Math. Program., Ser. A (2016) 160:547–555
DOI 10.1007/s10107-016-0994-6

SHORT COMMUNICATION

Newton's method may fail to recognize proximity to optimal points in constrained optimization

R. Andreani[1] · **J. M. Martínez**[1] · **L. T. Santos**[1]

Received: 18 March 2015 / Accepted: 10 February 2016 / Published online: 23 February 2016

Abstract We will show examples in which the primal sequence generated by the Newton–Lagrange method converges to a strict local minimizer of a constrained optimization problem but the gradient of the Lagrangian does not tend to zero, independently of the choice of the dual sequence.

Keywords Constrained optimization · Newton–Lagrange method · Sequential optimality conditions · Stopping criteria

Mathematics Subject Classification 90C30 · 90C46 · 90C55

1 Introduction

In this section we will consider the equality-constrained minimization problem

$$\text{Minimize} \quad f(x) \quad \text{subject to} \quad h(x) = 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth. We say that $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ is feasible and satisfies the Karush–Kuhn–Tucker (KKT) conditions if $h(x) = 0$ and $\nabla f(x) + \nabla h(x)\lambda = 0$, where $\nabla w(x) = [\nabla w_1(x), \ldots, \nabla w_m(x)]$ denotes the

✉ L. T. Santos
lucio@ime.unicamp.br

R. Andreani
andreani@ime.unicamp.br

J. M. Martínez
martinez@ime.unicamp.br

1 DMA/IMECC, University of Campinas – UNICAMP, Campinas, SP, Brazil

Jacobian matrix of a function $w : \mathbb{R}^n \to \mathbb{R}^m$. If x is a local minimizer of (1) and satisfies some constraint qualification, there exist Lagrange multipliers λ_i for which the KKT conditions are verified. In the absence of constraint qualifications, however, even a global minimizer may not satisfy KKT conditions at all. For example, in the problem

$$\text{Minimize} \quad x \quad \text{subject to} \quad x^2 = 0, \tag{2}$$

the global solution $x^* = 0$ does not satisfy the KKT conditions for any choice of the multiplier λ.

Fortunately, it can be proved [1,4] that, for any local minimizer x^* of (1), *there exists* a sequence $\{(x^k, \lambda^k)\}$ such that $\lim x^k = x^*$, $\lambda^k \in \mathbb{R}^m$, and $\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k\| = 0$, where the symbol $\|\cdot\|$ denotes an arbitrary norm.

Given an algorithm for constrained optimization and a sequence $\{x^k\}$ generated by this algorithm that converges to a local minimizer, we have three possibilities:

1. The algorithm generates, together with $\{x^k\}$, a dual sequence $\{\lambda^k\} \subset \mathbb{R}^m$ such that $\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k\| = 0$.
2. The algorithm generates, together with $\{x^k\}$, a dual sequence $\{\lambda^k\} \subset \mathbb{R}^m$, but the property "$\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k\| = 0$" fails to hold.
3. For every choice of the dual sequence $\{\lambda^k\} \subset \mathbb{R}^m$, the property "$\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k\| = 0$" does not hold.

Algorithms that satisfy the first property for all generated sequences $\{x^k\}$ that converge to local minimizers will be said to belong to the First Class. Analogously, an algorithm that may generate some sequence with the second property belongs to the Second Class and algorithms that may generate sequences with the third property belong to the Third Class.

If an algorithm belongs to the First Class, it is natural to employ the stopping criterion given by

$$\left\|\nabla f(x^k) + \nabla h(x^k)\lambda^k\right\| \le \varepsilon_1 \tag{3}$$

and

$$\left\|h(x^k)\right\| \le \varepsilon_2, \tag{4}$$

for small values of $\varepsilon_1, \varepsilon_2 > 0$, since we can guarantee that (3) and (4) hold if x^k is close enough to the minimizer.

