Computer Science Journal of Moldova, vol.**19**, no.3(57), 2011

A Smooth Newton Method for Nonlinear Programming Problems with Inequality Constraints

Vasile Moraru

Abstract

The paper presents a reformulation of the Karush-Kuhn-Tucker (KKT) system associated nonlinear programming problem into an equivalent system of smooth equations. Classical Newton method is applied to solve the system of equations. The superlinear convergence of the primal sequence, generated by proposed method, is proved. The preliminary numerical results with a problems test set are presented.

Keywords: nonlinear programming, KKT conditions, strict complementarity, Newton method, local superlinear convergence.

1 Introduction

We consider the following nonlinear programming problem:

$$\begin{cases} \text{minimize } f(x) \\ \text{subject to} \\ g_i(x) \geq 0, i = 1, 2, \ldots, m, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$, f, $g_i : \mathbb{R}^n \to \mathbb{R}$ are assumed to be twice continuously differentiable. Such problems have proved extremely useful in very many areas of activity, in science, engineering and management [1,2].

Let the *Lagrange function* of problem (1) be defined by

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x),$$

where $\lambda = (\lambda_1, \lambda_2, ..., \lambda_m)^T$ is the Lagrange multiplier vector.

We will denote by x^* any local solution of the problem (1), and by

$$\mathcal{A}(x) = \{i : g_i(x) = 0\}$$

the index set of active constraints at x.

The *Karush-Kuhn-Tucker system* (KKT system for short) associated with (1) is [3,4] :

$$\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \quad (\textit{stationarity}) \tag{2}$$

$$g_i(x) \geq 0, i = 1, 2, \ldots, m \quad (\textit{primal feasibility}) \tag{3}$$

$$\lambda_i \geq 0, i = 1, 2, \ldots, m \quad (\textit{dual feasibility}) \tag{4}$$

$$\lambda_i g_i(x) = 0, i = 1, 2, \ldots, m \quad (\textit{complementarity}) \tag{5}$$

This system is the local equivalent to the problem (1) whenever the inequalities (3) and (4) are satisfied and relations (2) and (5) comply with the conditions of regularity [3–5].

We assume that the following hypotheses hold for any local solution x^*.

Assumption A1. The active constraint gradients $\nabla g_i(x^*), i \in \mathcal{A}(x^*)$ are linearly independent (the assumption is called the *linear independence constraint qualification* (LICQ)).

Assumption A2. *Strict complementarity* holds at x^*, i.e. $\lambda_i^* > 0$ for all $i \in \mathcal{A}(x^*)$.

Assumption A3. The *strong second order sufficient condition* (SSOSC):

$$p^T \nabla_{xx}^2 L(x^*, \lambda^*) p \geq c \|p\|^2, c > 0, \text{for all } \ p \in T(x^*), \tag{6}$$

where $\nabla_{xx}^2 L(x^*, \lambda^*)$ is the Hessian matrix of the Lagrange function of the problem (1) and

$$T(x^*) = \left\{ p \in \mathbb{R}^n : p \neq 0, [\nabla g_i(x^*)]^T p = 0, \ i \in \mathcal{A}(x^*) \right\}.$$

It is well know that, under **Assumptions A1–A3**, x^* is an isolated local minimum and to solve (1) is equivalent to solve the KKT system (2) – (5).

KKT system (2) – (5) establishes necessary conditions [5,6] for solving *finite-dimensional variational inequality*:

$$[\nabla f(x^*)]^T (x - x^*) \geq 0, \quad \text{for all } x \in \Omega, \tag{7}$$

where $\Omega = \{x \in \mathbb{R}^n : g_i(x) \geq 0, i = 1, 2, \ldots, m\}$.

In the particular case when $\Omega = \{x \in \mathbb{R}^n : x_i \geq 0, i = 1, 2, \ldots, n\}$ the variational inequality problem (7) is equivalent to the following *complementarity problem*

$$\nabla f(x) \geq 0, x \geq 0, [\nabla f(x)]^T x = 0.$$

Furthermore, the KKT system (2) − (5) can be written as a *mixed complementarity problem* (MCP) [8]:

$$[F(z^*)]^T (z - z^*) \geq 0, \quad \text{for all} \quad z \in B,$$

where $z^T = \left(x^T, \lambda^T\right), B = \{z \in \mathbb{R}^n \times \mathbb{R}^m : \lambda \geq 0\}$ and

$$[F(z)]^T = \left([\nabla_x L(z)]^T, g_1(x), g_2(x), \ldots, g_m(x)\right).$$

The first major achievements obtained in the constrained optimization referred to the KKT systems. This has brought to the development of methods of optimization, a very active area with remarkable results [1–7]. Conditions (5), also called complementarity conditions, raise some difficult problems to solve directly the system of equations and inequalities (2)–(5). Depending on how KKT system (2)–(5) is used (i.e. how Lagrange multipliers are calculated and how the conditions of complementarity are ensured), some methods that can be used successfully in problems with medium or even large number of variables have been developed [9]:

- active-set methods [5,7,9,10],

- barrier/penalty and augmented Lagrangian methods [3,7, 10, 11, 12],
- sequential linear and quadratic programming methods [6,7,13,14],
- interior-point and trust region methods [15–17].

