Optimization Methods & Software, 2014
Vol. 29, No. 3, 646–657, http://dx.doi.org/10.1080/10556788.2013.841692

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

On the behaviour of constrained optimization methods when Lagrange multipliers do not exist

R. Andreani[a], J.M. Martínez[a], L.T. Santos[a*] and B.F. Svaiter[b]

[a]*DMA/IMECC, University of Campinas—UNICAMP, Campinas, SP, Brazil;* [b]*Institute of Pure and Applied Mathematics—IMPA, Rio de Janeiro, RJ, Brazil*

(*Received 10 October 2012; accepted 27 August 2013*)

Sequential optimality conditions are related to stopping criteria for nonlinear programming algorithms. Local minimizers of continuous optimization problems satisfy these conditions without constraint qualifications. It is interesting to discover whether well-known optimization algorithms generate primal–dual sequences that allow one to detect that a sequential optimality condition holds. When this is the case, the algorithm stops with a 'correct' diagnostic of success ('convergence'). Otherwise, closeness to a minimizer is not detected and the algorithm ignores that a satisfactory solution has been found. In this paper it will be shown that a straightforward version of the Newton–Lagrange (sequential quadratic programming) method fails to generate iterates for which a sequential optimality condition is satisfied. On the other hand, a Newtonian penalty–barrier Lagrangian method guarantees that the appropriate stopping criterion eventually holds.

Keywords: constrained optimization; Newton–Lagrange method; sequential optimality conditions; stopping criteria

AMS Subject Classification: 90C30; 90C46; 90C55

1. Introduction

Numerical optimization involves the design of algorithms that presumably find the lowest possible value of a function within some domain. Necessary optimality conditions are useful tools for that purpose. Points that satisfy necessary conditions, usually called 'stationary points', are probable local minimizers and, in many cases, even global minimizers of the problem. On the other hand, algorithms that guaranteedly find stationary points are less expensive than algorithms that provide certificates of global optimality, and can be used as auxiliary tools for global optimization methods.

The most popular optimality conditions (Karush–Kuhn–Tucker, or briefly, KKT) say that the gradient of the objective function must belong to a cone defined by the gradients of the constraints. This condition holds at a local minimizer if some 'constraint qualification' is fulfilled. In the absence of constraint qualifications the gradient of the objective function may not belong to the cone determined by the gradients of the constraints. This situation was studied in recent papers, where sequential optimality conditions, which *do not require constraint qualifications at all*, were defined and analysed [3,4,35].

*Corresponding author. Email: lucio@ime.unicamp.br

© 2013 Taylor & Francis

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

The case in which the KKT conditions do not hold at a local minimizer, due to the absence of constraint qualifications, deserves to be studied for several reasons. Lack of linear independence of constraints is common in applications. In some cases this is due to redundancy or modelling problems, and degeneration can be removed by pre-processing or model analysis. If, in spite of those procedures, degeneration remains, it becomes useful to employ algorithms that properly handle that anomaly. Many examples of 'abnormal' problems have been given in [7]. On the other hand, severe ill-conditioning of the constraints usually causes extremely big Lagrange multipliers and it is plausible to conjecture that methods that deal efficiently with the non-existence of multipliers will be successful when multipliers are meaningless due to their extreme size. Abnormality may also occur when the constraints specify that some lower level function should be minimized and when, without previous knowledge of the modeller, the problem is infeasible.

Many successful algorithms have been introduced in the last 15 years for large-scale nonlinear programming ([2,9,11–14,16–19,24,25,27,31–33,38–41] and many others). Augmented Lagrangians, sequential quadratic programming (SQP), interior point, filter and dynamic infeasibility techniques have been employed with that purpose. From a practical point of view, it is interesting to study the behaviour of these methods in the case of local minimizers for which the classical KKT conditions do not hold. A partial study, addressing external penalty and augmented Lagrangian methods, may be found in [3]. The main question concerns the fulfillment of sequential optimality conditions at limit points of sequences generated by the methods. The fulfillment of a sequential optimality condition guarantees that, in finite time, a stopping criterion based on that condition will be verified. A rigorous treatment of stopping criteria for bound-constrained problems may be found in [28].

In Section 2 we state the theoretical background concerning sequential optimality conditions and stopping criteria based on them. In Section 3 we prove negative results concerning the application of the straightforward SQP method to the nonlinear programming problem. In Section 4 we show that Newtonian methods based on barriers for inequalities and penalties for equalities behave well with respect to approximate KKT criteria, when multipliers do not exist. Conclusions are stated in Section 5.

Notation: The symbol $\|\cdot\|$ will denote the Euclidian norm. If $h : \mathbb{R}^n \to \mathbb{R}^m$ we denote $\nabla h(x) = (\nabla h_1(x), \dots, \nabla h_m(x))$ and $h'(x) = \nabla h(x)^{\mathrm{T}}$. The symbol $\mathbb{R}^n_+$ denotes the elements of $\mathbb{R}^n$ with non-negative components, while $\mathbb{R}^n_{++}$ will be the set of elements of $\mathbb{R}^n$ with strictly positive components. For all $x \in \mathbb{R}^n_{++}$, we denote by X the diagonal matrix with elements $x_1, \dots, x_n$. For $v \in \mathbb{R}^n$, the components of v_+ are defined by $(v_+)_i = \max\{0, v_i\}$, $i = 1, 2, \dots, n$. We define $\mathbf{e} = (1, \dots, 1)^{\mathrm{T}}$.