However, if an algorithm belongs to the Second Class the stopping criterion (3)–(4) may not hold (with the multipliers generated by the method) even when $\{x^k\}$ converges to a solution. If an algorithm belongs to the Third Class, the condition (3) may not hold for a convergent sequence $\{x^k\}$, even if we compute the dual sequence in the best possible way. In particular, the stopping criterion may not hold when, after computing x^k, we take λ^k as the minimizer of $\|\nabla f(x^k) + \nabla h(x^k)\lambda\|$ with respect to $\lambda \in \mathbb{R}^m$. This means that the algorithm may not recognize the proximity to a minimizer x^* even being able to approximate such solution with arbitrary precision.

In [2] we proved that Newton's (also called Newton–Lagrange) method belongs to the Second Class. For example, by formula (22) of [2], when one applies Newton's method to the nonlinear KKT system derived from the problem (2), the generated sequence $\{(x^k, \lambda^k)\}$ verifies $\lim |\nabla f(x^k) + \nabla h(x^k)\lambda^k| = 1/3$. However, the family of examples presented in [2] is not enough to include Newton's method in the third class of algorithms, since, in all these examples, replacing the natural multipliers λ^k generated by Newton's method with alternative suitable multipliers λ^k_{opt}, the fulfillment of $\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k_{opt}\| = 0$ takes place.

Therefore, the natural question is whether Newton's method may generate a sequence $\{x^k\}$ that converges to a local minimizer x^*, with the property that, for every choice of $\{\lambda^k\} \subset \mathbb{R}^m$, one has that $\lim \|\nabla f(x^k) + \nabla h(x^k)\lambda^k\| = 0$ does not hold.

In Sect. 2 we will prove that Newton's method in fact belongs to the Third Class. We will prove that, for a very simple problem with only equality constraints, sequences $\{x^k\}$ with the third property exist.

In Sect. 3, using a simple modification of the example in Sect. 2, we will prove that a similar property holds for the inequality constrained optimization problem. Moreover, we provide numerical evidence of such property for an example in which the feasible set does not coincide with its boundary.

2 An example

In this section we will present an equality constrained optimization problem and a sequence $\{x^k\}$ generated by Newton's method such that, for every choice of $\lambda^k \in \mathbb{R}^m$, $\|\nabla f(x^k) + \nabla h(x^k)\lambda^k\|$ does not tend to zero. From now on we define $\|\cdot\| = \|\cdot\|_2$. Let us define

$$\underline{\lambda}^k = \arg\min_{\lambda} \left\| \nabla f(x^k) + \nabla h(x^k)\lambda \right\|^2. \tag{5}$$

Clearly, for proving the desired property it is enough to prove that $\|\nabla f(x^k) + \nabla h(x^k)\underline{\lambda}^k\|_2$ does not tend to zero. Moreover, since $\nabla f(x^k) + \nabla h(x^k)\underline{\lambda}^k$ is the projection of $-\nabla f(x^k)$ on the null-space of $\nabla h(x^k)^T$, the fact that $\|\nabla f(x^k) + \nabla h(x^k)\underline{\lambda}^k\|_2$ does not tend to zero implies that $\{x^k\}$ is a sequence that does not detect the Approximate Gradient Projection (AGP) optimality condition introduced in [9] (see, also, [1,4]).

The problem is the following:

$$\text{Minimize} \quad x_1 \quad \text{subject to} \quad x_1^2 + x_2^2 = 0. \tag{6}$$

