ISSN 0965-5425, Computational Mathematics and Mathematical Physics, 2012, Vol. 52, No. 11, pp. 1504–1519. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © A.F. Izmailov, E.I. Uskov, 2012, published in Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 2012, Vol. 52, No. 11, pp. 1959–1975.

On the Influence of the Critical Lagrange Multipliers on the Convergence Rate of the Multiplier Method

A. F. Izmailov and E. I. Uskov

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia
e-mail: izmaf@ccas.ru, ydoom@narod.ru

Received May 30, 2012

Abstract—The paper is devoted to the analysis of the influence of the critical Lagrange multipliers on the convergence rate of the multiplier method and the efficiency of various techniques for accelerating the final stage of this method.

DOI: 10.1134/S0965542512110073

Keywords: mathematical programming problem, augmented Lagrangian, method of multipliers, superlinear convergence, critical Lagrange multiplier, stabilized Newton–Lagrange method.

1. INTRODUCTION. CRITICAL LAGRANGE MULTIPLIERS

Consider the problem

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

where $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is a twice differentiable function, while $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$ and $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice differentiable mappings.

Recall that $\bar{x} \in \mathbb{R}^n$ is called a stationary point of problem (1.1) if there exist $\bar{\lambda} \in \mathbb{R}^l$ and $\bar{\mu} \in \mathbb{R}^m$ such that the triple $(\bar{x}, \bar{\lambda}, \bar{\mu})$ satisfies the Karush–Kuhn–Tucker (KKT) system for this problem; i.e.,

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

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

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

The pair $(\bar{\lambda}, \bar{\mu})$ is called a Lagrange multiplier corresponding to the stationary point $\bar{x}$. The set of such pairs is denoted by $\mathcal{M}(\bar{x})$.

Denote by $A(\bar{x}) = \{i = 1, \ldots, m | g_i(\bar{x}) = 0\}$ the set of indices corresponding to the inequality constraints of problem (1.1) that are active at the feasible point $\bar{x}$ of this problem. We also introduce the partition of $A(\bar{x})$ into the subsets

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

Furthermore, set $N(\bar{x}) = \{1, \ldots, m\} \backslash A(\bar{x})$.

Below, we use the following second-order sufficient optimality condition for a stationary point $\bar{x}$ of problem (1.1) and the corresponding Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \backslash \{0\}, \tag{1.3}$$

where

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0, g'_{A(\bar{x})}(\bar{x})\xi \le 0, \langle f'(\bar{x}), \xi \rangle \le 0\}$$
$$= \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0, g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0, g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \le 0\}$$

is the critical cone of problem (1.1) at $\bar{x}$. Hereinafter, the symbol z_I stands for the subvector of z with the components z_i ($i \in I$), where I is a given finite set. Recall that, if condition (1.3) is fulfilled, then $\bar{x}$ is a strict local solution to problem (1.1).

In this paper, the concept of a critical Lagrange multiplier, introduced in [1], is of key importance. A multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is said to be *critical* if there exists a triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that $\xi \neq 0$ and the following relations are fulfilled:

$$\begin{aligned} &\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0, \\ &\zeta_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}), \xi\rangle = 0, \quad i \in A_0(\bar{x}, \bar{\mu}), \quad \zeta_{N(\bar{x})} = 0. \end{aligned} \tag{1.4}$$

Otherwise, this multiplier is said to be *noncritical*. It is easy to verify that, if at a point $\bar{x}$, the corresponding multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfies the second-order sufficient condition (1.3), then $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier.

Suppose that there are no inequality constraints in problem (1.1). Then the first two equalities in (1.4) imply the following: $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is a critical multiplier, if, and only if, there exists an element $\xi \in \ker h'(\bar{x})\backslash\{0\}$ such that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \mathrm{im}(h'(\bar{x}))^{\mathrm{T}}. \tag{1.5}$$

For this type of problem, the concept of a critical multiplier was introduced in [2]. In the case of full degeneracy (that is, when $h'(\bar{x}) = 0$), the fact that $\bar{\lambda}$ is critical means the singularity of the Hessian matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})$. In the general case, since $\mathrm{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, condition (1.5) is equivalent to the condition

$$\pi_{\ker h'(\bar{x})}\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi = 0,$$

where the symbol $\pi_{\ker h'(\bar{x})}$ denotes the orthoprojector onto $\ker h'(\bar{x})$. The symmetric matrix of the linear operator $\xi \longrightarrow \pi_{\ker h'(\bar{x})}\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi : \ker h'(\bar{x}) \longrightarrow \ker h'(\bar{x})$ is called the *reduced Hessian* (with respect to x) of the Lagrangian function. If $\bar{\lambda}$ is critical, then this matrix is singular.

Relations (1.2) have an immediate implication: if the multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfies the condition $\bar{\mu}_{A(\bar{x})\backslash A} = 0$ for some index set $A \subset A(\bar{x})$, then $\bar{x}$ is a stationary point of the following problem with equality constraints:

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

The pair $(\bar{\lambda}, \bar{\mu}_A)$ is the Lagrange multiplier corresponding to this stationary point. The concept of a critical Lagrange multiplier is supplemented by the following concept introduced in [3]: a multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is said to be *critical with respect to the index set* $A \subset A(\bar{x})$ if $\bar{\mu}_{A(\bar{x})\backslash A} = 0$ and $(\bar{\lambda}, \bar{\mu}_A)$ is a critical Lagrange multiplier corresponding to the stationary point $\bar{x}$ of problem (1.6). Otherwise, $(\bar{\lambda}, \bar{\mu})$ is *noncritical with respect to A*.

It is easy to see from the above definitions that, if $(\bar{\lambda}, \bar{\mu})$ is a critical multiplier, then there exists a subset $I_1 \subset A_0(\bar{x}, \bar{\mu})$ such that this multiplier is critical with respect to the index set $A = A_+(\bar{x}, \bar{\mu}) \cup I_1$. In particular, if the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$ (that is, $A_0(\bar{x}, \bar{\mu}) = \varnothing$) is fulfilled, then the noncriticality is equivalent to the noncriticality with respect to $A(\bar{x})$.

Evidently, the second-order sufficient condition (1.3) implies the corresponding second-order sufficient condition at the stationary point $\bar{x}$ of problem (1.6) with $A = A(\bar{x})$ for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}_{A(\bar{x})})$. Therefore, $(\bar{\lambda}, \bar{\mu})$ cannot be a critical Lagrange multiplier with respect to $A(\bar{x})$.

On the other hand, this multiplier can be critical with respect to narrower index sets $A \subset A(\bar{x})$ (see [3, Example 3.5]).

The augmented Lagrangian method (or, simply, the multiplier method) is one of the traditional approaches to solving problem (1.1). It goes back to the papers [4, 5] and is still actively developed at the present time (see also [6, 7]). This method underlies such successful solvers as LANCELOT (see [8]) and ALGENCAN (see [9]). The multiplier method has a number of attractive properties (see [10]). One of them is a strong theory of global convergence; namely, the stationarity of the limit points of trajectories of this method can be justified under very weak constraint qualifications (see [11, 12]) or even when the traditional regularity is lacking (see [13]), in particular, when the set $\mathcal{M}(\bar{x})$ can be unbounded. This makes the multiplier methods a promising global strategy for optimization problems with irregular constraints.

Another attractive feature of the multiplier method is its sound locally linear and even superlinear convergence without any regularity conditions. Namely, the following result was obtained in [14] (see also Theorem 1 below). Suppose that the initial approximation is chosen near the point $(\bar{x}, \bar{\lambda}, \bar{\mu})$ for which $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfies the second-order sufficient optimality condition (1.3). Then any method of this type generates a trajectory $\{(x^k, \lambda^k, \mu^k)\}$ that converges to a point $(\bar{x}, \lambda^*, \mu^*)$ for some $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$ close to $(\bar{\lambda}, \bar{\mu})$. Moreover, the convergence rate is linear, and even superlinear if the penalty parameter tends to infinity. This is consistent with the results on the convergence rate of the multiplier method under more traditional (and considerably more stringent) assumptions (see [6, Theorem 2.7; 7, Theorem 4.7.4]). In particular, the multiplier methods have the property of dual stabilization, which means that the dual trajectories of these methods usually converge even if the set of Lagrange multipliers is unbounded. This indicates that, on the whole, the multiplier methods can be considered a promising approach to the numerical solution of potentially irregular optimization problems, namely, an approach that combines global convergence with a high convergence rate.