Another way of solving KKT system is to use procedures based on the complementary functions [18,19]. The relationship (3) − (5) constitutes the essence of nonlinear complementarity problems. The KKT system (2) − (5) may be equivalently reformulated as solving the nonlinear system [20, 21]:

$$\nabla_x L(x, \lambda) = 0, \varphi(\lambda_i, g_i(x)) = 0, i = 1, 2, \ldots, m, \tag{8}$$

where φ is any NCP function. A function $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is called NCP function if the set of solutions of equation $\varphi(a, b)$ coincides with the set:

$$M = \{a, b \in \mathbb{R} : ab = 0, a \geq 0, b \geq 0\}.$$

Classical examples of NCP functions are the min function

$$\varphi(a, b) = \min(a, b)$$

and the Fischer-Burmeister function [22]

$$\varphi(a, b) = \sqrt{a^2 + b^2} - a - b.$$

Other functions of complementarity with applications in optimization can be found in the works [23, 24]. Most of them are nondifferentiable at the point $(0, 0)$. There are smooth functions of complementarity, for example, the function [23, 30]

$$\varphi(a, b) = 2ab - [\min(0, a + b)]^2.$$

This theme in optimization problems still presents challenges, which are based mainly on the application of the Newton method for systems of equations equivalent to system:

$$\nabla_x L(x,\lambda) = 0, g_i(x) = 0, i \in \mathcal{A}(x^*). \tag{9}$$

Relations (3)−(5) can be reformulated as the system of equalities by introducing auxiliary variables $y_1, y_2, ..., y_m$, so that (see [25,26,31,32]):

$$\begin{cases} [-\min(0,y_i)]^k - g_i(x) = 0, \\ [\max(0,y_i)]^k - \lambda_i = 0, \\ i = 1,2,\ldots,m, \end{cases} \tag{10}$$

where $k \geq 1$. In [31,32] it is considered $k = 1$, in [25] and [26] take $k = 2$, respectively $k = 3$.

Emphasize the role of regularity condition **A2**. Only in this case we can guarantee that in the vicinity of (x^*, λ^*) relations (3) − (5) are equivalent to the system of equations:

$$\begin{aligned} g_i(x) &= 0, \quad i \in \mathcal{A}(x^*), \\ \lambda_i &= 0, \quad i \notin \mathcal{A}(x^*). \end{aligned}$$

In other words, the **Assumption A2**, the KKT system (2) − (5) is locally equivalent to system of equations (9). Both the procedure (8) and procedure (10) do not require the explicit identification of the active constraints $\mathcal{A}(x^*)$.

In the present paper a KKT system transformation in a system of smooth equations, which can be solved by classical Newton method is considered. The paper is organized as follows. In Section 2 we define the functions that ensure the nonsingularity of Jacobian at a solution. Section 3 presents the algorithm of Newton method. Superlinear convergence in terms of primal variables is proved in Section 4. Some numerical results are given in Section 5 and conclusion is drawn in Section 6.

Throughout this paper, $\mathbb{R}^n$ denotes the space of n-dimensional real column vectors and the superscript "T " denotes transpose. For convenience, we use (x, λ, y) to denote the column vector $\left(x^T, \lambda^T, y^T\right)^T$.

Given two vectors $x, y \in \mathbb{R}^n$, $x^T y$ denotes the Euclidian scalar product and $\|\bullet\|$ denotes the Euclidian vector norm. The identity matrix is denoted by I. For any $\alpha > 0, \beta > 0, \alpha = o(\beta)$ (respectively $\alpha = O(\beta)$) means α/β tends to zero (respectively α/β is uniformly bounded) as $\beta \to 0$.

2 Smoothing Reformulation of the Karush-Kuhn-Tucker System

We define two functions $u, v : \mathbb{R} \to \mathbb{R}_+$ with the following properties:

P1.

$$u(x) = \begin{cases} = 0, & \text{for all } x \leq 0, \\ > 0, & \text{for all } x > 0, \end{cases}$$
$$v(x) = \begin{cases} > 0, & \text{for all } x < 0, \\ = 0, & \text{for all } x \geq 0. \end{cases}$$

P2. $u(x)$ and $v(x)$ is at least twice continuously differentiable on $\mathbb{R}$.

P3. $u(x) = 0$ and $v(x) = 0$ if and only if $u'(x) = 0$, respectively $v'(x) = 0$ for all $x \neq 0$.