2. Background

We will consider nonlinear programming problems given in the form

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

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. Under suitable constraint qualifications [1,5,6,10,37], if x is a local minimizer of (1), there exist Lagrange multipliers $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that the following KKT conditions are satisfied:

$$\nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu = 0 \tag{2}$$

and

$$\mu_i g_i(x) = 0, \quad \text{for all } i = 1, \dots, p. \tag{3}$$

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

In the absence of constraint qualifications, even global minimizers may fail to satisfy the KKT conditions. However, at every local minimizer x^*, the following 'approximate Karush–Kuhn–Tucker' (AKKT) conditions hold [4]: for all $\delta, \varepsilon > 0$ there exist $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p_+$ such that $\|x - x^*\| \leq \delta$

$$\|\nabla f(x) + \nabla h(x)\lambda + \nabla g(x)\mu\| \leq \varepsilon, \tag{4}$$

$$|\min\{-g_i(x), \mu_i\}| \leq \varepsilon, \quad \text{for all } i = 1, \dots, p, \tag{5}$$

$$\|h(x)\| \leq \varepsilon \quad \text{and} \quad \|g(x)_+\| \leq \varepsilon. \tag{6}$$

Moreover, if f, h and g satisfy a plausible additional smoothness condition, the following approximate complementarity property, stronger than (5), takes place [3]:

$$|\lambda_i h_i(x)| \leq \varepsilon, \quad \text{for all } i = 1, \dots, m \quad \text{and} \quad |\mu_i g_i(x)| \leq \varepsilon, \quad \text{for all } i = 1, \dots, p. \tag{7}$$

If $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$ satisfies (4), (5) and (6), we say that AKKT(ε) hold at (x, λ, μ). Analogously, the complementary approximate KKT conditions (CAKKT(ε)) hold at $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$ when the inequalities (4), (6) and (7) are fulfilled.

Both (4–5–6) and (4–7–6) define natural stopping criteria for constrained optimization algorithms. However, the question about the suitability of a practical algorithm with respect to the conditions above arises. We say that an algorithm is 'suitable' with respect to the AKKT condition if it generates a sequence $\{x^k, \lambda^k, \mu^k\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$ such that for all $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that (4)–(6) hold with $x = x^k$, $\lambda = \lambda^k$ and $\mu = \mu^k$. Analogously, an algorithm is suitable with respect to the CAKKT condition if it generates a sequence $\{x^k, \lambda^k, \mu^k\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$ such that for all $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that (4), (7) and (6) hold with $x = x^k$, $\lambda = \lambda^k$ and $\mu = \mu^k$.

An algorithm may generate a sequence $\{x^k\}$ that converges to a minimizer x^* without satisfying suitability conditions. This means that the generated primal–dual sequence $\{x^k, \lambda^k, \mu^k\}$ fails to satisfy the stopping criteria defined by approximate KKT conditions for some $\varepsilon > 0$ and for all $k \in \mathbb{N}$. The practical consequence of this fact is that a proper diagnostic of convergence cannot be obtained. As a consequence, many unnecessary iterations may be performed without satisfying the stopping criterion, and the algorithm may stop by excess of iterations without realizing that iterates very close to a solution have been already obtained.

In [3] it was shown that, when we apply the SQP (Newton–Lagrange) [23,36] method to

$$\text{Minimize } x \quad \text{subject to } x^2 = 0, \tag{8}$$

we obtain that the generated primal sequence converges to the solution $x^* = 0$, but the KKT residual $r^k = \nabla f(x^k) + \nabla h(x^k)\lambda^k$ converges to $\frac{1}{3}$.

This state of facts led us to study the behaviour of the Newton–Lagrange method when applied to the KKT system in the case that no constraint qualification holds at the minimizer of the optimization problem and the KKT conditions do not hold. In the following section we will show that the behaviour observed in (8) can be expected for most problems of this class.

3. Newton's method and the fulfillment of approximate KKT conditions

Let us consider the optimization problem

$$\text{Minimize } f(x) \quad \text{subject to } h(x) = 0, \tag{9}$$

where $f, h : \mathbb{R}^1 \to \mathbb{R}^1$ are continuously differentiable functions.

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

If $f(x) = x$ and $h(x) = x^2$, it can be shown that, independently of the initial approximation, if (x_k, λ_k) is generated by the Newton–Lagrange method, one has that the limit of the KKT residual $r_k = f'(x_k) + \lambda_k h'(x_k)$ is always $\frac{1}{3}$ [3]. If $f(x) = x$ and $h(x) = x^3$ then $\lim r_k = \frac{7}{19}$. In this section we will show that a pattern for this behaviour exists.

Our objective in this section is not to show that, in the absence of regularity, the Newton's method may not converge. Of course, this is well known and examples of this situation should be futile. We are interested in cases in which the primal sequence $\{x^k\}$ *does converge* to the true solution but convergence *cannot be detected* by the computation of the KKT residual.