However, on a closer inspection, the situation is not so unambiguously reassuring. It is known (see [2, 3, 15, 16]) that the following phenomenon occurs when Newton-type methods are applied to optimization problems with irregular constraints: dual trajectories are attracted by Lagrange multipliers that are critical with respect to some index set $A \subset A(\bar{x})$. Such multipliers and, in particular, those violating the second-order sufficient condition are often attractors for dual trajectories. It is this phenomenon that is usually responsible for the loss of the superlinear convergence of these methods. As shown in [17], an attraction to critical Lagrange multipliers is also observed for the multiplier methods. In this paper, we examine how typical is the indicated phenomenon for methods of this class and what is its influence on their convergence rate.

Another well-known problem concerning the multiplier methods is that the growth of the penalty parameter, which is required for attaining the superlinear convergence, leads to the deterioration of the conditioning for subproblems in these methods. This increases the computational costs of their solution. A higher convergence rate may not compensate these costs, and a method converging superlinearly may be inferior in terms of total efficiency to a method with a bounded sequence of the penalty parameters.

As a possible resolution of this problem, one can switch at the final stage of the method (when there are reasons to assume that the current approximation is sufficiently close to the solution) to some local fast algorithm. This approach to accelerating the multiplier methods was proposed in [10], and it is implemented in the current version of ALGENCAN: when the residual of the KKT system (1.2) at the current point is sufficiently small, the set of inequality constraints that are active at the desired solution is estimated, the corresponding problem with equality constraints is formed, and Newton's method is applied to its Lagrange system. However, as noted above, the problem is that, in the case of irregular constraints, the superlinear convergence of this Newton–Lagrange method is usually lost anyway because its dual trajectories are attracted by critical Lagrange multipliers.

A natural alternative is that, instead of the Newton–Lagrange method, its stabilized version, first proposed in [18] for problems with inequality constraints, is used as an accelerator. A survey of subsequent publications devoted to this method was given in [19], while, in [1], the following result on the local convergence of the stabilized Newton–Lagrange method for problems with equality constraints is obtained. Let $\bar{\lambda}$ be a noncritical Lagrange multiplier corresponding to a stationary point $\bar{x}$. Suppose that the initial approximation is close to the point $(\bar{x}, \bar{\lambda})$. Then the method generates a trajectory $\{(x^k, \lambda^k)\}$ that superlinearly converges to the point $(\bar{x}, \lambda^*)$ for some $\lambda^* \in \mathcal{M}(\bar{x})$ close to $\bar{\lambda}$. In particular, for this method, an

attraction to critical Lagrange multipliers is not so persistent. As noted in [16], there exist extensive sets of dual points with the property that such an attraction does not occur if these points are used as initial approximations. The question, certainly, is whether these attractive properties of the stabilized Newton–Lagrange method are sufficient for using it as an accelerator in the multiplier methods: if the dual trajectories of the latter method are attracted by critical multipliers, then the points on these trajectories may never fall into the region of initial approximations that are suitable for a successful application of this accelerator. Thus, the role of critical Lagrange multipliers is crucial here as well. The comparative efficiency of two indicated accelerators is also a subject of this paper.

2. THE MULTIPLIER METHOD AND ITS ACCELERATORS

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 maximum is taken componentwise. The multiplier method consists in that unconstrained minimization problems are successively solved for the function $L_c(\cdot, \lambda, \mu)$ for fixed values of λ and μ, which are iteratively updated. Note that, if there are simple constraints (for instance, lower and upper bounds for the components of x) among the constraints of problem (1.1), then they are often not incorporated into the augmented Lagrangians but are treated directly. In this case, a subproblem of the multiplier method consists in minimizing the function $L_c(\cdot, \lambda, \mu)$ subject to simple direct constraints (see [11]). In particular, exactly this approach is implemented in ALGENCAN (see [9]). To simplify our presentation, we assume here that there are no direct constraints. Up to this simplification, the algorithm presented below is identical to that underlying ALGENCAN (see [11, Algorithm 3.1]; see also some recent modifications and improvements proposed in [12, 20]).

Algorithm 1

Choose the scalar parameters $\bar{\lambda}_{\min}$, $\bar{\lambda}_{\max}$, $\bar{\lambda}_{\min} \le \bar{\lambda}_{\max}$, $\bar{\mu}_{\max} \ge 0$, $c_0 > 0$, $\theta \in [0, 1)$, and $\rho > 1$. Set $k = 0$.

Step 1. Choose a vector $(\bar{\lambda}^k, \bar{\mu}^k) \in \mathbb{R}^l \times \mathbb{R}^m$ such that $\bar{\lambda}_{\min} \le \bar{\lambda}_j^k \le \bar{\lambda}_{\max}$ ($j = 1, ..., l$) and $0 \le \bar{\mu}_j^k \le \bar{\mu}_{\max}$ ($j = 1, ..., m$). (For $k \ge 1$, it is natural to choose $(\bar{\lambda}^k, \bar{\mu}^k)$ as the projection of the vector (λ^k, μ^k) found at Step 2 of this algorithm on the set $\otimes_{j=1}^{l}[\bar{\lambda}_{\min}, \bar{\lambda}_{\max}] \times \otimes_{j=1}^{m}[0, \bar{\mu}_{\max}]$.)

Choose $\varepsilon_k \ge 0$ and calculate $x^{k+1} \in \mathbb{R}^n$ as an approximate solution to the problem

$$L_{c_k}(x, \bar{\lambda}^k, \bar{\mu}^k) \longrightarrow \min, \quad x \in \mathbb{R}^n. \tag{2.1}$$

Namely, x^{k+1} is found as a point satisfying the condition

$$\left\|\frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \bar{\lambda}^k, \bar{\mu}^k)\right\|_\infty \le \varepsilon_k.$$

Step 2. Set

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

Step 3. If $k = 0$ or the condition

$$\|(h(x^{k+1}), \min\{\mu^{k+1}, -g(x^{k+1})\})\|_\infty \le \|(h(x^k), \min\{\mu^k, -g(x^k)\})\|_\infty \tag{2.2}$$

is fulfilled, choose an arbitrary $c_{k+1} \ge c_k$. Otherwise, choose c_{k+1} so that $c_{k+1} \ge \rho c_k$. Increase by one the index k of the current step and go to Step 1.

ALGENCAN uses the following values of the parameters: $\bar{\lambda}_{\min} = -10^{20}$, $\bar{\lambda}_{\max} = \bar{\mu}_{\max} = 10^{20}$, $\theta = 0.5$, and $\rho = 10$. The initial value of the penalty parameter c_0 is determined using a special initialization procedure, and it depends on the values of f, h, and g at the current point (see [21]). The subsequent values of the penalty parameter are chosen at Step 3 of Algorithm 1 as follows: $c_{k+1} = c_k$ if condition (2.2) is fulfilled and $c_{k+1} = \rho c_k$ otherwise. However, we note that both Algorithm 1 and the theory of its global convergence developed in [12] allow that the penalty parameter is arbitrarily increased if condition (2.2) is fulfilled and is increased to a larger extent than above if (2.2) is violated.

We also note that, in addition to the procedure described in Algorithm 1, ALGENCAN uses a number of heuristics aimed to increase its efficiency (see [21]). In particular, restarts are used for the penalty parameter; that is, if the current approximation is "nearly feasible" but the residual of the other constraints in the KKT system (1.2) is not sufficiently small, then the next value of the penalty parameter is again calculated with the help of the initialization procedure. Furthermore, a procedure for scaling the problem is incorporated into ALGENCAN: instead of the original problem, Algorithm 1 is applied to its modified variant

$$\tilde{f}(x) \longrightarrow \min, \quad \tilde{h}(x) = 0, \quad \tilde{g}(x) \le 0, \tag{2.3}$$

where $\tilde{f}(x) = f(x)/\max\{1, \|f'(x^0)\|_\infty\}$, $\tilde{h}_i(x) = h_i(x)/\max\{1, \|h_i'(x_0)\|_\infty\}$ $(i \in 1, \ldots, l)$, $\tilde{g}_i(x) = g_i(x)/\max\{1, \|g_i'(x^0)\|_\infty\}$ $(i \in 1, \ldots, m)$, and $x^0 \in \mathbb{R}^n$ is the initial approximation. Naturally, the transition to problem (2.3) changes both the primal and dual trajectories of the method. On the other hand, $\bar{x}$ is a stationary point of the original problem (1.1) if and only if it is a stationary point of the modified problem (2.3), even though the corresponding Lagrange multipliers can change.