The functions $u(x)$ and $v(x)$ so defined form a complementarity pair in the sense that the two functions are complementary to one another (if one is zero at a point, then the other is necessarily nonzero at that point):

$$\begin{cases} u(x) = 0 \Longleftrightarrow v(x) > 0, \\ v(x) = 0 \Longleftrightarrow u(x) > 0, \end{cases} \tag{11}$$

i.e. $u(x) \times v(x) = 0$ for all $x \in \mathbb{R}$.

The following can serve as an example of functions $u(x)$ and $v(x)$ that satisfy properties **P1 - P3**:

$$\begin{cases} u(x) = x^{\beta} \max(0, x) = \frac{1}{2}\left(x^{\beta+1} + |x|\, x^{\beta}\right), \\ v(x) = (-1)^{\gamma+1} x^{\gamma} \min(0, x) = \frac{(-1)^{\gamma+1}}{2}\left(x^{\gamma+1} - |x|\, x^{\gamma}\right), \end{cases} \tag{12}$$

where $\beta \geq 2, \gamma \geq 2$ are any fixed parameters.

The functions $u(x)$ and $v(x)$ as defined by the formulas (12) are continuously differentiable:

$$\begin{aligned} u'(x) &= \frac{\beta+1}{2}(x^{\beta} + |x|\, x^{\beta-1}), \\ u''(x) &= \frac{\beta(\beta+1)}{2}(x^{\beta-1} + |x|\, x^{\beta-2}), \end{aligned}$$

$$\begin{aligned} v'(x) &= \frac{(-1)^{\gamma+1}(\gamma+1)}{2}\left(x^{\gamma} + |x|\, x^{\gamma-1}\right), \\ v''(x) &= \frac{(-1)^{\gamma+1}\gamma(\gamma+1)}{2}\left(x^{\gamma-1} + |x|\, x^{\gamma-2}\right). \end{aligned}$$

By entering the auxiliary variables $y_1, y_2, \ldots, y_m$ the KKT system (2) − (5) may be transformed into an equivalent system of smooth nonlinear equations:

$$\left\{ \begin{array}{c} \nabla_x L(x,\lambda) = 0, \\ u(y_1) - g_1(x) = 0, \\ \vdots \\ u(y_m) - g_m(x) = 0, \\ v(y_1) - \lambda_1 = 0, \\ \vdots \\ v(y_m) - \lambda_m = 0. \end{array} \right. \tag{13}$$

It is easily found that $\lambda_i > 0$ for all $i \in \mathcal{A}(x^*)$. Indeed, according to (11) if $g_i(x) = 0$, i.e. $u(y_i) = 0$, then $\lambda_i = v(y_i) > 0$. Therefore $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ solves KKT system (2) − (5) or system (9) if and only if $(x^*, \lambda^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ solves the system of equations (13).

For any $y = (y_1, y_2, \ldots, y_m)^T \in \mathbb{R}^m$ define

$$U(y) = (u(y_1), u(y_2), \ldots, u(y_m))^T,$$

$$V(y) = (v(y_1), v(y_2), \ldots, v(y_m))^T,$$

$$U'(y) = diag\left(u'(y_1), u'(y_2), \ldots, u'(y_m)\right),$$

$$V'(y) = diag\left(v'(y_1), v'(y_2), \ldots, v'(y_m)\right).$$

Matrices $U'(y)$ and $V'(y)$ are diagonal matrices of dimension $m \times m$ with elements $u'(y_i)$, respectively $v'(y_i)$, $i = 1, 2, \ldots, m$.

We denote the Jacobian matrix of a mapping

$$G(x) = (g_1(x), g_2(x), \ldots, g_m(x))^T : \mathbb{R}^n \to \mathbb{R}^m$$

by $G'(x)$:

$$G'(x) = \begin{pmatrix} [\nabla g_1(x)]^T \\ [\nabla g_2(x)]^T \\ \vdots \\ [\nabla g_m(x)]^T \end{pmatrix}$$

Let $F'(x, \lambda, y)$ denote the Jacobian of

$$F(x, \lambda, y) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ u(y_1) - g_1(x) \\ \vdots \\ u(y_m) - g_m(x) \\ v(y_1) - \lambda_1 \\ \vdots \\ v(y_m) - \lambda_m \end{pmatrix}.$$

Then the Jacobian matrix $F'(x, \lambda, y)$ has the form

$$F'(x, \lambda, y) = \begin{pmatrix} \nabla^2_{xx}(Lx, \lambda) & -[G'(x)]^T & O_{n\times m} \\ -G'(x) & O_{m\times n} & U'(y) \\ O_{m\times n} & -I_m & V'(y) \end{pmatrix},$$

where I_m is the identity matrix of order m and O represents the null matrix, with subscripts indicating their dimensions.

The matrix $F'(x, \lambda, y)$ has order $(n+2m) \times (n+2m)$.