The KKT (with feasibility) conditions for (9) are

$$f'(x) + \lambda h'(x) = 0, \quad h(x) = 0. \tag{10}$$

Applying Newton's method to (10), we obtain the following iteration scheme:

$$\begin{bmatrix} f''(x_k) + \lambda_k h''(x_k) & h'(x_k) \\ h'(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \lambda_{k+1} \end{bmatrix} = - \begin{bmatrix} f'(x_k) \\ h(x_k) \end{bmatrix}, \tag{11}$$

where $d_k = x_{k+1} - x_k$. For simplicity, we will denote $h_k = h(x_k)$. The iterations are well defined under the condition that $h'_k \neq 0$ for all $k = 0, 1, 2, \ldots$ Let us assume that $\lim x_k = x_*$. Therefore, $\lim d_k = 0$ and, from (11), we have that $h(x_*) = 0$. If $h'(x_*) \neq 0$, i.e. x_* is a regular point, then it is a well-known result that $\lim \lambda_k = \lambda_*$ with $f'(x_*) + \lambda_* h'(x_*) = 0$. Let $\nu > 1$ be such that $h^{(j)}(x_*) = 0$ for $j = 1, \ldots, \nu - 1$ but $h^{(\nu)}(x_*) \neq 0$. From Taylor's theorem we can write

$$h_k = c_\nu e_k^\nu + o(e_k^\nu), \tag{12}$$

$$h'_k = \nu c_\nu e_k^{\nu-1} + o(e_k^{\nu-1}), \tag{13}$$

$$h''_k = \nu(\nu - 1) c_\nu e_k^{\nu-2} + o(e_k^{\nu-2}), \tag{14}$$

where $c_\nu = h^{(\nu)}(x_*)/\nu! \neq 0$ and $e_k = x_k - x_*$. Now, from Equation (11)

$$\lambda_{k+1} = -\frac{1}{h'_k}[f'_k + (f''_k + \lambda_k h''_k) d_k] = -\frac{1}{h'_k}\left[f'_k + f''_k d_k - \lambda_k \frac{h''_k h_k}{h'_k}\right]. \tag{15}$$

Thus,

$$\lambda_{k+1} h'_{k+1} = \frac{h'_{k+1}}{h'_k}\left[-f'_k - f''_k d_k + \lambda_k h'_k \frac{h''_k h_k}{(h'_k)^2}\right], \tag{16}$$

and then, defining

$$y_k = \lambda_k h'_k, \quad \alpha_k = \frac{h'_{k+1}}{h'_k}, \quad \beta_k = f'_k + f''_k d_k \quad \text{and} \quad \gamma_k = \frac{h''_k h_k}{(h'_k)^2}, \tag{17}$$

we can establish the following linear iteration:

$$y_{k+1} = \alpha_k[-\beta_k + \gamma_k y_k]. \tag{18}$$

Let us compute the limits of the sequences $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\gamma_k\}$. Clearly, $\lim \beta_k = f'(x_*)$. Now

$$\begin{aligned} \lim \gamma_k &= \lim \frac{[\nu(\nu-1)c_\nu e_k^{\nu-2} + o(e_k^{\nu-2})][c_\nu e_k^\nu + o(e_k^\nu)]}{[\nu c_\nu e_k^{\nu-1} + o(e_k^{\nu-1})]^2} \\ &= \lim \frac{\nu(\nu-1)c_\nu^2}{[\nu c_\nu]^2} = 1 - \frac{1}{\nu} \equiv \gamma_* \end{aligned} \tag{19}$$

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

and

$$\lim \alpha_k = \lim \frac{\nu c_\nu e_{k+1}^{\nu-1} + o(e_{k+1}^{\nu-1})}{\nu c_\nu e_k^{\nu-1} + o(e_k^{\nu-1})} = \lim \left[\frac{e_{k+1}}{e_k}\right]^{\nu-1} = \lim \left[1 + \frac{d_k}{e_k}\right]^{\nu-1}$$
$$= \lim \left[1 - \frac{h_k}{h_k' e_k}\right]^{\nu-1} = \lim \left[1 - \frac{c_\nu e_k^\nu + o(e_k^\nu)}{[\nu c_\nu e_k^{\nu-1} + o(e_k^{\nu-1})]e_k}\right]^{\nu-1}$$
$$= \lim \left[1 - \frac{1}{\nu}\right]^{\nu-1} = \gamma_*^{\nu-1}. \tag{20}$$

Observe that γ_* is the well-known linear convergence rate for Newton's method in the case of a zero with multiplicity ν. Since $\lim |\alpha_k \gamma_k| = \gamma_*^\nu < 1$, the sequence $\{y_k\}$ defined by (18) converges to some y_*. Therefore, it is easy to show that

$$y_* = \frac{\gamma_*^{\nu-1}}{\gamma_*^\nu - 1} f'(x_*), \tag{21}$$

and then, the KKT residual $r_k = f_k' + \lambda_k h_k' = f_k' + y_k$ converges to

$$r_* = f'(x_*) + y_* = \left[1 + \frac{\gamma_*^{\nu-1}}{\gamma_*^\nu - 1}\right] f'(x_*) \equiv \Gamma_\nu f'(x_*). \tag{22}$$

It follows that $r_* = 0$ only if $f'(x_*) = 0$ (x_* is an unconstrained stationary point of f) or $\nu = 1$ ($\Gamma_1 = 0$), which corresponds to regularity of x_*. In particular, we have $\Gamma_2 = \frac{1}{3}$ and $\Gamma_3 = \frac{7}{19}$. Moreover, Γ_ν is an increasing function of ν with $\lim_{\nu\to\infty} \Gamma_\nu = (e-2)/(e-1) \approx 0.418$, where e is Euler's number.

Therefore, we proved that, when we employ the Newton–Lagrange method for solving (9), even in the situation of convergence to a solution of the problem, the approximate KKT conditions fail to be satisfied (thus, the corresponding optimality stopping criterion never takes place) except in the trivial cases in which the point is regular or the limit is an unconstrained stationary point. As a consequence, there are strong reasons to introduce more stable forms of Newton's method for solving this problem.