The KKT (with feasibility) conditions for (6) are

$$g + 2\lambda x = 0 \quad \text{and} \quad \|x\|^2 = 0, \tag{7}$$

where $x \in \mathbb{R}^2$, $\lambda \in \mathbb{R}$ and $g = (1, 0)^T$. Clearly, $x^* = (0, 0)^T$ is the global solution and there are no Lagrange multipliers associated to it. Applying Newton's method to (7), we obtain the following iteration scheme:

$$\begin{bmatrix} 2\lambda^k I & 2x^k \\ 2(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} & - x^k \\ \lambda^{k+1} & - \lambda^k \end{bmatrix} = - \begin{bmatrix} g + 2\lambda^k x^k \\ \|x^k\|^2 \end{bmatrix}, \tag{8}$$

where I is the identity matrix of order 2. Rearranging the equations, we find

$$x^{k+1} = \frac{1}{2}\left(x^k + \frac{x_2^k}{\lambda^k \|x^k\|^2}\begin{bmatrix} -x_2^k \\ x_1^k \end{bmatrix}\right) \quad \text{and} \quad \lambda^{k+1} = \frac{1}{2}\left(\lambda^k - \frac{x_1^k}{\|x^k\|^2}\right). \tag{9}$$

The iterations are well defined under the condition that $\lambda^k \|x^k\| \neq 0$, for all $k = 0, 1, 2, \ldots$. Moreover, the least-square Lagrange multiplier and its associated KKT residual are given by

$$\underline{\lambda}^k = \arg\min_{\lambda} \left\| g + 2\lambda x^k \right\|^2 = -\frac{x_1^k}{2\|x^k\|^2}, \tag{10}$$

and

$$\underline{r}^k = \left\| g + 2\underline{\lambda}^k x^k \right\| = \frac{|x_2^k|}{\|x^k\|}. \tag{11}$$

From (9), if we define a new variable $y = 2\lambda x$, we have, after some algebraic manipulation, that

$$y^{k+1} = a^k \left[-g + \left(1 - a^k\right) y^k\right], \quad \text{with} \quad a^k = \frac{1}{2} - \frac{y_1^k}{\|y^k\|^2}. \tag{12}$$

Observe that the condition $\lambda^k \|x^k\| \neq 0$ is equivalent to $\|y^k\| \neq 0$. Define $\alpha = -(\sqrt{5}+1)/4 \approx -0.809$ and $\beta = \sqrt{(5-\sqrt{5})/40} \approx 0.263$.

Lemma 1 *There exists* $0 < \epsilon < \min\{|\alpha|, \beta^2\}$ *such that if* $|y_1^0 - \alpha| < \epsilon$ *and* $|(y_2^0)^2 - \beta^2| < \epsilon$ *then the iteration scheme* (12) *is well defined,* $\lim y_1^k = \alpha$*, and* $\lim |y_2^k| = \beta$.

Proof Let us define $z = (y_1, (y_2)^2)^T$ and $z^* = (\alpha, \beta^2)^T$. Therefore, the iteration scheme (12) is transformed to $z^{k+1} = \Phi(z^k)$, where

$$\Phi(z) = \begin{bmatrix} a(z)[-1 + (1 - a(z))z_1] \\ a(z)^2(1 - a(z))^2 z_2 \end{bmatrix}, \quad \text{with} \quad a(z) = \frac{1}{2} - \frac{z_1}{z_1^2 + z_2}. \tag{13}$$

It is easy to check that $\Phi(z^*) = z^*$, i.e., z^* is a fixed point of function Φ. The Jacobian matrix of Φ at z^* is given by

$$\nabla\Phi(z^*) = \frac{1}{8}\begin{bmatrix} \sqrt{5} - 3 & 5 - \sqrt{5} \\ -10 & 5\sqrt{5} - 7 \end{bmatrix}, \tag{14}$$

which has two complex eigenvalues with modulus $(\sqrt{5}-1)/2 \approx 0.618 < 1$. Therefore, from the Fixed Point Theorem, there exists $0 < \delta < \min\{|\alpha|, \beta^2\}$ such that if $\|z^0 - z^*\| < \delta$ then the iterations are well defined and $\lim z^k = z^*$. Therefore, there exists $0 < \epsilon < \delta$ such that $|y_1^0 - \alpha| < \epsilon$ and and $|(y_2^0)^2 - \beta^2| < \epsilon$ implies that $\|z^0 - z^*\| < \delta$, and then $\lim y_1^k = \lim z_1^k = \alpha$ and $\lim |y_2^k| = \lim \sqrt{|z_2^k|} = \beta$, and the proof is complete. □