The local convergence and the convergence rate of Algorithm 1 are specified in the following theorem proved in [14]. We define the residual function $\sigma: \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} \longrightarrow \mathbb{R}_+$ of the KKT system (1.2) by the formula

$$\sigma(x, \lambda, \mu) = \left\| \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), \min\{\mu, -g(x)\} \right) \right\|. \tag{2.4}$$

Theorem 1. *Assume that the function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ and the mappings $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$ and $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice differentiable in a neighborhood of a point $\bar{x} \in \mathbb{R}^n$ and their second derivatives are continuous at this point. Assume also that $\bar{x}$ is a stationary point of problem* (1.1) *satisfying the second-order sufficient condition* (1.3) *for some $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$. Let $\varepsilon: \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be an arbitrary function such that $\varepsilon(t) = o(t)$.*

Then, for every $M > 0$, there exists $\bar{c} > 0$ such that, for any $c_0 \ge \bar{c}$ and any initial approximation $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$ that is sufficiently close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, Algorithm 1 with $\varepsilon_k = \varepsilon(\sigma(x^k, \lambda^k, \mu^k))$ generates a trajectory $\{(x^k, \lambda^k, \mu^k)\}$ such that the condition

$$\left\| (x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k) \right\| \le M\sigma(x^k, \lambda^k, \mu^k)$$

is fulfilled for each $k = 0, 1, \ldots$. Every such trajectory converges to a point $(\bar{x}, \lambda^*, \mu^*)$ for some $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$. Moreover, the convergence is linear and even superlinear if $c_k \longrightarrow +\infty$.

Now, we discuss the construction of an accelerator for Algorithm 1 that was proposed in [10] and is, up to slight modifications, implemented in ALGENCAN. For the prescribed accuracy (stopping criterion) $\varepsilon > 0$, the accelerator is run when Algorithm 1 attains the given accuracy level $\hat{\varepsilon} > \varepsilon$, namely, when the current approximation (x^k, λ^k, μ^k) satisfies the condition

$$\sigma(x^k, \lambda^k, \mu^k) \le \hat{\varepsilon} \tag{2.5}$$

(ALGENCAN uses (2.4) with the infinity norm), but the stopping criterion is not yet fulfilled; that is, (2.5) is false if $\hat{\varepsilon}$ is replaced by ε. Otherwise, the algorithm terminates because an approximation (x^k, λ^k, μ^k) of the required quality is already found by the multiplier method.

If the run of the accelerator fails, then all the approximations produced by it are discarded, the value of $\hat{\varepsilon}$ is reduced (still remaining larger than ε), and the execution of Algorithm 1 is resumed from the point where it was interrupted. This continues until the new accuracy level $\hat{\varepsilon}$ is attained; then the accelerator is run again, and so on. A run of the accelerator is regarded as a failure if its K iteration steps, where K is a given integer, do not yield an approximation $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that

$$\sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \le \varepsilon. \tag{2.6}$$

The run of the accelerator from the point $(\hat{x}, \hat{\lambda}, \hat{\mu}) = (x^k, \lambda^k, \mu^k)$ is realized as follows. First, the operating index set $\hat{A} \subset \{1, \ldots, m\}$ is formed. It estimates the set of indices corresponding to the inequality constraints that are active at the desired solution:

$$\hat{A} = \{i = 1, \ldots, m \,|\, g_i(\hat{x}) \ge -\hat{\varepsilon}\}. \tag{2.7}$$

Inequality constraints with the indices $i \in \{1, ..., m\}\backslash\hat{A}$ are discarded, while the remaining inequality constraints are replaced by the equalities

$$g_i(x) + \frac{1}{2}s_i^2 = 0, \quad i \in \hat{A},$$

where s_i ($i \in \hat{A}$) are auxiliary variables. According to [10], the introduction of auxiliary variables makes the accelerator considerably more robust, compensating the negative effect caused by the possible inclusion in $\hat{A}$ the indices of constraints that are not active at the solution.

Without loss of generality, we thereafter assume that $\hat{A} = \{1, ..., \hat{m}\}$ for some $\hat{m} \in \{0, 1, ..., m\}$. Define the mapping $\hat{h} : \mathbb{R}^n \times \mathbb{R}^{\hat{m}} \longrightarrow \mathbb{R}^l \times \mathbb{R}^{\hat{m}}$ by the formula

$$\hat{h}(x, s) = \left(h(x), g_1(x) + \frac{1}{2}s_1^2, \ldots, g_{\hat{m}}(x) + \frac{1}{2}s_{\hat{m}}^2\right). \tag{2.8}$$

Then we solve the problem

$$f(x) \longrightarrow \min, \quad \hat{h}(x, s) = 0, \tag{2.9}$$

applying the Newton–Lagrange method (NLM), which is the Newton method as applied to the Lagrange system

$$\frac{\partial \hat{L}}{\partial(x, s)}((x, s), (\lambda, \nu)) = 0, \quad \hat{h}(x, s) = 0 \tag{2.10}$$

for problem (2.9). Here, $\hat{L} : (\mathbb{R}^n \times \mathbb{R}^{\hat{m}}) \times (\mathbb{R}^l \times \mathbb{R}^{\hat{m}}) \longrightarrow \mathbb{R}$ is the Lagrangian of problem (2.9):

$$\hat{L}((x, s), (\lambda, \nu)) = f(x) + \langle(\lambda, \nu), \hat{h}(x, s)\rangle. \tag{2.11}$$

Let $((x^k, s^k), (\lambda^k, \nu^k)) \in (\mathbb{R}^n \times \mathbb{R}^{\hat{m}}) \times (\mathbb{R}^l \times \mathbb{R}^{\hat{m}})$ be the current approximation to a solution to system (2.10). Then, according to NLM, the next approximation $((x^{k+1}, s^{k+1}), (\lambda^{k+1}, \nu^{k+1}))$ is found as the solution to the linear system

$$\begin{pmatrix} \dfrac{\partial^2 \hat{L}}{\partial(x, s)^2}((x^k, s^k), (\lambda^k, \nu^k)) & (\hat{h}'(x^k, s^k))^{\mathrm{T}} \\ \hat{h}'(x^k, s^k) & 0 \end{pmatrix} \begin{pmatrix} (x, s) - (x^k, s^k) \\ (\lambda, \nu) - (\lambda^k, \nu^k) \end{pmatrix}$$
$$= -\begin{pmatrix} \dfrac{\partial \hat{L}}{\partial(x, s)}((x^k, s^k), (\lambda^k, \nu^k)) \\ \hat{h}(x^k, s^k) \end{pmatrix}. \tag{2.12}$$

As an initial approximation, NLM uses the point $((\hat{x}, s^0), (\hat{\lambda}, \nu^0))$, where $s_i^0 = \sqrt{2\max\{0, -g_i(\hat{x})\}}$ and $\nu_i^0 = \hat{\mu}_i$ for $i \in \hat{A}$. For each $k = 1, ..., K$, NLM verifies the condition

$$\hat{\sigma}((x^k, s^k), (\lambda^k, \nu^k)) \le \varepsilon, \tag{2.13}$$

where $\hat{\sigma} : (\mathbb{R}^n \times \mathbb{R}^{\hat{m}}) \times (\mathbb{R}^l \times \mathbb{R}^{\hat{m}}) \longrightarrow \mathbb{R}_+$ is the residual of the Lagrange system (2.10). It is defined by the formula

$$\hat{\sigma}((x, s), (\lambda, \nu)) = \left\|\left(\frac{\partial \hat{L}}{\partial(x, s)}((x, s), (\lambda, \nu)), \hat{h}(x, s)\right)\right\|. \tag{2.14}$$

If condition (2.13) is fulfilled for a certain k, then condition (2.6) is verified for $\tilde{x} = x^k$, $\tilde{\lambda} = \lambda^k$, $\tilde{\mu}_i = \max\{0, \nu_i^k\}$ for $i \in \hat{A}$, and $\tilde{\mu}_i = 0$ for $i \in \{1, ..., m\}\backslash\hat{A}$. If this condition is fulfilled, then the algorithm terminates because the approximation $(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ has the required quality. Otherwise, the next approximation is found from (2.12) for $k < K$. If $k = K$, then the current run of the accelerator is declared as a failure.

By default, ALGENCAN uses the values $K = 10$, $\varepsilon = 10^{-8}$, and the initial accuracy level $\hat{\varepsilon} = 10^{-4}$. A possible rule for reducing $\hat{\varepsilon}$ is given in [10]. Moreover, the accelerator is applied in ALGENCAN to the original problem even if Algorithm 1 is run for the modified variant (2.3). In the latter case, the dual initial approximation $(\hat{\lambda}, \hat{\mu})$ is modified for the accelerator in an appropriate way.