The next theorem is true:

Theorem 1. ***Under the Assumption A1-A3 for*** $z^* = (x^*, \lambda^*, y^*)$ ***Jacobian matrix*** $F'(z^*)$ ***is nonsingular.***

Proof. Let $d = (p, q, r) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ satisfy $F'(z^*)\, d = 0$. Then

$$\nabla_{xx}^2 L(x^*, \lambda^*)\, p - \left[G'(x^*)\right]^T q = 0, \tag{14}$$

$$-G'(x^*)\, p + U'(y^*)\, r = 0, \tag{15}$$

$$-q + V'(y^*)\, r = 0. \tag{16}$$

By multiplying equation (14) on the left side by p^T we have

$$p^T \nabla_{xx}^2 L(x^*, \lambda^*)\, p = p^T \left[G'(x^*)\right]^T q = q^T \left[G'(x^*)\right]^T p. \tag{17}$$

There is at least an index k such that $u(y_k^*) = 0$ (because otherwise the $g_i(x^*) > 0$ for any $i = 1, 2, \ldots, m$, i.e. x^* is a point of unconstrained minimum). As $g_k(x^*) = u(y_k^*) = 0$ yields $k \in \mathcal{A}(x^*)$. As we have that **Property P3** and $u'(y_k^*) = 0$, which together with (15) gives us

$$[\nabla g_k(x^*)]^T p = 0 \text{ for all } k \in \mathcal{A}(x^*), \tag{18}$$

i.e. $p \in T(x^*)$.

For $s \notin \mathcal{A}(x^*)$ we have $u(y_s^*) \neq 0$ and so $v(y_s^*) = 0$ where the $v'(y_s^*) = 0$; then from the (16) result $q_s = 0$. This together with the relationship (18) gives us $q^T G'(x^*)\, p = 0$. From equation (16) together with the **Assumption A3** we have $p = 0$. So, from equations $(14) - (16)$ one obtains

$$\sum_{i=1}^{m} q_i \nabla g_i \left(x^*\right) = 0, \tag{19}$$

$$u' \left(y_i^*\right) r_i = 0 \ \text{ for all } i = 1, 2, \ldots, m, \tag{20}$$

$$q_i = v' \left(y_i^*\right) r_i \ \text{ for all } i = 1, 2, \ldots, m. \tag{21}$$

We have $\lambda_i^* = 0, v\left(y_i^*\right) = 0, v'\left(y_i^*\right) = 0, u'\left(y_i^*\right) = 0$ for all $i \notin \mathcal{A}\left(x^*\right)$. From (20) and (21) $q_i = 0, r_i = 0$, for all $i \notin \mathcal{A}\left(x^*\right)$. It is true that

$$\sum_{i \in \mathcal{A}(x^*)} q_i \nabla g_i \left(x^*\right) = 0,$$

where, under the **Assumption A1**, $q_i = 0$ for all $i \in \mathcal{A}\left(x^*\right)$. For $i \in \mathcal{A}\left(x^*\right)$ we have $v'\left(y_i^*\right) \neq 0$ and from (21) it results that $r_i = 0$. Thus, the **Assumption A1-A3** implies that $p = 0, q = 0, r = 0$. Therefore, the Jacobian matrix $F'\left(x^*, \lambda^*, y^*\right)$ is nonsingular. ■

3 Local Newton method

The best-known method for solving nonlinear systems of equations is the Newton method [27,28,29]. Newton's method has very attractive theoretical and practical properties, because of its rapid convergence: under the nonsingularity of the Jacobian matrix it will converge locally superlinearly. If in addition, the Jacobian is Lipshitz continuous, then the convergence is quadratic.

Let $\left(x^{(0)}, \lambda^{(0)}, y^{(0)}\right)$ be given sufficiently close to $\left(x^*, \lambda^*, y^*\right)$. Given that $x^{(k)}$, $\lambda^{(k)}$, $y^{(k)}$ is the k-th iterate of the Newton method for the system (13), a new estimate $x^{(k+1)}$, $\lambda^{(k+1)}$, $y^{(k+1)}$ is defined by solving the following linear system:

$$\left\{\begin{array}{c}\nabla^2_{xx}(Lx^{(k)},\lambda^{(k)})(x-x^{(k)})-\left[G'\left(x^{(k)}\right)\right]^T(\lambda-\lambda^{(k)})=\\=-\nabla_x(Lx^{(k)},\lambda^{(k)}),\\-G'(x^{(k)})(x-x^{(k)})+U'(y^{(k)})(y-y^{(k)})=\\=-U(y^{(k)})+G(x^{(k)}),\\-\lambda+\lambda^{(k)}+V'(y^{(k)})(y-y^{(k)})=-V(y^{(k)})+\lambda^{(k)}.\end{array}\right. \tag{22}$$

The system (22) consists of $(n+2m)$ equations and $(n+2m)$ unknowns. From the last equation of the system (22) we have

$$\lambda=V(y^{(k)})+V'(y^{(k)})(y-y^{(k)}). \tag{23}$$

Substituting (23) in the first equation, the system (22) becomes

$$\left\{\begin{array}{c}\nabla^2_{xx}L\left(x^{(k)},\lambda^{(k)}\right)\left(x-x^{(k)}\right)-\left[G'\left(x^{(k)}\right)\right]^T V'\left(y^{(k)}\right)\left(y-y^{(k)}\right)=\\=-\nabla f\left(x^{(k)}\right)+\left[G'\left(x^{(k)}\right)\right]^T V\left(y^{(k)}\right),\\-G'\left(x^{(k)}\right)\left(x-x^{(k)}\right)+U'\left(y^{(k)}\right)\left(y-y^{(k)}\right)=\\=-U\left(y^{(k)}\right)+G\left(x^{(k)}\right).\end{array}\right. \tag{24}$$

The system (24) is from $(n+m)$ equations with $(n+m)$ unknowns.