Lemma 2 *There exists* $\epsilon > 0$ *such that if* $|2\lambda^0 x_1^0 - \alpha| < \epsilon$ *and* $|(2\lambda^0 x_2^0)^2 - \beta^2| < \epsilon$ *then, in the iteration scheme* (9), $\lim |\lambda^k| = \infty$ *and* $\lim \|x^k\| = 0$.

Proof Let $\epsilon > 0$ be given by Lemma 1 and choose x^0 and λ^0 such that $|2\lambda^0 x_1^0 - \alpha| < \epsilon$ and $|(2\lambda^0 x_2^0)^2 - \beta^2| < \epsilon$. Therefore, $\lim y_1^k = \lim 2\lambda^k x_1^k = \alpha$, $\lim |y_2^k| = \lim |2\lambda^k x_2^k| = \beta$, and, from Eq. (9),

$$\lambda^{k+1} = \frac{1}{2}\left(\lambda^k - \frac{x_1^k}{\|x^k\|^2}\right) = \frac{1}{2}\left(\lambda^k - \frac{2\lambda^k y_1^k}{\|y^k\|^2}\right) = \lambda^k\left(\frac{1}{2} - \frac{y_1^k}{\|y^k\|^2}\right) = a^k\lambda^k, \quad (15)$$

where

$$\lim a_k = \frac{1}{2} - \frac{\alpha}{\alpha^2 + \beta^2} = \frac{\sqrt{5}+1}{2} \approx 1.618. \quad (16)$$

Then, for k sufficiently large, $|\lambda^{k+1}| > 1.5|\lambda^k|$, and so $\lim |\lambda^k| = \infty$. Now, since $\lim \|2\lambda_k x^k\|$ is finite, we must have $\lim \|x^k\| = 0$. □

Theorem 1 *Consider the iteration scheme* (9). *For any given* $\lambda^0 > 0$ *there exists* $u_1 > \ell_1 > 0$ *and* $u_2 > \ell_2 > 0$, *all of them depending on* λ^0, *such that if* $\ell_1 < -x_1^0 < u_1$ *and* $\ell_2 < |x_2^0| < u_2$, *then* $\lim x^k = (0, 0)^T$ *and* $\lim \underline{r}^k = (\sqrt{5} - 1)/4 > 0$. *Moreover,* $x_1^k < 0$ *and* $\lambda^k > 0$ *for all* $k = 0, 1, 2, \ldots$.

Proof Let $\epsilon > 0$ be given by Lemma 1. Since $\epsilon < \min\{|\alpha|, \beta^2\}$ we can define

$$\ell_1 = \frac{|\alpha| - \epsilon}{2\lambda^0}, \quad u_1 = \frac{|\alpha| + \epsilon}{2\lambda^0}, \quad \ell_2 = \frac{\sqrt{\beta^2 - \epsilon}}{2\lambda^0}, \quad \text{and} \quad u_2 = \frac{\sqrt{\beta^2 + \epsilon}}{2\lambda^0}. \quad (17)$$

Therefore, if $\ell_1 < -x_1^0 < u_1$ then $|2\lambda^0 x_1^0 - \alpha| < \epsilon$, and if $\ell_2 < |x_2^0| < u_2$ then $|(2\lambda^0 x_2^0)^2 - \beta^2| < \epsilon$. So, from Lemmas 1 and 2, $\lim 2\lambda^k x_1^k = \alpha$, $\lim |2\lambda^k x_2^k| = \beta$, and $\lim x^k = (0, 0)^T$. Now,

$$\lim r^k = \lim \|g + 2\lambda^k x^k\| = \sqrt{(1+\alpha)^2 + \beta^2} = \sqrt{\frac{5 - 2\sqrt{5}}{5}} \approx 0.325 > 0, \quad (18)$$

and

$$\lim \underline{r}^k = \lim \frac{|x_2^k|}{\|x^k\|} = \lim \frac{|2\lambda^k x_2^k|}{\|2\lambda^k x^k\|} = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} = \frac{\sqrt{5} - 1}{4} \approx 0.309 > 0. \quad (19)$$