Having in mind the possible improvement of the behavior of the above scheme for problems with irregular constraints, we also consider the variant of the accelerator in which NLM as applied to problem (2.9) is replaced by the stabilized Newton–Lagrange method (SNLM); that is, instead of (2.12), the linear system

$$\begin{pmatrix} \dfrac{\partial^2 \hat{L}}{\partial (x,s)^2}((x^k, s^k), (\lambda^k, \nu^k)) & (\hat{h}'(x^k, s^k))^{\mathrm{T}} \\ \hat{h}'(x^k, s^k) & -\sigma_k I \end{pmatrix} \begin{pmatrix} (x,s) - (x^k, s^k) \\ (\lambda, \nu) - (\lambda^k, \nu^k) \end{pmatrix} = -\begin{pmatrix} \dfrac{\partial \hat{L}}{\partial (x,s)}((x^k, s^k), (\lambda^k, \nu^k)) \\ \hat{h}(x^k, s^k) \end{pmatrix} \tag{2.15}$$

is solved. Here, $\sigma_k \geq 0$ is the stabilization parameter. The choice of a value for this parameter is based on the estimate for the distance from $((x^k, s^k), (\lambda^k, \nu^k))$ to the solution set of the Lagrange system (2.10) for problem (2.9). Namely, we assume that $\sigma_k = \hat{\sigma}((x^k, s^k), (\lambda^k, \nu^k))$.

From the results of [1], we can derive the following theorem about SNLM as applied to problem (2.9). This theorem shows that, if SNLM is run as an accelerator of Algorithm 1 near a stationary point and the dual approximation is sufficiently close to a noncritical Lagrange multiplier that satisfies the strict complementarity condition, then the accelerator has the superlinear convergence.

Theorem 2. *Assume that the function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ and the mappings $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$ and $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice differentiable in a neighborhood of a point $\bar{x} \in \mathbb{R}^n$ and their second derivatives are continuous at this point. Assume also that $\bar{x}$ is a stationary point of problem* (1.1) *and $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is a noncritical Lagrange multiplier satisfying the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$.*

Take an arbitrary $\hat{\varepsilon} > 0$ and any initial approximation $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ sufficiently close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Define $\hat{A}$ as in (2.7) *with $\hat{x} = x^0$. Set $s_i^0 = \sqrt{2\max\{0, -g_i(x^0)\}}$ and $\nu_i^0 = \mu_i^0$ for $i \in \hat{A}$. Then there exists a unique trajectory $\{((x^k, s^k), (\lambda^k, \nu^k))\}$ such that, for each $k = 0, 1, \ldots$, the point $\{((x^{k+1}, s^{k+1}), (\lambda^{k+1}, \nu^{k+1}))\}$ satisfies system* (2.15) *with $\hat{h}$ defined as in* (2.8), *$\hat{L}$ defined as in* (2.11), *and $\sigma_k = \hat{\sigma}((x^k, s^k), (\lambda^k, \nu^k))$ defined as in* (2.14). *This trajectory converges superlinearly to the point $((\bar{x}, \bar{s}), (\lambda^*, \nu^*))$, where $\bar{s}_i = \sqrt{-2g_i(\bar{x})}$, $\nu_i^* = \mu_i^*$ $(i \in \hat{A})$ for some point $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$ such that $\mu_i^* = 0$ for $i \in \{1, \ldots, m\} \backslash \hat{A}$.*

Proof. A sufficient proximity of x^0 and $\bar{x}$ ensures the inclusion $A(\bar{x}) \subset \hat{A}$. It follows that $\bar{x}$ is a stationary point of the problem

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad g_{\hat{A}}(x) \leq 0, \tag{2.16}$$

while $(\bar{\lambda}, \bar{\mu}_{\hat{A}})$ is the corresponding noncritical Lagrange multiplier satisfying the strict complementarity condition. By applying Theorem 2 in [1] to problem (2.16), we obtain the required assertion.

Note that, if the strict complementarity condition is violated, then one cannot expect that SNLM as applied to problem (2.9) converges superlinearly. Indeed, in this case, Proposition 5 in [1] is applicable to problem (2.16), which implies that, in problem (2.9), the multiplier $(\bar{\lambda}, \bar{\mu}_{\hat{A}})$, corresponding to the stationary point $(\bar{x}, \bar{s})$, is critical.

The comparative behavior of NLM and SNLM in the presence of a critical Lagrange multiplier is clearly demonstrated by the following simple example.

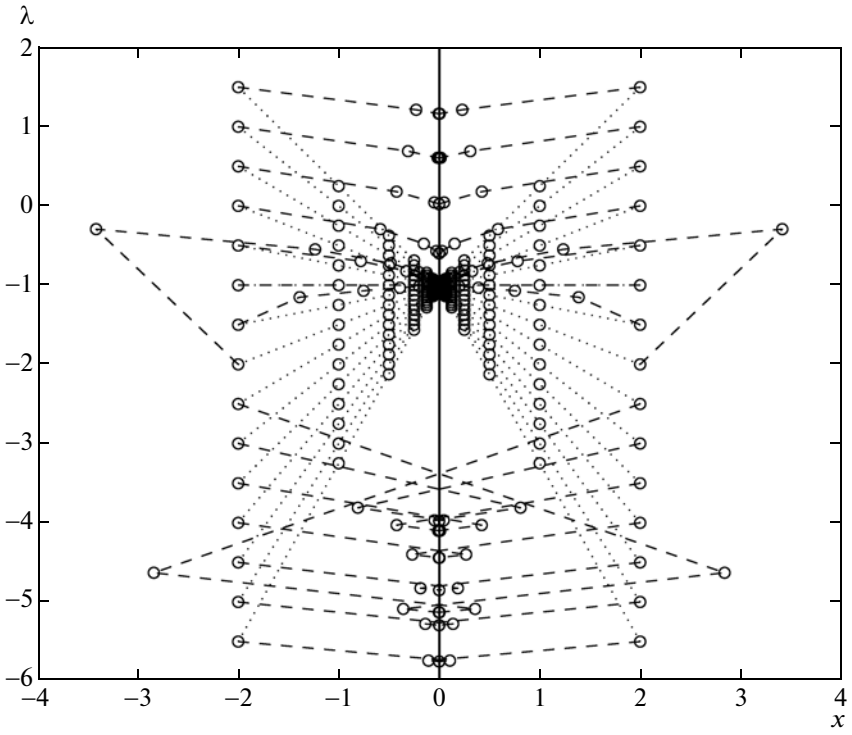

Fig. 1.

Example 1 (Test 20101 from the DEGEN collection (see [22])). Let $n = l = 1$, $m = 0$, $f(x) = x^2$, and $h(x) = x^2$. The only feasible point (and, hence, the only solution) of problem (1.1) with these data is $\bar{x} = 0$. Moreover, $\mathcal{M}(\bar{x}) = \mathbb{R}$, while the only critical multiplier is $\bar{\lambda} = -1$.

In Fig. 1, the vertical solid line corresponds to the set $\{\bar{x}\} \times \mathcal{M}(\bar{x})$. The dotted lines show the trajectories of NLM, while the dashed lines show the trajectories of SNLM. The runs were performed from the points (x^0, λ^0), where x^0 was either -2 or 2, while λ^0 took various values in the interval $[-5.5, 0.5]$. The trajectories of NLM always converged to $(\bar{x}, \bar{\lambda})$ with the linear rate. For SNLM, the behavior of the dual trajectory depends on the dual initial point. If $\lambda^0 \in (-2, -0.4)$ (the interval here is indicated approximately), then the convergence is to $(\bar{x}, \bar{\lambda})$ and the rate is linear. For other values of λ^0 that were used, we observed the superlinear convergence to $(\bar{x}, \lambda^*)$ for some $\lambda^* \neq \bar{\lambda}$.

In Fig. 2, the region of initial approximations (x^0, λ^0) from which SNLM converges to the critical Lagrange multiplier is shown in grey.

3. NUMERICAL RESULTS

In this section, we present the results of numerical tests in which we investigated the influence of critical multipliers on, first, the convergence rate of ALGENCAN with different control procedures for the penalty parameter and, second, the behavior of two accelerators of this solver that were described in the preceding section (that is, NLM and SNLM).

For these tests, we used version 2.3.7 of the ALGENCAN package with the AMPL interface. The package was compiled employing the MA57 library (see [23]). This library, designed for the efficient solution of sparse systems of linear equations, makes it possible to significantly increase the overall efficiency of the solver for large-scale problems.

We used the test collections DEGEN (see [22]) and MacMPEC (see [24]) (both collections exist as AMPL models).

