov, "The Josephy–Newton Method for Semismooth Generalized Equations and Semismooth SQP for Optimization," Set-Valued Var. Anal., 2012, DOI 10.1007/s11228-012-0218-z.