Finally, the fact that $x_1^k < 0$ and $\lambda^k > 0$ for all k follows from $x_1^0 < 0$ and (9). □

Remark 1 By (8), since $\lambda^k > 0$ for all k, x^{k+1} is the unique minimizer of the strictly convex quadratic with Hessian $\lambda^k I$ and gradient g, subject to $2(x^k)^T(x - x^k) = -\|x^k\|^2$. In the general case, this is the reason for the denomination Sequential Quadratic Programming (SQP) usually given to several variations of the Newton–Lagrange method [10].

Remark 2 A different Newton-like method arises if, at every iteration, one replaces λ^k with the least-square approximation $\underline{\lambda}^k$. Such replacement obviously modifies the primal sequence which makes the analysis much more complicated. We do not have a rigorous mathematical characterization of this modified method in the general case. However, motivated by a question of an anonymous referee, we considered the counter-example analyzed in this section with the replacement $\lambda^k \leftarrow \underline{\lambda}^k$ and we verified that, in this particular example, the gradient of the Lagrangian at primal-dual sequences generated by the algorithm always tends to zero. Obviously, this algorithm is more expensive than the straight Newton–Lagrange method but deserves further research.

Remark 3 Denoting the Golden number by $\phi = (\sqrt{5}+1)/2$, we have that $\alpha = -\phi/2$ and $\lim \underline{r}^k = 1/(2\phi) = \sin 18^o$.

3 Problems with inequality constraints

Consider the general nonlinear programming problem,

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{subject to } h_i(x) = 0, \ i \in \mathcal{E}, \\ &\qquad\qquad\quad h_i(x) \leq 0, \ i \in \mathcal{I}, \end{aligned} \tag{20}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and the functions $h_i : \mathbb{R}^n \to \mathbb{R}$ are all sufficiently smooth and $\mathcal{E} \cup \mathcal{I} = \{1, \ldots, m\}$. According to the Local Sequential Quadratic Programming (SQP) method (see, for example, [10], p. 533) given $x^k \in \mathbb{R}^n$ and $\lambda^k \in \mathbb{R}^m$, with $\lambda_i^k \geq 0$ for all $i \in \mathcal{I}$, x^{k+1} is defined as a solution of the quadratic programming problem

$$\begin{aligned} &\text{Minimize } \tfrac{1}{2}(x-x^k)^T H_k (x-x^k) + \nabla f(x^k)^T (x-x^k) \\ &\text{subject to } \nabla h_i(x^k)^T (x-x^k) + h_i(x^k) = 0, \ i \in \mathcal{E}, \\ &\qquad\qquad\quad \nabla h_i(x^k)^T (x-x^k) + h_i(x^k) \leq 0, \ i \in \mathcal{I}, \end{aligned} \tag{21}$$

where H_k is the Hessian of the Lagrangian, given by

$$H_k = \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k). \tag{22}$$

Moreover, λ^{k+1} is defined as the vector of Lagrange multipliers corresponding to the solution of (21). Clearly, if the solution of this subproblem exists, we have that $\lambda_i^{k+1} \geq 0$ for all $i \in \mathcal{I}$.