The DEGEN collection consists of 109 small mathematical programming problems with irregular constraints. We used 98 of these problems for our tests. The other 11 problems were dropped for the fol-

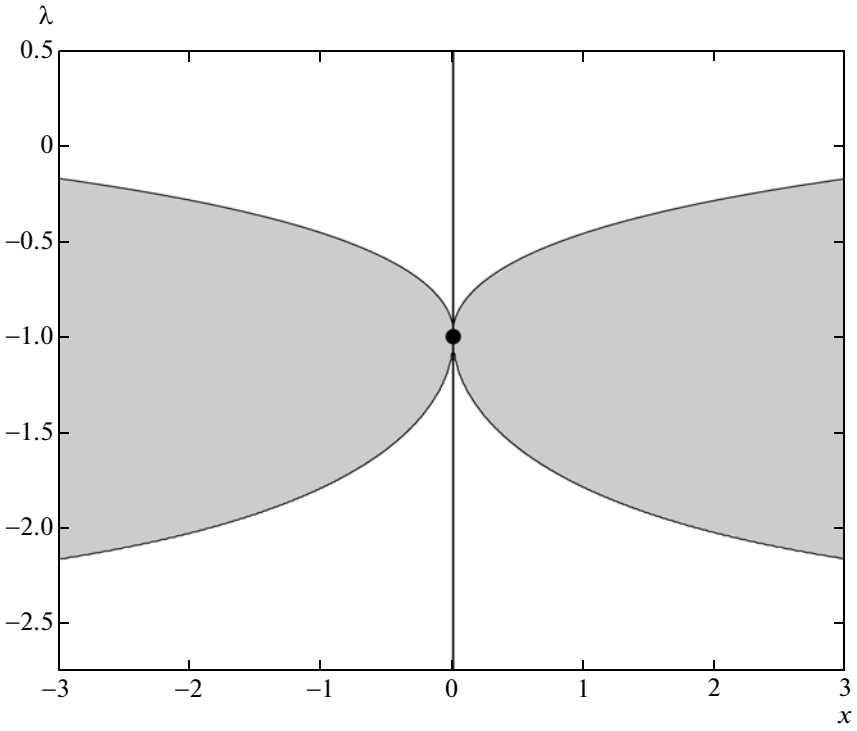

Fig. 2.

lowing reasons: the objective functions in problems 20205, 20223, 30204, and 40207 are unbounded on their feasible sets (although they have stationary points at which the constraints are irregular). Problems 2DD01-50h, 2DD01-50v, 2DD01-500h, and 2DD01-500v are too large compared to the other problems. Problems 20201, 40210, and 40211 have only simple constraints in consequence of which no outer iteration is performed in ALGENCAN for these problems. Note that, among the remaining 98 problems, 52 problems have Lagrange multipliers that are critical with respect to some index sets $A \subset A(\bar{x})$. Another 33 problems have no critical multipliers. For 13 problems, there are no Lagrange multipliers corresponding to the solutions, which is entirely possible in the case of irregular constraints.

MacMPEC is a collection of 180 optimization problems with complementarity constraints. These are problems of the form

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad G(x) \ge 0, \quad H(x) \ge 0, \quad \langle G(x), H(x) \rangle \le 0, \tag{3.1}$$

where the complementarity constraints are stated with the help of the mappings G, $H : \mathbb{R}^n \longrightarrow \mathbb{R}^s$. Problem (3.1) is of course a particular case of the general mathematical programming problem (1.1). On the other hand, it is well known (e.g., see [25, Section 4.3]) that the presence of complementarity constraints inevitably leads to irregularity. MacMPEC contains both small academic examples and problems of practical origin, including large-scale ones.

We used 161 problems from MacMPEC for our tests. Thirteen problems were discarded because they have the so-called mixed complementarity constraints. One problem contains a binary variable. The other five discarded problems have no feasible points.

3.1. Experiment with the Rules for the Penalty Parameter: Critical Multipliers and Convergence Rate

We begin by discussing the experiment that aimed to compare the original ALGENCAN as given by Algorithm 1 and the accompanying comments (in what follows, this basic variant is denoted by M1) and its modification (thereafter denoted by M2) in which the penalty parameter, which was set to one at the first iteration step, was then doubled at each subsequent step. Thus, by Theorem 1, we can expect that M2 converges superlinearly.

For both M1 and M2, we used the following settings. First, the accelerators were not used because the aim of this experiment was to examine the behavior of the "pure" multiplier method. Second, at the first

Table 1. Results for problems with critical multipliers from the DEGEN collection

Problem	SL	CM	Problem	SL	CM
10201	●	●	20219	●	●
10203	●	●	20226	●	●
10206	●	●	20227	●	●
10207	●	●	20302	●	●
10301		●	20303	●	
20101	●		20304	●	
20103	●	●	20307		●
20105	○	○	20309		●
20107	●		2DD01-5v	●	
20109	●	●	30210	●	
20202	●	●	30211	●	
20203		●	30301	●	
20204	●		30401	●	●
20207	●	●	40201	●	
20209	○	○	40202	●	
20210	●	●	40205	●	●
20211	●		40206	●	●
20213	○	●	40208	●	●
20214		●	40401	●	
20216	●	●	40402	●	●
20217	○	●			

outer iteration, the bound for the number of inner iteration steps (that is, steps of the method used for solving subproblems (2.1)) was increased to 100 (while the default value is 10). Moreover, in the tests based on the DEGEN collection, the scaling was switched off because it changes the set of Lagrange multipliers corresponding to the solution.

For each problem in the DEGEN collection, both M1 and M2 were run ten times from random initial points. For each run, we analyzed the rate at which the distance to the primal solution decays and the attraction of the dual trajectory to critical multipliers. (The DEGEN collection provides the solutions and critical multipliers for its models.)

For the tests based on the DEGEN collection, the results produced by M1 and M2 were virtually identical. This is related to the fact that, for many DEGEN problems, Algorithm 1, in which subproblems are solved exactly (that is, $\varepsilon_k = 0$ for all k), finds the exact solution in a finite number of steps. For such problems, the behavior of the algorithm is to a considerable extent determined by the accuracy of solving subproblems and is almost independent of the rule for changing the penalty parameter. Consequently, below, we present only the results produced by M2.

Among 98 problems used in our experiment, the convergence to solutions violating constraint qualifications was observed for 80 problems. Below, we discuss the results obtained exactly for these problems. For the other problems, either the runs failed or the convergence to nondegenerate solutions occurred.

For all the problems that have no Lagrange multipliers corresponding to the solution (there are 12 such problems of 80), the primal trajectory converged linearly, while the norm of dual approximations grew to infinity. For almost all problems with the property that only noncritical multipliers correspond to the solution (there are 27 such problems of 80), the primal trajectory converged superlinearly, while the dual trajectory converged to some Lagrange multiplier.

Table 1 shows the results obtained for problems with the property that at least one critical multiplier corresponds to the solution (41 of 80 problems). The SL column registers the cases of the superlinear convergence of primal trajectories, while the CM column indicates the cases of the convergence to critical multipliers. The black circle means that the corresponding behavior was observed in all the runs, and the

white circle signifies that this behavior occurred only in a portion of the runs. As can be seen from the table, the convergence to critical multipliers is typical, but the superlinear convergence of primal trajectories is usually preserved. On the other hand, the loss of superlinear convergence is always related to the convergence to critical multipliers.

For the MacMPEC problems, a single run was performed for each variant of the algorithm. The initial points were those indicated for the models of this collection. For most of these problems, neither the solution nor the corresponding Lagrange multipliers are known. Consequently, here, the convergence rate is understood as the rate of decay of the residual for the KKT system, which is calculated at each step of ALGENCAN.

For M2, we obtained the following results: for 28 problems of 161, the runs were failures; of the remaining 133 problems, only 16 problems exhibited linear convergence; for the other problems, the convergence was superlinear.

The basic variant M1 turned out to be considerably more robust: its runs were failures for only ten problems. As should be expected, the superlinear convergence was missing much more often than for M2, namely, for 98 runs of 151. However, for 30 of these problems, the linear convergence was very rapid: at each iteration step, the residual of the KKT system was reduced more than tenfold. The superlinear convergence was registered for 53 problems.

The results presented above suggest the following conclusions. If the penalty parameter grows infinitely, then the multiplier method often preserves the superlinear convergence in spite of the irregularity of constraints and the attraction of dual trajectories by critical Lagrange multipliers (if they exist). However, as should be expected, the infinite growth of the penalty parameter has a negative effect upon the robustness of the method.

3.2. Experiment with the Rules for the Penalty Parameter: Overall Efficiency

As noted above, the high convergence rate of the outer iterative process in the multiplier method may not compensate the effort spent for attaining this rate. A rapid growth of the costs of solving subproblems may have a decisive negative influence on the overall efficiency (and even the robustness) of the method.