4. Newton-like methods that satisfy approximate KKT conditions

Let us consider the problem

$$\text{Minimize } f(x) \quad \text{subject to } h(x) = 0, \quad x \geq 0, \tag{23}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth. With the employment of slack variables every nonlinear programming problem can be reduced to the form (23). Given $\rho > 0$, $\mu > 0$, we define the penalty–barrier function $\Phi_{\rho,\mu}$ by

$$\Phi_{\rho,\mu}(x) = f(x) + \frac{\rho}{2}\|h(x)\|^2 - \mu \sum_{i=1}^n \log(x_i), \tag{24}$$

for all $x \in \mathbb{R}^n_{++}$. If $x \notin \mathbb{R}_{++}$, we define $\Phi_{\rho,\mu}(x) = \infty$. For all $x \in \mathbb{R}^n_{++}$ we have

$$\nabla \Phi_{\rho,\mu}(x) = \nabla f(x) + \rho \nabla h(x) h(x) - \mu X^{-1}\mathbf{e}. \tag{25}$$

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

Thus, the optimality condition $\nabla \Phi_{\rho,\mu}(x) = 0$ is

$$\nabla f(x) + \rho \nabla h(x) h(x) - \mu X^{-1}\mathbf{e} = 0. \tag{26}$$

Condition (26) is obviously equivalent to

$$\nabla f(x) + \nabla h(x)\lambda - z = 0, \quad h(x) - \frac{\lambda}{\rho} = 0, \quad XZ\mathbf{e} - \mu\mathbf{e} = 0. \tag{27}$$

The application of Newton's method to the nonlinear system (27) yields

$$\begin{bmatrix} H_k + D_k & \nabla h(x^k) & -I \\ h'(x^k) & -\dfrac{I}{\rho} & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{bmatrix} x - x^k \\ \lambda - \lambda^k \\ z - z^k \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) - \nabla h(x^k)\lambda^k + z^k \\ -h(x^k) + \dfrac{\lambda^k}{\rho} \\ -X_k Z_k \mathbf{e} + \mu \mathbf{e} \end{bmatrix}, \tag{28}$$

where

$$H_k \equiv \nabla^2 L(x^k, \lambda^k) = \nabla^2 f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k), \tag{29}$$

and D_k is a correction matrix that, as we will see later, may be used to guarantee good descent properties of the direction generated by (28).

The linear system (28) is equivalent to

$$\begin{bmatrix} H_k + D_k & \nabla h(x^k) & -I \\ h'(x^k) & -\dfrac{I}{\rho} & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{bmatrix} d^k \\ \lambda \\ z \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ -h(x^k) \\ \mu \mathbf{e} \end{bmatrix}, \tag{30}$$

where $d^k = x - x^k$. By the last block of equations of (30) we have that

$$z = X_k^{-1}(\mu \mathbf{e} - Z_k d^k). \tag{31}$$

Replacing (31) in the first block of (30), this system becomes

$$\begin{bmatrix} H_k + D_k + X_k^{-1} Z_k & \nabla h(x^k) \\ h'(x^k) & -\dfrac{I}{\rho} \end{bmatrix} \begin{bmatrix} d^k \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) + \mu X_k^{-1}\mathbf{e} \\ -h(x^k) \end{bmatrix}. \tag{32}$$

Now, we want to verify whether the direction d^k provided by (32) is a descent direction for $\Phi_{\rho,\mu}$. From (32), we have $\lambda = \rho[h(x^k) + h'(x^k)d^k]$. Therefore, by the first block of (32)

$$(H_k + D_k + X_k^{-1} Z_k) d^k + \nabla h(x^k)\rho[h(x^k) + \nabla h(x^k)^T d^k] = -\nabla f(x^k) + \mu X_k^{-1}\mathbf{e}, \tag{33}$$

and thus,

$$\begin{aligned} [H_k + D_k + X_k^{-1} Z_k + \rho \nabla h(x^k) \nabla h(x^k)^{\mathrm{T}}] d^k &= -\nabla f(x^k) + \mu X_k^{-1}\mathbf{e} - \rho \nabla h(x^k) h(x^k) \\ &= -\nabla \Phi_{\rho,\mu}(x^k). \end{aligned} \tag{34}$$

This means that the fact that d^k is a descent direction for $\Phi_{\rho,\mu}$ is linked to the positive definiteness of $[H_k + D_k + X_k^{-1} Z_k + \rho \nabla h(x^k) \nabla h(x^k)^{\mathrm{T}}]$. On the other hand, the positive definiteness of this matrix is connected with the inertia of the matrix of the system (32). If this matrix does not have

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

the correct inertia, we can add a suitable diagonal matrix to the upper-left block in order to obtain the desired descent property.

This discussion motivates the definition of the following algorithm, the description of which follows the essential lines of [41], except that external penalties are used instead of filters.

ALGORITHM 1 *Let* $(x^0, \lambda^0, z^0) \in \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$ *be the initial approximation to the primal–dual solution. Let* $\mu_0 > 0$ *and* $\rho_0 > 0$ *be the initial barrier and penalty parameters, respectively. We initialize safeguarding parameters* $\beta_0^{\max} \geq 1 \geq \beta_0^{\min} > 0$ *and* $\theta_0 > 0$. *The required accuracy for convergence will be given by* $\varepsilon > 0$.

Step 1. (*Initialization*) *Initialize the outer iteration counter* $j \leftarrow 0$ *and the global iteration counter* $k \leftarrow 0$. *Set* $\tau_0 = \max\{0.99, 1 - \mu_0\}$.