According to Theorem 1, in the neighborhood of z^* the system of equations (24) admits the unique solution $\left(x^{(k+1)},y^{(k+1)}\right)$.

On the other hand, we see that system (13) can be transformed into an equivalent system which only contains x and y:

$$\left\{\begin{array}{l}\nabla_x L\left(x,V(y)\right)=0,\\-G\left(x\right)+U\left(y\right)=0.\end{array}\right. \tag{25}$$

Linearizing the system (25), we obtain the same system of linear equations (24), except that in the Hessian matrix $\nabla^2_{xx}L\left(x^{(k)},\lambda^{(k)}\right)$ we have $\lambda^{(k)}=V\left(y^{(k)}\right)$.

From (23) it follows that

$$\begin{aligned}\lambda^{(k+1)} &= V\left(y^{(k)}\right) + V'\left(y^{(k)}\right)\left(y^{(k+1)} - y^{(k)}\right) = \\ &= V\left(y^{(k+1)}\right) + o\left(\left\|y^{(k+1)} - y^{(k)}\right\|\right).\end{aligned}$$

So, both local direct linearization of the system (13) and linearization of the system (25) are close enough. The difference may appear only if you are not near the solution.

Thus we can define the following algorithm.

Algorithm 1. ***Local version of Newton's method.***

Step 1. *Let* $x^{(0)} \in \mathbb{R}^n$, $y^{(0)} \in \mathbb{R}^m$, $\lambda^{(0)} = V\left(y^{(0)}\right)$, $\varepsilon > 0$ and $k = 0$.

Step 2. *If*

$$\max\left\{\left\|\left[G'(x^{(k)})\right]^T V(y^{(k)}) - \nabla(fx^{(k)})\right\|, \left\|U(y^{(k)}) - G(x^{(k))}\right\|\right\} < \varepsilon,$$

STOP.

Otherwise, let $x^{(k+1)} \in \mathbb{R}^n, y^{(k+1)} \in \mathbb{R}^m$ be a solution of linear system (24).

Step 3. *Let the multiplier vector*

$$\lambda^{(k+1)} = V(y^{(k)}) + V'(y^{(k)})\left(y^{(k+1)} - y^{(k)}\right),$$

and $k = k + 1$.

Go to Step 2.

Remark 1. The **Algorithm 1** generates a sequence of pairs $\left(x^{(k)}, y^{(k)}\right)$ – the solution of system of linear equations (24) which consists of $(n + m)$ equations with $(n + m)$ unknowns. The Lagrange multiplier λ is determined as a function of y through the formula (23). Therefore the algorithm can be considered a primal-dual method.

Remark 2. After reformulation, **Algorithm 1** can be used for solving the complementarity problems and variational inequality problems.

Remark 3. As it is well known, Newton method possesses just local convergence, it is very sensitive to the choice of initial approximations and is not convergent if it is not sufficiently close to the solution. There are several ways to modify the method to ensure its global convergence:

- Damped Newton's method [27],
- The Levenberg-Marquardt scheme,
- Trust-regions approach [5,28].

4 Primal superlinear convergence

Newton's method for solving system of linear equations (24) has the advantage of high convergence of couple $\left(x^{(k)}, y^{(k)}\right)$. In the following we show that the rate of convergence for the sequence of primal variables $\left\{x^{(k)}\right\}$ is also superlinear.

In addition to the **Properties P1-P3**, we assume that the functions $u(x)$ and $v(x)$ satisfy the following **Property P4**:

P4.

$$u(x) \times v'(x) = u'(x) \times v(x) = 0 \quad \text{for all } x \in \mathbb{R}.$$

It is not difficult to see that the functions from example (12) also satisfy the **Property P4**.