To compare the efficiency of multiplier methods with different procedures for controlling the penalty parameter, we added another two variants of Algorithm 1 to the variants M1 and M2, described in the preceding section. The variant M3 is analogous to M2, but the penalty parameter is increased by a factor of $\rho = 10$ at each iteration step. In the variant M4, the penalty parameter is also increased tenfold at each iteration step; however, similarly to the basic ALGENCAN, restarts for this parameter are used and the value of c_0 is determined with the help of the initialization procedure (see Section 2).

Similar experiments were described in [20, 21]. The authors investigated the influence of different procedures for controlling the penalty parameter on the robustness and efficiency of ALGENCAN. In [20], the basic ALGENCAN was compared with its modifications in which the penalty parameter grows at each iteration step. The influence of restarts and various values of the parameter ρ was examined in [21]. However, in contrast to this paper, the tests in [20, 21] were performed for problems with regular constraints.

In the experiment under discussion, the accelerator was switched off for all variants of the algorithm. For the rest, the default settings of ALGENCAN were set up. In particular, scaling was used, and the bound for the number of inner iteration steps at the first outer iteration was kept at 10.

The comparison of the various variants of ALGENCAN was conducted in the following four characteristics: the number of outer and inner iteration steps and the number of evaluations of the objective function and the constraints. The number of outer iteration steps reflects the convergence rate of the algorithm, while the other characteristics can be regarded as different indicators of its efficiency.

The results of the tests based on the MacMPEC collection are shown in Fig. 3 in the form of the so-called "performance profiles" (see [26]). For each algorithm, we present in Fig. 3 the graph of the function defined as follows. Its value at the point $\tau \in [1, +\infty]$ specifies the fraction of problems in the series for which the result of the corresponding algorithm (that is, the number of iteration steps or the number of function evaluations) was at most τ times worse than the best result over the algorithms under comparison. It is assumed that the result of an unsuccessful run is infinitely many times worse than that of any successful run. In particular, the value of this function at $\tau = 1$ yields the fraction of runs for which the result of the corresponding algorithm was the best one. Its value for large τ characterizes the robustness of the algorithm, that is, the fraction of successful runs. The solid, dashed, dot-dashed, and dotted lines correspond to M1, M2, M3, and M4, respectively.

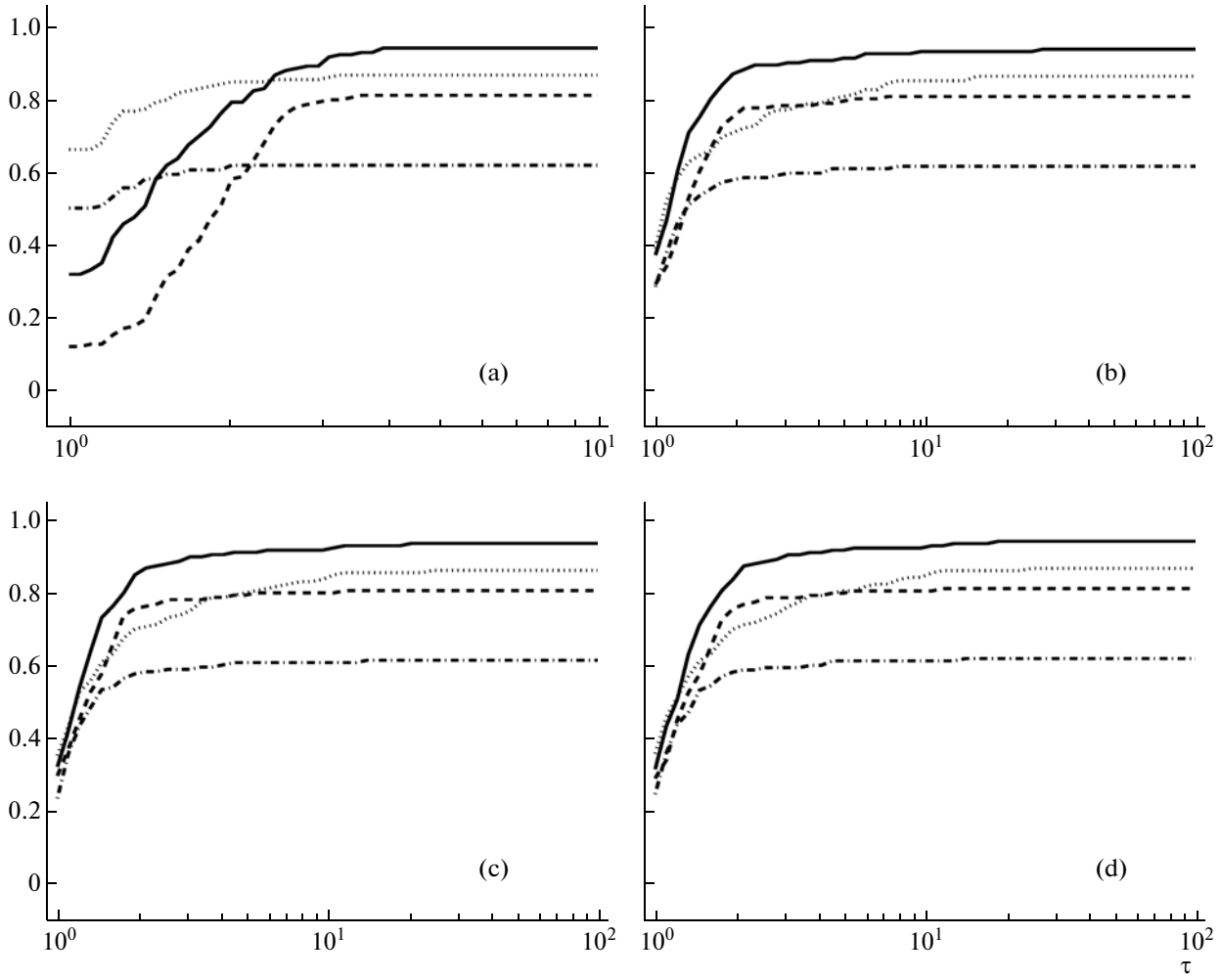

Fig. 3. (a) outer iterations, (b) inner iterations, (c) objective function evaluations, (d) constraint evaluations.

As can be seen from Fig. 3, the robustness of all the variants with the penalty parameter increasing at each iteration step is inferior to that of the basic ALGENCAN. In M2–M4, the penalty parameter more often attains the value 10^{20}, and, in such cases, the run is regarded as a failure. Furthermore, the robustness of M4 is superior to that of M2 and (especially) to the robustness of M3. Thus, the adaptive initialization of the penalty parameter and restarts increase the probability of a successful run.

Figure 3a also demonstrates that M3 and (especially) M4 are superior to the basic variant M1 in terms of the number of outer iteration steps. This indicates a higher convergence rate of M3 and (especially) M4. By contrast, the results of M2 are worse than those of M1. Although the penalty parameter grows at each step of M2, this growth is too slow.

On the other hand, Figs. 3b–3d show that all the variants with the penalty parameter increasing at each iteration step (even those that have a higher convergence rate) are no more efficient than the basic ALGENCAN. Note also that M4 is slightly more efficient than M2 and M3.

To present the results obtained for the DEGEN collection, we modify the construction from [26] to take into account the fact that, for these problems, each algorithm is run repeatedly. Namely, for each algorithm, we present the graph of the function $\pi : [1, +\infty) \longrightarrow [0, 1]$ defined as follows. Denote by k_p the average result per successful run and by s_p the fraction of successful runs of the corresponding algorithm for the pth problem in the collection. Let r_p be the minimum value of k_p taken over all algorithms. Then,