Step 2. (*Check Optimality of Subproblem*) *If*

$$\|\nabla \Phi_{\rho_j,\mu_j}(x^k)\|_\infty \leq 10 \max\left\{\mu_j, \frac{1}{\rho_j}\right\}, \tag{35}$$

replace

$$\lambda^k = \rho_k h(x^k), \quad z^k = \mu_j X_k^{-1}\boldsymbol{e}, \tag{36}$$

and go to Step 3. *Else*, *go to Step* 5.

Step 3. (*Check approximate KKT*) *Define the KKT residual*

$$r^k = \nabla f(x^k) + \nabla h(x^k)\lambda^k - z^k. \tag{37}$$

If

$$\max\{\|r^k\|_\infty, \|h(x^k)\|_\infty, \|X_k Z_k \boldsymbol{e}\|_\infty\} \leq \varepsilon, \tag{38}$$

stop.

Step 4. (*Define new inner parameters*) *Compute*

$$\mu_{j+1} = \min\left\{\frac{\mu_j}{5}, \mu_j^{1.5}\right\}, \quad \rho_{j+1} = \max\{5\rho_j, \rho_j^{1.5}\}, \tag{39}$$

$\theta_{j+1} \in (0, \theta_j]$, $\beta_{j+1}^{\min} \in (0, \beta_j^{\min}]$, $\beta_{j+1}^{\max} \geq \beta_j^{\max}$ *and*

$$\tau_{j+1} = \max\{0.99, 1 - \mu_{j+1}\}. \tag{40}$$

Set $j \leftarrow j + 1$.

Step 5. (*Newton Step*) *Solve the linear system*

$$\begin{bmatrix} H_k + X_k^{-1}Z_k + D_k & \nabla h(x^k) \\ h'(x^k) & -\dfrac{I}{\rho_j} \end{bmatrix} \begin{bmatrix} d^k \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) + \mu_j X_k^{-1}\boldsymbol{e} \\ -h(x^k) \end{bmatrix}, \tag{41}$$

where $D_k = c_k I$ $(c_k > 0)$ *is obtained by an inertia-correction procedure* (*see* [41]) *that guarantees that* d^k *is a descent direction for* Φ_{ρ_j,μ_j}. *Compute*

$$z = X_k^{-1}(\mu_j \boldsymbol{e} - Z_k d^k). \tag{42}$$

Step 6. (*Check adequacy of the search direction*) *If* $\|d^k\| = 0$, *replace*

$$\lambda^k \leftarrow \rho_j h(x^k), \quad z^k \leftarrow X_k^{-1}\mu_j \boldsymbol{e}, \tag{43}$$

and go to Step 2. *Note that test* (35) *can be in fact skipped because it will be automatically satisfied in this case.*

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

If $\nabla\Phi_{\rho_j,\mu_j}(x^k)^{\mathrm{T}}d^k \geq -\theta_j\|\nabla\Phi_{\rho_j,\mu_j}(x^k)\| \cdot \|d^k\|$, *replace*

$$d^k \leftarrow -\nabla\Phi_{\rho_j,\mu_j}(x^k) \tag{44}$$

and go to Step 7. *If* $\|d^k\| > \beta_j^{\max}\|\nabla\Phi_{\rho_j,\mu_j}(x^k)\|$, *replace*

$$d^k \leftarrow d^k\beta_j^{\max}\|\nabla\Phi_{\rho_j,\mu_j}(x^k)\|/\|d^k\|. \tag{45}$$

If $\|d^k\| < \beta_j^{\min}\|\nabla\Phi_{\rho_j,\mu_j}(x^k)\|$, *replace*

$$d^k \leftarrow d^k\beta_j^{\min}\|\nabla\Phi_{\rho_j,\mu_j}(x^k)\|/\|d^k\|. \tag{46}$$

Step 7. (*Backtracking*) *Obtain, by means of backtracking, starting with the trial step* $t = 1$, *a step* $t_k > 0$ *such that*

$$\Phi_{\rho_j,\mu_j}(x^k + t_kd^k) \leq \Phi_{\rho_j,\mu_j}(x^k) + 0.001t_k\nabla\Phi_{\rho_j,\mu_j}(x^k)^{\mathrm{T}}d^k. \tag{47}$$

Define $x^{k+1} = x^k + t_kd^k$, $\lambda^{k+1} = \lambda^k + t_k(\lambda - \lambda^k)$, $z^{k+1} = z^k + t_k^z(z - z^k)$, *where*

$$t_k^z = \max\{t \in (0, 1] \mid z^k + t(z - z^k) \geq (1 - \tau_j)z^k\}. \tag{48}$$

Set $k \leftarrow k + 1$ *and go to Step* 2.

Many variations for the inertia-correction process may be employed. Modifications that preserve order of convergence in the sense of [20] deserve to be analysed. The theoretical results considered here are not affected by different choices of D_k. The employment of (44)–(46) requires some additional explanation. Although the Newton direction computed at Step 5 is necessarily a descent direction for the merit function, this direction might not satisfy the sufficient requirements for guaranteeing global convergence of the (essentially unconstrained) subproblem. Such requirements (angle condition and proportionality with respect to the gradient) are tested at Step 6 and the direction is corrected in the case that at least one of them does not hold.

LEMMA 4.1 *Algorithm 1 is well defined and generates iterates* $x^k \in \mathbb{R}^n_{++}$, $\lambda^k \in \mathbb{R}^m$ *and* $z^k \in \mathbb{R}^n_{++}$.