Theorem 2. ***Let us suppose that the standard Assumptions A1-A3 are satisfied. Assume also that functions*** $u(x)$ ***and*** $v(x)$ ***satisfy the Properties P1-P4. Then the sequence*** $\left\{x^{(k)}\right\}$ ***generated by Algorithm 1 for problem (1) converges locally to*** x^* ***superliniarly.***

Proof. The linear system (24) together with (23) are equivalent to the system (22). Let $x^{(k+1)}$, $\lambda^{(k+1)}$ and $y^{(k+1)}$ be solution for the system equations (22). It is easy to follow the relations obtained from the first equation of system (22):

$$-\nabla_{xx}^2 L\left(x^{(k)},\lambda^{(k)}\right)\left(x^{(k+1)}-x^{(k)}\right)=\nabla_x L\left(x^{(k)},\lambda^{(k+1)}\right)=$$

$$=\nabla_x L\left(x^*,\lambda^{(k+1)}\right)+\left[\nabla_x L\left(x^{(k)},\lambda^{(k+1)}\right)-\nabla_x L\left(x^*,\lambda^{(k+1)}\right)\right]=$$

$$=\nabla_x L\left(x^*,\lambda^{(k+1)}\right)+\nabla_{xx}^2 L\left(x^*,\lambda^{(k+1)}\right)\left(x^{(k+1)}-x^*\right)+$$

$$+o\left(\left\|x^{(k)}-x^*\right\|\right)=\nabla_x L\left(x^*,\lambda^*\right)+\nabla_{x\lambda}^2 L\left(x^*,\lambda^*\right)\left(\lambda^{(k+1)}-\lambda^*\right)+$$

$$=\nabla_x L\left(x^*,\lambda^*\right)+\nabla_{x\lambda}^2 L\left(x^*,\lambda^*\right)\left(\lambda^{(k+1)}-\lambda^*\right)+$$

$$+\nabla_{xx}^2 L\left(x^*,\lambda^{(k+1)}\right)\left(x^{(k+1)}-x^*\right)+o\left(\left\|x^{(k)}-x^*\right\|\right)=$$

$$=\left[G'\left(x^*\right)\right]^T\left(\lambda^{(k+1)}-\lambda^*\right)+\nabla_{xx}^2 L\left(x^*,\lambda^*\right)\left(x^{(k+1)}-x^*\right)+$$

$$=\left[\nabla_{xx}^2 L\left(x^*,\lambda^{(k+1)}\right)-\nabla_{xx}^2 L\left(x^*,\lambda^*\right)\right]\left(x^{(k)}-x^*\right)+$$

$$+o\left(\left\|x^{(k)}-x^*\right\|\right)=\left[G'\left(x^*\right)\right]^T\left(\lambda^{(k+1)}-\lambda^*\right)+$$

$$+\nabla_{xx}^2 L\left(x^*,\lambda^*\right)\left(x^{(k+1)}-x^*\right)-\nabla_{xx}^2 L\left(x^*,\lambda^*\right)\left(x^{(k+1)}-x^{(k)}\right)+$$

$$+o\left(\left\|x^{(k)}-x^*\right\|\right),$$

where we have

$$\left[\nabla_{xx}^2 L\left(x^*,\lambda^*\right)-\nabla_{xx}^2 L\left(x^{(k)},\lambda^{(k)}\right)\right]\left(x^{(k+1)}-x^*\right)=\left[G'\left(x^*\right)\right]^T\times \tag{26}$$

$$\times\left(\lambda^{(k+1)}-\lambda^*\right)+\nabla_{xx}^2 L\left(x^*,\lambda^*\right)\left(x^{(k+1)}-x^*\right)+o\left(\left\|x^{(k)}-x^*\right\|\right).$$

Taking into consideration the **Properties P1-P3** from the last equation of system (22) and from (23), one obtains:

$$\lambda_i^{(k+1)}g_i\left(x^{(k+1)}\right)=\left[v\left(y_i^{(k)}\right)+v'\left(y_i^{(k)}\right)\left(y_i^{(k+1)}-y_i^{(k)}\right)\right]\times$$

$$\times g_i\left(x^{(k+1)}\right)+\left[\nabla g_i\left(x^{(k+1)}\right)\right]^T\left(x^{(k+1)}-x^{(k)}\right)+\left(x^{(k+1)}-x^{(k)}\right)=$$

$$=\left[v\left(y_i^{(k)}\right)+v'\left(y_i^{(k)}\right)\left(y_i^{(k+1)}-y_i^{(k)}\right)\right]\times$$

$$\times\left[u\left(y_i^{(k)}\right)+u'\left(y_i^{(k)}\right)\left(y_i^{(k+1)}-y_i^{(k)}\right)\right]+o\left(\left\|x^{(k)}-x^*\right\|\right).$$

So,

$$\lambda_i^{(k+1)}g_i\left(x^{(k+1)}\right)=o\left(\left\|x^{(k)}-x^*\right\|\right)\quad\text{for any } i. \tag{27}$$

On the other hand,

$$\lambda_i^{(k+1)}g_i\left(x^{(k+1)}\right)=\lambda_i^*g_i\left(x^*\right)+g_i\left(x^*\right)\left(\lambda_i^{(k+1)}-\lambda_i^*\right)+$$

$$+\lambda_i^* \left[\nabla g_i\left(x^*\right)\right]^T \left(x^{(k+1)} - x^*\right) + o\left(\left\|x^{(k+1)} - x^*\right\|\right).$$