$$\pi(\tau) = \frac{1}{P} \sum_{p \in R(\tau)} s_p,$$

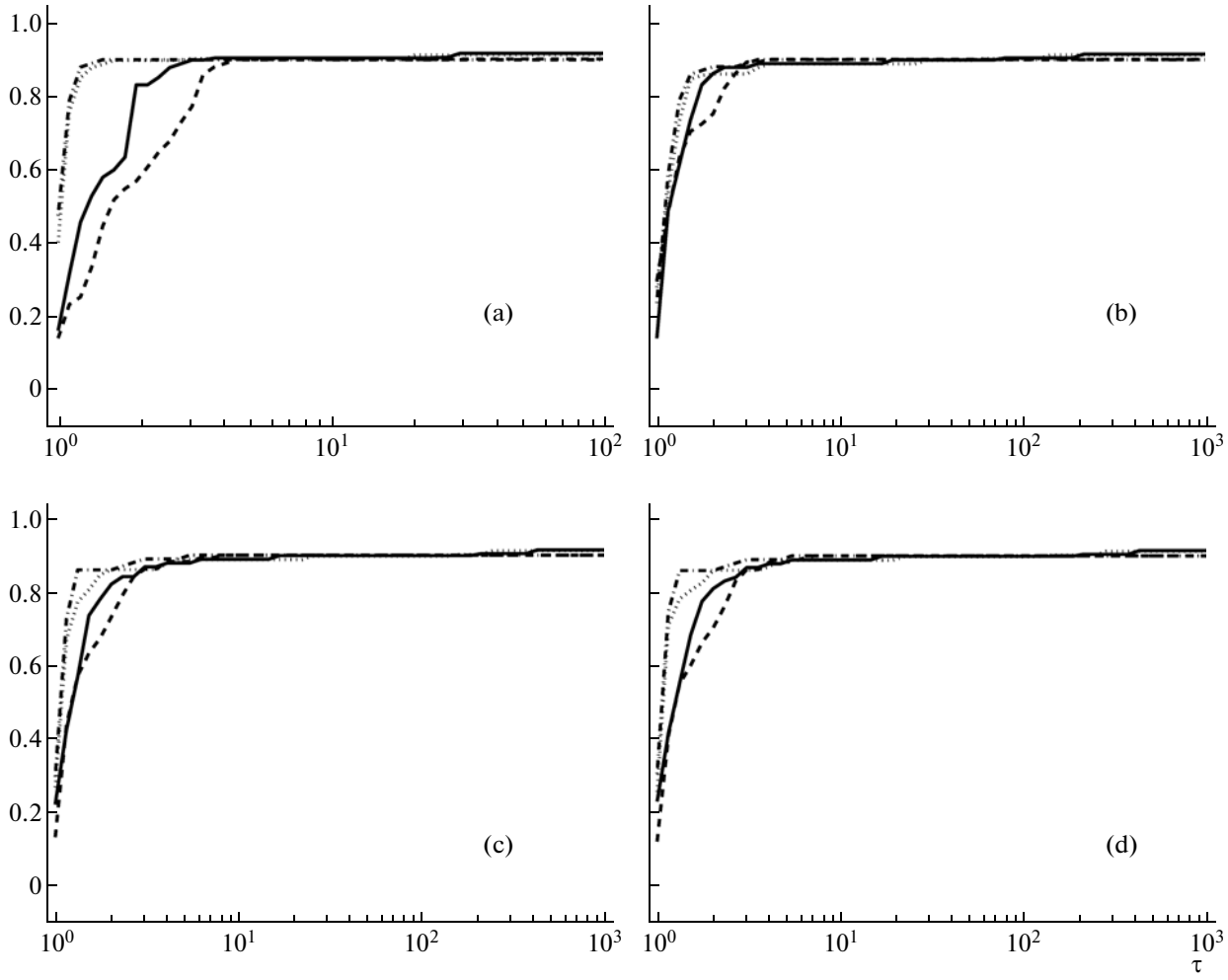

Fig. 4. (a) outer iterations, (b) inner iterations, (c) objective function evaluations, (d) constraint evaluations.

where P is the total number of problems in the collection and $R(\tau)$ is the set of problems for which the result of the corresponding algorithm is at most τ times worse than the best result; that is,

$$R(\tau) = \{p = 1, \ldots, P | k_p \leqslant \tau r_p\}.$$

It can be seen from Fig. 4 (where the correspondence between the lines and the variants is the same as in Fig. 3), that, for problems in the DEGEN collection, all the variants have about the same robustness. The results concerning the number of outer iteration steps (see Fig. 4a) repeat the situation for MacMPEC, although the distinctions between M1 and M2 and between M3 and M4 are much less significant. However, Figs. 4b–4d show that, in contrast to MacMPEC, M3 and M4 are also slightly superior to M1 and M2 in terms of other characteristics.

These results confirm the fears stated above. In numerous cases, the high convergence rate of the multiplier method, which is attained even for problems with irregular constraints at the expense of an infinite growth of the penalty parameter, does not lead to the high overall efficiency of this method.

3.3. Experiment with the Accelerators

In closing, we present the results of comparing two variants of ALGENCAN with an accelerator activated, namely, the basic variant with NLM as an accelerator and the variant using SNLM as an accelerator.

For the DEGEN collection, the tests were conducted in the same way as in Section 3.1. For ten of 98 problems, all the runs of both variants either were failures or resulted in the convergence to a solution at which the constraints were regular. For another 14 problems, neither accelerator was switched on because ALGENCAN managed to attain the required accuracy (that is, to find an approximation satisfy-

Table 2. Results for problems with critical multipliers from the DEGEN collection

Problem	SA	SL	CM	Problem	SA	SL	CM
10201	●/●	/●	●/	20214	●/●	/	●/●
10203	●/●	/	●/●	20216	●/●	/○	●/○
10206	●/●	/●	●/	20217	●/●	/	●/●
10207	●/●	/●	●/	20219	●/●	/	●/●
10301	/	/	/	20226	●/●	/○	●/●
10403	○/○	●/●	/	20227	●/●	/	●/●
20101	●/●	●/●	/	20302	●/●	/○	●/○
20103	●/●	/○	●/○	20303	●/●	○/●	○/
20105	●/●	●/●	/	20307	●/	/	●/
20107	●/●	○/●	○/	20309	●/●	/	●/●
20109	●/●	/	●/●	20312	○/○	/	●/●
20202	●/●	/	●/●	2DD01-5v	●/●	○/○	○/○
20203	○/○	/	●/●	30211	○/○	●/●	/
20204	●/●	○/●	○/	30301	●/●	●/●	○/
20207	●/●	/	●/●	30401	●/●	○/●	○/
20209	○/○	○/○	○/○	40201	○/○	●/●	/
20210	●/●	/	●/●	40206	/	/	/
20211	●/●	●/●	/	40208	/	/	/
20213	/●	/	/●	40402	●/●	/●	●/

ing condition (2.6)) before test (2.5) for activating the accelerator was fulfilled so that only steps of the multiplier method were used. Therefore, we present the results obtained for the remaining 74 problems. For some of these problems, the accelerator was activated at several outer iteration steps (that is, its runs were failures at all such steps with the possible exception of the last step). In such cases, only the trajectory generated by the accelerator at the last outer iteration was analyzed.

As mentioned above, there are 12 problems with no Lagrange multiplier corresponding to the solution. For eleven such problems, the runs of the accelerators always were failures. For the remaining problem 2DD01-5h, all the runs of both accelerators were successful; moreover, the primal convergence was always superlinear.

Table 2 shows the results obtained for problems having critical multipliers (38 of 74 problems). The results for NLM and SNLM are separated by slashes. The SA column marks the cases of successful runs of the accelerators, that is, the cases where the solution was found on an accelerator stage. The SL column registers the cases of the superlinear convergence of primal trajectories, while the CM column indicates the cases where the dual trajectories were attracted by critical multipliers. Black and white circles are interpreted as in Table 1.

It can be seen from the table that virtually all the runs were successful for both accelerators. However, their convergence rate and the behavior of dual trajectories were somewhat different.

For NLM, we virtually always observed the convergence to critical multipliers. There were only several cases where primal trajectories converged superlinearly; in almost all of these cases, dual trajectories converged to noncritical multipliers. For SNLM, the superlinear convergence of primal trajectories occurred significantly more often, while dual trajectories less often converged to critical multipliers. Still, the cases of such convergence, as well as the cases of linear primal convergence, were prevailing. Similarly to the situation with the NLM, in virtually all cases of superlinear convergence, dual trajectories converged to noncritical multipliers. Thus, the convergence to critical multipliers in most cases destroys superlinear convergence for both accelerators under discussion.