Proof Recall that $x^0 \in \mathbb{R}^n_{++}$. We want to show that, if the stopping criterion (38) does not hold at iteration k, then the next iterate $x^{k+1} \in \mathbb{R}^n_{++}$ is computed in finite time. This is trivial when $\|d^k\| = 0$ because $x^{k+1} = x^k$ in that case. Thus, we need to prove that Step 7 (backtracking) is successful in the process of finding t_k that fulfills (47). Using an inductive argument we may assume that $x^k \in \mathbb{R}^n_{++}$. Then, for $t > 0$ small enough, we have that $x^k + td^k \in \mathbb{R}^n_{++}$ and $\Phi_{\rho_j,\mu_j}(x^k + td^k) < \infty$. So, (47) represents a standard Armijo descent condition for the function Φ_{ρ_j,μ_j}. Its fulfillment for small t_k follows from well-known arguments of unconstrained optimization [36, Chapter 3]. Finally, by construction, $\lambda^k \in \mathbb{R}^m$ and $z^k \in \mathbb{R}^n_{++}$ for all k. ■

From now on we assume that the sequence $\{x^k\}$ generated by Algorithm 1 is bounded. (This assumption is trivially verified when one replaces the constraints $x \geq 0$ by $\ell \leq x \leq u$ and we perform the corresponding modifications on the algorithm.) By construction, $x^k \in \mathbb{R}^n_{++}$ for all $k \in \mathbb{N}$. Employing traditional arguments of unconstrained optimization, we will prove that, for all $j = 0, 1, 2, \ldots$, there exists $k = k(j)$ such that (35) holds.

LEMMA 4.2 *For all* $j = 0, 1, 2, \ldots$ *there exists* $k = k(j)$ *such that the criterion* (35) *holds.*

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

Proof Assume, by contradiction, that there exists j such that (35) fails to hold for all k. Then, there exists k_0 such that, for all $k \geq k_0$, the test (35) is performed and fails to hold. By the boundedness of the whole sequence, there exists a convergent subsequence $\{x^k\}_{k\in K}$ contained in $\{x^k\}_{k\geq k_0}$, where K is an infinite subset of $\mathbb{N}$. Since $\Phi_{\rho_j,\mu_j}(x^k) \leq \Phi_{\rho_j,\mu_j}(x^{k_0})$ for all $k \geq k_0$, the limit point x^* of $\{x^k\}_{k\in K}$ belongs to $\mathbb{R}^n_{++}$. (Otherwise the subsequence $\{\Phi_{\rho_j,\mu_j}(x^k)\}_{k\in K}$ would tend to infinity.) If the sequence $\{t_k\}_{k\in K}$ is bounded away from zero, (46) and (44) imply that $\lim_{k\in K} \Phi_{\rho_j,\mu_j}(x^k) = -\infty$, which is impossible. Therefore $\lim_{k\in K} t_k = 0$. This implies that there exists $\bar{t}_k > t_k$ such that $\lim_{k\in K} \bar{t}_k = 0$ and

$$\Phi_{\rho_j,\mu_j}(x^k + \bar{t}_k d^k) > \Phi_{\rho_j,\mu_j}(x^k) + 0.001\bar{t}_k \nabla\Phi_{\rho_j,\mu_j}(x^k)^T d^k \tag{49}$$

for all $k \in K$. Now, since $x^* \in \mathbb{R}^n_{++}$ and $\{\|d^k\|\}_{k\in K}$ is bounded, it turns out that $x^k + \bar{t}_k d^k \in \mathbb{R}^n_{++}$ for $k \in K$ large enough. Therefore, using the mean value theorem in the inequality

$$\Phi_{\rho_j,\mu_j}(x^k + \bar{t}_k d^k) - \Phi_{\rho_j,\mu_j}(x^k) \geq 0.001\bar{t}_k \nabla\Phi_{\rho_j,\mu_j}(x^k)^{\mathrm{T}} d^k, \tag{50}$$

invoking (44) and (46), and taking limits for $k \in K$, we finally find that $\|\nabla\Phi_{\rho_j,\mu_j}(x^*)\| = 0$. This means that (35) holds for k large enough. ■

In the following lemma we prove that every limit point of the sequence $\{x^{k(j)}\}$ is a stationary point of the squared infeasibility norm $\|h(x)\|^2$.

LEMMA 4.3 *Assume that x^* is a limit point of $\{x^{k(j)}\}$. Then x^* is a stationary point of $\|h(x)\|^2$ subject to $x \geq 0$.*

Proof By Lemma 4.2 we have that, for all $j = 0, 1, 2, \ldots$, there exists $k(j)$ such that

$$\|\nabla\Phi_{\rho_j,\mu_j}(x^{k(j)})\| \leq 10 \max\left\{\mu_j, \frac{1}{\rho_j}\right\}. \tag{51}$$

By (39), $\lim \mu_j = 0$ and $\lim \rho_j = \infty$, and then, by (25)

$$\lim[\nabla f(x^{k(j)}) + \rho_j \nabla h(x^{k(j)})h(x^{k(j)}) - \mu_j X^{-1}_{k(j)}\mathbf{e}] = 0. \tag{52}$$

Consequently,

$$\lim\left[\frac{1}{\rho_j}\nabla f(x^{k(j)}) + \nabla h(x^{k(j)})h(x^{k(j)}) - \frac{\mu_j}{\rho_j} X^{-1}_{k(j)}\mathbf{e}\right] = 0. \tag{53}$$