From here and from (27), taking into consideration the **Assumption A2**, we have for all $i \in \mathcal{A}\left(x^*\right)$:

$$\left[\nabla g_i\left(x^*\right)\right]^T \left(x^{(k+1)} - x^*\right) = -\delta_i^{(k)}, \tag{28}$$

where

$$-\delta_i^{(k)} = o\left(\left\|x^{(k+1)} - x^*\right\|\right) + o\left(\left\|x^{(k+1)} - x^{(k)}\right\|\right).$$

As the system of vectors $\left\{\nabla g_i\left(x^*\right)\right\}$, $i \in \mathcal{A}\left(x^*\right)$, is linearly independent (**Assumption A2)**, there is a vector $\gamma^{(k)} \in \mathbb{R}^n$ such that:

$$\left[\nabla g_i\left(x^*\right)\right]^T \gamma^{(k)} = \delta_i^{(k)} \tag{29}$$

and

$$\left\|\gamma^{(k)}\right\| = o\left(\left\|x^{(k+1)} - x^*\right\|\right) + o\left(\left\|x^{(k+1)} - x^{(k)}\right\|\right).$$

Let now

$$p^{(k)} = x^{(k+1)} - x^* + \gamma^{(k)} \in \mathbb{R}^n.$$

Then (28) and (29) shows that

$$\left[\nabla g_i\left(x^*\right)\right]^T p^{(k)} = 0 \;\text{ for all } i \in \mathcal{A}\left(x^*\right),$$

i.e. $p^{(k)} \in T\left(x^*\right)$. We also notice that

$$\left[p^{(k)}\right]^T \left[G'\left(x^*\right)\right]^T \left(\lambda^{(k+1)} - \lambda^*\right) = 0. \tag{30}$$

Indeed

$$\left[p^{(k)}\right]^T \left[G'\left(x^*\right)\right]^T \left(\lambda^{(k+1)} - \lambda^*\right) =$$

$$= \sum_{i \in \mathcal{A}(x^*)} \left(\lambda_i^{(k+1)} - \lambda_i^*\right) \left[\nabla g_i\left(x^*\right)\right]^T p^{(k)} = 0.$$

From (29) and (30) we have

$$\left[p^{(k)}\right]^T \left[\nabla_{xx}^2 L\left(x^*, \lambda^*\right) - \nabla_{xx}^2 L\left(x^{(k)}, \lambda^{(k)}\right)\right] \left(x^{(k+1)} - x^*\right) =$$

$$= \left[p^{(k)}\right]^T \nabla_{xx}^2 L\left(x^*, \lambda^*\right) \left(x^{(k+1)} - x^*\right) + o\left(\left\|x^{(k+1)} - x^*\right\|\right) \left\|p^{(k)}\right\|.$$

Finally, the last relationship together with **Assumption A2**, gives us

$$c\left\|p^{(k)}\right\|^2 \leq \left[p^{(k)}\right]^T \nabla_{xx}^2 L\left(x^*, \lambda^*\right) p^{(k)} =$$

$$\left[p^{(k)}\right]^T \left[\nabla_{xx}^2 L\left(x^*, \lambda^*\right) - \nabla_{xx}^2 L\left(x^{(k)}, \lambda^{(k)}\right)\right] \left(x^{(k+1)} - x^*\right) +$$

$$+ \left[p^{(k)}\right]^T \nabla_{xx}^2 L\left(x^*, \lambda^*\right) \gamma^{(k)} + o\left(\left\|x^{(k)} - x^*\right\|\right) \left\|p^{(k)}\right\| =$$

$$= o\left(\left\|x^{(k+1)} - x^*\right\|\right) \left\|p^{(k)}\right\| + O\left(\left\|p^{(k)}\right\| \left\|\gamma^{(k)}\right\|\right) + o\left(\left\|x^{(k)} - x^*\right\|\right) \times$$

$$\times \left\|p^{(k)}\right\| = o\left(\left\|x^{(k+1)} - x^*\right\|\right) \left\|p^{(k)}\right\| + o\left(\left\|x^{(k)} - x^*\right\|\right) \left\|p^{(k)}\right\|,$$

where

$$\left\|p^{(k)}\right\| = o\left(\left\|x^{(k+1)} - x^*\right\|\right) + o\left(\left\|x^{(k)} - x^*\right\|\right).$$

So,

$$\left\|x^{(k+1)} - x^*\right\| = \left\|p^{(k)} - \gamma^{(k)}\right\| \leq \left\|p^{(k)}\right\| + \left\|\gamma^{(k)}\right\| =$$

$$= o\left(\left\|x^{(k+1)} - x^*\right\|\right) + o\left(\left\|x^{(k)} - x^*\right\|\right).$$

The last relationship implies that

$$\left\|x^{(k+1)} - x^*\right\| = o\left(\left\|x^{(k)} - x^*\right\|\right),$$

This completes the proof.■

5 Test examples

In this section, we give some numerical results. The algorithm described in this paper was implemented by a Maple code and tested on a selection of problems from [2] and [33]. As functions $u(x)$ and $v(x)$, there were taken the concrete functions (12), where $\beta = \gamma = 2$.