Let us apply this method to the problem defined by

$$\text{Minimize} \quad x_1 \quad \text{subject to} \quad x_1^2 + x_2^2 \leq 0. \tag{23}$$

Therefore, $n = 2$, $m = 1$, $\mathcal{E} = \emptyset$, $\mathcal{I} = \{1\}$, $f(x_1, x_2) = x_1$, and $h(x_1, x_2) = x_1^2 + x_2^2$. Assume that $x_1^k < 0$ and $\lambda^k > 0$. Then, $H_k = 2\lambda^k I$ and, since this matrix is positive definite, the problem (21) has a unique solution. The only constraint in (21) is, in this case,

$$2(x^k)^T(x - x^k) + \|x^k\|^2 \leq 0, \tag{24}$$

and the objective function is

$$Q_k(x) = \lambda^k \|x - x^k\|^2 + \left(x_1 - x_1^k\right). \tag{25}$$

Writing the optimality conditions that corresponds to the minimization of $Q_k(x)$ subject to (24) we note that the constraint (24) is necessarily active at the solution and the following equations are satisfied,

$$2\lambda^k(x^{k+1} - x^k) + (1, 0)^T + 2\lambda^{k+1}x^k = 0, \tag{26}$$

and

$$2(x^k)^T(x^{k+1} - x^k) + \|x^k\|^2 = 0. \tag{27}$$

Simple calculations show that these equations are equivalent to (8) and the solution (x^{k+1}, λ^{k+1}) is given by (9). Therefore, by induction, if $x_1^0 < 0$ and $\lambda^0 > 0$ we have that the solution of the quadratic subproblem that defines the SQP iteration is well-defined, unique, the linear constraint is active, $x_1^k < 0$, and $\lambda^k > 0$ for all $k = 0, 1, 2, \ldots$. Since the recurrence relation (8) is the same as in the equality constrained case, the arguments that follows equation (9) in Sect. 2 can be repeated here, showing that, choosing x_1^0 as in Theorem 1, the norm of $\nabla f(x^k) + \underline{\lambda}^k \nabla h(x^k)$ does not tend to zero, independently of the choice of $\underline{\lambda}^k$.

Let us now consider the following problem,

$$\text{Minimize} \quad x_1 \quad \text{subject to} \quad -x_1^3 + x_2^2 \leq 0. \tag{28}$$

The point $x^* = (0, 0)$ is the global minimum with no Lagrange multipliers associated to it. The SQP iteration (21) for the above problem takes the form

$$\begin{aligned} &\text{Minimize } \lambda_k \left[-3x_1^k\left(x_1 - x_1^k\right)^2 + \left(x_2 - x_2^k\right)^2\right] + \left(x_1 - x_1^k\right) \\ &\text{subject to } -3\left(x_1^k\right)^2\left(x_1 - x_1^k\right) + 2x_2^k\left(x_2 - x_2^k\right) \leq \left(x_1^k\right)^3 - \left(x_2^k\right)^2. \end{aligned} \tag{29}$$

The solution of (29) defines x^{k+1} and $\lambda^{k+1} \geq 0$ is the corresponding Lagrange multiplier. If we take $x_1^0 < 0$ and $\lambda^0 > 0$, it is not difficult to show by induction that for all $k > 0$, $x_1^k < 0$ and $\lambda^k > 0$ (the linear constraint is always active.) The KKT residual is then given by

$$r^k = \sqrt{\left(1 - 3\lambda^k\left(x_1^k\right)^2\right)^2 + \left(2\lambda^k x_2^k\right)^2}, \tag{30}$$

and, for the least-square solution, we have that

$$\underline{\lambda}^k = \arg\min_{\lambda} \left(1 - 3\lambda\left(x_1^k\right)^2\right)^2 + \left(2\lambda x_2^k\right)^2 = \frac{3\left(x_1^k\right)^2}{9\left(x_1^k\right)^4 + 4\left(x_2^k\right)^2} \geq 0, \tag{31}$$

and

$$\underline{r}^k = \sqrt{\left(1 - 3\underline{\lambda}^k \left(x_1^k\right)^2\right)^2 + \left(2\underline{\lambda}^k x_2^k\right)^2} = \frac{2\left|x_2^k\right|}{\sqrt{9\left(x_1^k\right)^4 + 4\left(x_2^k\right)^2}}. \tag{32}$$