Suppose that $x^*_i > 0$. Since $\lim \mu_j = 0$ and $\lim \rho_j = \infty$, we obtain, by (53), that

$$[\nabla h(x^*)h(x^*)]_i = 0. \tag{54}$$

Consider now the case in which $x^*_i = 0$. Since $\lim \nabla f(x^{k(j)})/\rho_j = 0$ and $[\mu_j X^{-1}_{k(j)}\mathbf{e}]_i > 0$ for all $j = 0, 1, 2, \ldots$, it follows from (53) that

$$[\nabla h(x^*)h(x^*)]_i \geq 0. \tag{55}$$

Since $\nabla\|h(x)\|^2 = 2\nabla h(x)h(x)$, (54) and (55) imply the desired result. ■

Lemma 4.3 says that, in terms of feasibility, Algorithm 1 enjoys the best property that can be expected for an affordable method: to find points that, being stationary for $\|h(x)\|^2$, are probable local minimizers of infeasibility. No algorithm can guarantee feasibility without further assumptions. In the following theorem we prove that, in the case that the algorithm finds almost feasible

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

points, the AKKT stopping criterion takes place. Note that this desirable property does not hold for the Newton–Lagrange method, as shown in Section 4.

THEOREM 4.4 *Assume that for infinitely many indices j we have $\|h(x^{k(j)})\|_\infty \leq \varepsilon$. Then, there exists k such that the stopping criterion (38) holds.*

Proof By Lemma 4.2, for all $j = 0, 1, 2, \ldots$ we have that

$$\|\nabla \Phi_{\rho_j,\mu_j}(x^{k(j)})\|_\infty \leq 10 \max\left\{\mu_j, \frac{1}{\rho_j}\right\}. \tag{56}$$

From (39), we have that $\lim \mu_j \to 0$ and $\lim \rho_j \to \infty$, therefore, for j large enough

$$\|\nabla \Phi_{\rho_j,\mu_j}(x^{k(j)})\|_\infty \leq \varepsilon. \tag{57}$$

Therefore, by (25) and (36)

$$\|\nabla f(x^{k(j)}) + \nabla h(x^{k(j)})\lambda^{k(j)} - z^{k(j)}\|_\infty \leq \varepsilon. \tag{58}$$

Therefore, the thesis follows from (35) and (36). ■

Theorem 4.4 says that, even in the case that KKT conditions do not hold at the solution of the problem, if the algorithm converges to a feasible point, the stopping convergence criterion (38) is satisfied. Observe that (38) corresponds to the AKKT optimality condition defined in [4], but it is close to the stronger CAKKT condition of [3], since, in (38), the product form of the complementarity is used instead of the weaker 'minimum' form (5).

5. Conclusions

The negative results presented in Section 3 of this paper indicate that the Newton–Lagrange method is essentially inadequate for dealing with optimization problems in which KKT conditions do not hold. In this case, one wishes to 'solve' a nonlinear system that has no solutions at all. Consequently, the sense in which these systems should be 'solved' corresponds to approximate KKT criteria [3,4]. Practical optimization algorithms always test (approximately) the KKT conditions for reporting 'convergence'. This test reflects the feeling that, in some sense, the KKT conditions should hold at the solution, even if Lagrange multipliers tend to infinity. Such feeling is supported by theoretical results [3,4], since we know that local minimizers of constrained optimization problems really satisfy sequential optimality conditions. As a consequence, the analysis of algorithms from the point of view of approximate fulfillment of KKT conditions, without regularity assumptions, seems to be a theoretical issue with practical relevance.

Here we proved that the penalty–barrier Newtonian algorithm generates sequences that satisfy an approximate KKT condition. The same conclusion probably holds for the null-space primal–dual algorithm of Liu and Yuan [30] and the exact augmented Lagrangian methods of [19] and [16]. According to Section 2 of this paper, the conclusion seems to be negative for algorithms strongly based in the 'pure' form of SQP, although the 'suitability' of stabilized sequential quadratic programming methods [21,22,29,42,43] and other stabilized Newton methods [8,15,34] with respect to approximate KKT conditions remains to be an open problem. Approximate KKT conditions should also be studied from the point of view of image space analysis [26].

Let us finish this paper emphasizing that our theoretical results are inspired by the necessity of explaining practical aspects of implementable optimization methods. Constrained optimization methods stop when approximate KKT conditions are verified. The question that we wish to answer in every particular case is: does this algorithm reliably stop? The answer is easy when KKT conditions hold, but, what about the case in which Lagrange multipliers do not exist? A lot of research can be expected analysing well-established practical methods with respect to this property.

Funding

Work supported by the Brazilian research agencies CAPES, CNPQ, FAPESP (grant 06/53768–0), PRONEX—Optimization (PRONEX—CNPq / FAPERJ E–26 /171.164/2003—APQ1) and CEPID—Center on Industrial Mathematics (FAPESP).