It should be noted that, in the presence of inequality constraints, a similar effect can be observed even in the case when only noncritical multupliers are associated with the solution. The reason is that NLM or SNLM is applied to the corresponding problem (2.9) with auxiliary variables, and this problem can contain critical multipliers. Moreover, it is easy to verify that, in a certain sense, the presence of critical multipliers in problem (2.9) is typical. Indeed, from [1, Proposition 5], we can derive the following conclu-

Table 3. Results for problems without critical multipliers from the DEGEN collection

Problem	SA	SL	Problem	SA	SL
00301	●/●	●/●	20222	○/●	●/●
00302	○/○	/	20225	○/○	○/○
00501	/	/	30202	●/○	●/●
10202	●/●	●/●	30205	●/●	●/●
10204	/●	/●	30206	○/○	●/●
10205	○/○	/	30303	●/●	●/●
20102	●/●	●/●	40203	/	/
20104	○/●	/●	40204	●/●	●/●
20106	/	/	01001	●/●	●/●
20110	●/●	●/●	10401	○/○	○/○
20111	/	/	10404	○/○	○/○
20221	●/●	●/●	10701	○/○	○/○

sion: if $A(\bar{x}) \subset \hat{A}$, then any Lagrange multiplier $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ violating the strict complementarity condition (that is, $\mu_i = 0$ for some $i \in A(\bar{x})$) is associated with the critical Lagrange multiplier $(\lambda, \mu_{\hat{A}})$ of problem (2.9). Furthermore, it is evident from (1.2) that the set

$$\mathcal{M}_A(\bar{x}) = \{\mu_{A(\bar{x})} | (\lambda, \mu) \in \mathcal{M}(\bar{x})\}$$

is the intersection of an affine set in $\mathbb{R}^{|A(\bar{x})|}$ with the nonnegative orthant. Consequently, the inequality $\mu_{A(\bar{x})} > 0$ can hold for all $\mu_{A(\bar{x})} \in \mathcal{M}_A(\bar{x})$ only in the case where $\mathcal{M}_A(\bar{x})$ consists of a single element. This implies that the vector μ is the same for all multipliers $(\lambda, \mu) \in \mathcal{M}(\bar{x})$, which is atypical for problems with irregular constraints.

Table 3 shows the results obtained for problems with the property that only noncritical multipliers correspond to the solution (24 of 74 problems). Note that, for almost all of these problems, the corresponding problems with auxiliary variables have critical multipliers. At the same time, the overall picture is somewhat different from the one given by Table 2. The results for both accelerators under testing almost not differ. The fraction of unsuccessful accelerator runs is fairly considerable, yet successful runs clearly prevail. Moreover, for successful runs, primal trajectories virtually always converged superlinearly.

The behavior of the accelerators demonstrated by Tables 2 and 3 is related to another heuristics implemented in ALGENCAN. It consists in an adaptive perturbation of the diagonal entries of the matrices that appear on the left-hand of systems (2.12) and (2.15). The purpose of these perturbations is to overcome the possible singularity of these matrices (see [10]). Perturbations of the diagonal entries of the matrix in (2.12) make the behavior of NLM closer to that of SNLM. For some problems, this results in that the primal trajectories of NLM converge superlinearly. If this heuristics is switched off, then the behavior of the accelerators becomes closer to what is expected of them. In particular, primal trajectories of NLM often exhibit linear convergence and the corresponding dual trajectories converge to critical Lagrange multipliers of problem (2.9). In some cases, the primal convergence of NLM remains superlinear, which is related to the presence of auxiliary variables. Nevertheless, if all the variables, including the auxiliary ones, are taken into account, then the primal convergence is linear, and the same is true of the convergence of residual (2.14) of the Lagrange system (2.10).

For the MacMPEC collection, the testing was conducted in the same way as in Section 3.1. The results obtained for NLM are as follows. For 22 of 161 problems, no accelerator runs were made. For another 29 problems, all the accelerator runs were failures. For 82 of the remaining 110 problems, the residual of the Lagrange system converged superlinearly along the trajectory of the accelerator. As for SNLM, it was never run for 22 problems. For 39 problems, all the accelerator runs were failures. For most of the remaining 100 problems (namely, for 74 problems), we registered superlinear convergence. Note that, in many cases, the superlinear convergence of NLM is related to the above heuristics implemented in ALGENCAN. When this heuristics was switched off, the convergence was almost always linear.

The comparison of the combinations of ALGENCAN with the two accelerators, analogous to the experiment discussed in Section 3.2, also did not reveal any significant distinctions in their behavior.

Summing up, we can say that no significant improvement in either efficiency or robustness of the solver is obtained when the standard accelerator NLM used in ALGENCAN is replaced with SNLM.

4. CONCLUDING REMARKS

From the theoretical viewpoint, the multiplier methods are attractive as a strategy for solving optimization problems with potentially irregular constraints. However, the high costs of solving subproblems (especially at late iteration steps, where a high accuracy is required of the solutions to these subproblems, and for large values of the penalty parameter) do not allow one to effectively combine, for problems in this class, the global convergence of such methods with a rapid convergence. Moreover, the simplest modifications of the multiplier methods based on the use various accelerators (such as the Newton–Lagrange method and its stabilized variant) are also inadequate for attaining this goal. The main reason is the fact that the dual trajectories of both multiplier methods and accelerators are attracted by critical Lagrange multipliers.

Thus, to design an actually efficient globally converging algorithm for optimization problems with irregular constraints, it is not sufficient to use the stabilized Newton–Lagrange method as an accelerator of the multiplier methods. Other principles are required for integrating these methods. The development of such principles is the subject of the future research of these authors.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 10-01-00251).

REFERENCES

1. A. F. Izmailov and M. V. Solodov, "Stabilized SQP Revisited," Math. Program. **133**, 93–120 (2012).
2. A. F. Izmailov, "On the Analytical and Numerical Stability of Critical Lagrange Multipliers," Comput. Math. Math. Phys. **45**, 930–946 (2005).
3. A. F. Izmailov and M. V. Solodov, "Examples of Dual Behavior of Newton-Type Methods on Optimization Problems with Degenerate Constraints," Comput. Optim. Appl. **42**, 231–264 (2009).
4. M. R. Hestenes, "Multiplier and Gradient Methods," J. Optim. Theory Appl. **4**, 303–320 (1969).
5. M. J. D. Powell, "A Method for Nonlinear Constraints in Minimization Problems," *Optimization* (Academic, New York, 1969), pp. 283–298.
6. D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic, New York, 1982; Radio i Svyaz', Moscow, 1987).
7. A. F. Izmailov and V. M. Solodov, *Numerical Optimization Methods*, 2nd. ed. (Fizmatlit, Moscow, 2008) [in Russian].
8. LANCELOT, http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml.
9. ALGENCAN, http://www.ime.usp.br/~egbirgin/tango/.
10. E. G. Birgin and J. M. Martnez, "Improving Ultimate Convergence of an Augmented Lagrangian Method," Optim. Meth. Software **23** (2), 177–195 (2008).
11. R. Andreani, E. G. Birgin, J. M. Martnez, and M. L. Schuverdt, "On Augmented Lagrangian Methods with General Lower-Level Constraints," SIAM J. Optim. **18**, 1286–1309 (2007).
12. R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, "A Relaxed Constant Positive Linear Dependence Constraint Qualification and Applications," Math. Program. **135**, 255–273 (2012).
13. A. F. Izmailov, M. V. Solodov, and E. I. Uskov, "Global Convergence of Augmented Lagrangian Methods Applied to Optimization Problems with Degenerate Constraints, Including Problems with Complementarity Constraints," SIAM J. Optimizat. (to appear).
14. 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).
15. A. F. Izmailov and M. V. Solodov, "On Attraction of Newton-Type Iterates to Multipliers Violating Second-Order Sufficiency Conditions," Math. Program. **117**, 271–304 (2009).
16. A. F. Izmailov and M. V. Solodov, "On Attraction of Linearly Constrained Lagrangian Methods and of Stabilized and Quasi-Newton SQP Methods to Critical Multipliers," Math. Program. **128**, 231–257 (2011).
17. A. F. Izmailov, "On the Limiting Properties of Dual Trajectories in the Lagrange Multipliers Method," Comput. Math. Math. Phys. **51**, 1–20 (2011).
18. S. J. Wright, "Superlinear Convergence of a Stabilized SQP Method to a Degenerate Solution," Comput. Optim. Appl. **11**, 253–275 (1998).
19. M. M. Golishnikov and A. F. Izmailov, "Newton-type Methods for Constrained Optimization with Nonregular Constraints," Comput. Math. Math. Phys. **46**, 1299–1319 (2006).
20. E. G. Birgin, D. Fernández, and J. M. Martínez, "On the Boundedness of Penalty Parameters in an Augmented Lagrangian Method with Lower Level Constraints," Optim. Methods Software **27**, 1001–1024 (2012).
21. E. G. Birgin and J. M. Martínez, "Augmented Lagrangian Method with Nonmonotone Penalty Parameters for Constrained Optimization," Comput. Optim. Appl. **51**, 941–965 (2012).
22. DEGEN, http://w3.impa.br/~optim/DEGEN_collection.zip.
23. The HSL Mathematical Software Library, http://www.hsl.rl.ac.uk.
24. MacMPEC, http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC.
25. A. F. Izmailov, *Sensitivity in Optimization* (Fizmatlit, Moscow, 2006) [in Russian].
26. E. Dolan and J. More, "Benchmarking Optimization Software with Performance Profiles," Math. Program. **91** (2), 201–213 (2002).