We have applied the SQP procedure (29), for a huge number of initial dual pairs (x^0, λ^0) with $x_1^0 < 0$ and $\lambda^0 > 0$. With no exception we obtained that $\lim x^k = x^*$, $\lim \lambda^k = +\infty$, $\lim r^k = +\infty$, $\lim \underline{\lambda}^k = +\infty$, and $\lim \underline{r}^k = 1$. These numerical experiments suggest that the SQP method applied to problem (28) generates a primal sequence $\{x^k\}$ such that the norm of the KKT residue does not converge to zero for any choice of the dual sequence. Therefore, there seems to be enough numerical evidence that (28) provides another example of a situation in which SQP generates a sequence that converges to a solution for which the approximate KKT stopping criterion is not satisfied, independently of the choice of the dual sequence.

4 Conclusions

The examples presented in this paper put an end on the question about the behavior of the local Newton–Lagrange method with respect to the existence sequences generated by the algorithm on which Approximate KKT and Approximate Gradient Projection optimality conditions can be detected. The answer turned out to be negative. The Newton–Lagrange method may converge to minimizers without detecting those *necessary* optimality conditions at all. On the other hand, we know that AKKT and AGP certainly hold [1].

Many open questions arise: With respect to Newton's method the question is whether the observed "anomaly" can be "fixed" by means of stabilization [5,6,11,12] or globalization approaches [7,8]. With respect to many other methods the question is whether they generate sequences for which sequential optimality conditions as AKKT or AGP can be detected. The answer is positive for some Augmented Lagrangian and Penalty-like methods [2,3] but remains unclear for most potentially useful algorithms introduced in the last 20 years.

Practical algorithms for constrained optimization stop, declaring convergence, when approximate feasibility, approximate complementarity, and approximate annihilation of the Lagrangian gradient take place. Here we showed that the Newton–Lagrange method, even converging to an isolated global minimizer, may fail to detect approximate annihilation of the Lagrangian gradient. Moreover, we showed that this inconvenience is associated with the primal sequence $\{x^k\}$ and cannot be overcome by means of alternative choices of the dual sequence $\{\lambda^k\}$. The practical consequence of this property could be that, in nontrivial problems, the presence of a very close local minimizer to the current iteration would not be detected at all.

Acknowledgments We wish to acknowledge the associate editor and two anonymous referees for insightful comments that helped us a lot to improve this paper. This work was kindly supported by the Brazilian research agencies CNPq (303013/2013–3, 305078/2013–5), PRONEX-CNPq/FAPERJ (E–26/111.449/2010–APQ1), and FAPESP (2010/10133–0, 2013/05475–7, 2013/07375–0).

References

1. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
2. Andreani, R., Martínez, J.M., Santos, L.T., Svaiter, B.F.: On the behavior of constrained optimization methods when Lagrange multipliers do not exist. Optim. Methods Softw. **29**, 646–657 (2014)
3. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**, 3533–3554 (2010)
4. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)
5. Fernández, D., Solodov, M.V.: Stabilized sequential quadratic programming: a survey. Pesquisa Oper. **34**, 591–604 (2013)
6. Izmailov, A., Solodov, M.V.: Stabilized SQP revisited. Math. Program. **133**, 93–120 (2012)
7. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Combining stabilized SQP with the augmented Lagrangian algorithm. Comput Optim Appl. **62**, 405–429 (2015)
8. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Globalizing stabilized sequential quadratic programming method by smooth primal-dual exact penalty function. J Optim Theor Appl, 1–31 (2016)
9. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. **118**, 117–133 (2003)
10. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, Berlin (2006)
11. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
12. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**, 470–497 (2002)