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013

References

[1] R. Andreani, J.M. Martínez, and M.L. Schuverdt, *On the relation between the constant positive linear dependence condition and quasinormality constraint qualification*, J. Optim. Theory Appl. 125 (2005), pp. 473–485.
[2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim. 18 (2007), pp. 1286–1309.
[3] R. Andreani, J.M. Martínez, and B.F. Svaiter, *A new sequential optimality condition for constrained optimization and algorithmic consequences*, SIAM J. Optim. 20 (2010), pp. 3533–3554.
[4] R. Andreani, G. Haeser, and J.M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization 60 (2011), pp. 627–641.
[5] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva, *Two new weak constraint qualifications and applications*, SIAM J. Optim. 22 (2012), pp. 1109–1135.
[6] 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 (2012), pp. 255–273.
[7] A.V. Arutyunov, *Optimality Conditions—Abnormal and Degenerate Problems*, Kluwer, Dordrecht, 2000.
[8] M.C. Bartholomew-Biggs, *Recursive quadratic programming methods based on the augmented Lagrangian*, Math. Program. Study 31 (1987), pp. 21–41.
[9] H.Y. Benson, D.F. Shanno, and R.J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming – Filter methods and merit functions*, Comput. Optim. Appl. 23 (2002), pp. 257–272.
[10] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
[11] R.H. Bielschowsky and F.A.M. Gomes, *Dynamic control of infeasibility in equality constrained optimization*, SIAM J. Optim. 19 (2008), pp. 1299–1325.
[12] R.H. Byrd, J.Ch. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Program. 89 (2000), pp. 149–185.
[13] R.H. Byrd, J. Nocedal, and R.A. Waltz, *KNITRO—An integrated package for nonlinear optimization*, in *Large-Scale Nonlinear Optimization*, G. Di Pillo and M. Roma, eds., Springer, New York, 2006, pp. 35–59.
[14] A.R. Conn, N.I.M. Gould, and Ph.L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, PA, 2000.
[15] L. Contesse-Becker, *Extended convergence results for the method of multipliers for non-strictly binding inequality constraints*, J. Optim. Theory Appl. 79 (1993), pp. 273–310.
[16] G. Di Pillo and S. Lucidi, *An augmented Lagrangian function with improved exactness properties*, SIAM J. Optim. 12 (2001), pp. 376–406.
[17] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi, *An exact augmented Lagrangian function for nonlinear programming with two-sided constraints*, Comput. Optim. Appl. 25 (2003), pp. 57–83.
[18] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi, *A truncated Newton method in an augmented Lagrangian framework for nonlinear programming*, Comput. Optim. Appl. 45 (2010), pp. 311–352.
[19] G. Di Pillo, G. Liuzzi, and S. Lucidi, *An exact penalty-Lagrangian approach for large-scale nonlinear programming*, Optimization 60 (2011), pp. 223–252.
[20] J.Y. Fan and Y.X. Yuan, *On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption*, Computing 34 (2005), pp. 23–39.
[21] D. Fernández, *A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems*, Math. Program. (2011), pp. 199–223. doi:10.1007/s10107-011-0493-8
[22] D. Fernández and M. Solodov, *Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems*, Math. Program. 125 (2010), pp. 47–73.
[23] R. Fletcher, *Practical Methods of Optimization*, Academic Press, London, 1987.
[24] R. Fletcher, S. Leyffer, and Ph.L. Toint, *On the global convergence of a filter-SQP algorithm*, SIAM J. Optim. 13 (2002), pp. 44–59.
[25] R. Fletcher, N.I.M. Gould, S. Leyffer, Ph.L. Toint, and A. Wächter, *Global convergence of trust-region SQP-filter algorithms for general nonlinear programming*, SIAM J. Optim. 13 (2002), pp. 635–659.
[26] F. Giannessi, *Separation of Sets and Optimality Conditions*, Springer, New York, 2005.

[27] N.I.M. Gould and Ph.L. Toint, *Nonlinear programming without a penalty function or a filter*, Math. Program. 122 (2010), pp. 155–196.

[28] S. Gratton, M. Mouffe, and Ph.L. Toint, *Stopping rules and backward error analysis for bound-constrained optimization*, Numer. Math. 119 (2011), pp. 163–187.

[29] A. Izmailov and M. Solodov, *Stabilized SQP revisited*, Math. Program. (2010). pp. 93–120. doi:10.1007/s10107-010-0413-3

[30] X.W. Liu and Y.X. Yuan, *A null-space primal–dual interior-point algorithm for nonlinear optimization with nice convergence properties*, Math. Program. 125 (2010), pp. 163–193.

[31] X.W. Liu and Y.X. Yuan, *A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization*, SIAM J. Optim. 21 (2011), pp. 545–571.

[32] L. Luksan, C. Matonoha, and J. Vlcek, *Interior point methods for large-scale nonlinear programming*, Optim. Methods Softw. 20 (2003), pp. 569–582.

[33] L. Luksan, C. Matonoha, and J. Vlcek, *Algorithm 896: LSA: Algorithms for large-scale optimization*, ACM Trans. Math. Softw. 36 (2009), pp. 16:1–16:29.

[34] J.M. Martínez and L.T. Santos, *Some new theoretical results on recursive quadratic programming algorithms*, J. Optim. Theory Appl. 97 (1998), pp. 435–454.

[35] J.M. Martínez and B.F. Svaiter, *A practical optimality condition without constraint qualifications for nonlinear programming*, J. Optim. Theory Appl. 118 (2003), pp. 117–133.

[36] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[37] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM J. Optim. 10 (2000), pp. 963–981.

[38] A. Schiela and A. Guenther, *An interior point algorithm with inexact step computation in function space for state constrained optimal control*, Numer. Math. 119 (2011), pp. 373–407.

[39] C. Shen, W. Xue, and D. Pu, *A filter SQP algorithm without a feasibility restoration phase*, Comput. Appl. Math. 28 (2009), pp. 167–194.

[40] C. Shen, S. Leyffer, and R. Fletcher, *Nonmonotone filter method for nonlinear optimization*, Comput. Optim. Appl. 52 (2012), pp. 583–607.

[41] A. Wächter and L.T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program. 106 (2006), pp. 25–57.

[42] S.J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl. 11 (1998), pp. 253–275.

[43] S.J. Wright, *Modifying SQP for degenerate problems*, SIAM J. Optim. 13 (2002), pp. 470–497.

Downloaded by [UNICAMP], [Lucio T. Santos] at 08:33 10 December 2013