Example 1. (**Problem:16 [33, p. 39]).**

$$\begin{cases} f(x) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 \to \min \\ \text{s.t. } x_1 + x_2^2 \geq 0, \quad x_1^2 + x_2 \geq 0, \\ -2 \leq x_1 \leq 0.5, \quad x_2 \leq 1. \end{cases}$$

The starting points: $x^{(0)} = (-2, 1)$, $y^{(0)} = (-1, 1)$.
The optimal solution:

$$\begin{aligned} x^* &= (0.5,\ 0.25),\ f(x^*) = 0.25, \mathcal{A}(x^*) = (4), \\ y^* &= (0.82548, 0.7937, 1.3572, -1.0, 0.90856), \\ \lambda^* &= (0.0,\ 0.0,\ 0.0, 1.0,\ 0.0). \end{aligned}$$

Example 2. (**Problem:43 [33, p. 66]).**

$$\begin{cases} f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \to \min \\ \text{s.t. } 8 - x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 \geq 0, \\ 10 - x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 \geq 0, \\ 5 - 2x_1^2 - x_2^2 - x_3^2 - x_4^2 - 2x_1 + x_2 + x_4 \geq 0. \end{cases}$$

The starting points: $x^{(0)} = (0,0,0,0)\,,\, y^{(0)} = (-1,1,-1)$.
The optimal solution:

$$\begin{aligned} x^* &= (0.135\,90,\ 1.092\,7, 1.857\,8,\ \ -1.138\,7\ )\,,\ f(x^*) = -43.716, \\ y^* &= (-1.010\,3, 0.817\,48, -1.103\,4), \\ \lambda^* &= (1.031\,1,\ 0.0,\ 1.343\,3)\,,\ \mathcal{A}(x^*) = (1,3) \end{aligned}$$

Example 3. (Problem:64 [33, p. 86]).

$$\begin{cases} f(x) = 5x_1 + \frac{50000}{x_1} + 20x_2 + \frac{72000}{x_2} + 10x_2 + \frac{144000}{x_3} \to \min \\ \text{s.t. } 1 - \frac{4}{x_1} - \frac{32}{x_2} - \frac{120}{x_3} \geq 0, \quad 0.00001 \leq x_1 \leq 300, \\ 0.00001 \leq x_2 \leq 300, \quad 0.00001 \leq x_3 \leq 300. \end{cases}$$

The starting points: $x^{(0)} = (1,1,1)\,,\, y^{(0)} = (-1,1,-1,1,1,1,1)$.
The optimal solution:

$$\begin{aligned} x^* &= (108.73,\ 85.126,\ 204.32)\,,\ f(x^*) = 6299.8424, \\ y^* &= (-13.160,\ 4.773\,0,\ 5.761\,6,\ 4.399,\ 5.989\,6,\ 5.889\,9,\ 4.573\,7)\,, \\ \lambda^* &= (2279.0,\ 0,\ 0,0,\ 0,0,0)\,,\ \mathcal{A}(x^*) = (1)\,. \end{aligned}$$

6 Conclusion

We have presented and tested smoothing Newton methods for solving nonlinear optimization problems, with requirements compared to those of the classical Newton method. The basic idea of the methods is to replace the KKT system by a system which is appropriate and equivalent to the one on which we apply Newton's method. It is shown that by reasonable choice of functions $u(x)$ and $v(x)$ we obtain effective methods for solving the nonlinear programming problems. The methods are simple and can be applied with different functions $u(x)$ and $v(x)$ that satisfy the **Properties P1-P4.** The numerical results show that the proposed method produces fast local convergence. From the practical point of view, the possibility to relax the assumptions of strict complementarity conditions (**Assumption A2**) remains an open question.

References

[1] N.I.M. Gould, Ph.L Toint, *How mature is nonlinear optimization?* In Applied mathematics entering the 21st century: ICIAM 2003 Congress, SIAM, Philadelphia (2004), pp. 141–161.

[2] N. Andrei, *Models, Test Problems and Applications for Mathematical Programming.* Technical Press, Bucharest, 2003 (in Romanian).

[3] D.P. Berstsekas, *Nonlinear programming.* 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

[4] M.S. Bazaraa, H. D. Sheraly,C.M. Shetty *Nonlinear programming. Theory and algorithms.* 3 rd. edition, Wiley Interscience. A John Wiley & Sons, Inc. Publication, 2006.

[5] R. Fletcher, *Practical methods of optimization.* Second edition, on Wiley & Sons, New-York, 1990 (republished in paperback 2000).

[6] B.N. Pshenichnyi, Yu.M. Danilin, *Numerical methods in extremal problems.* Mir, Moscow, 1978 (translated from the Russian).

[7] J. Nocedal, S.J. Wright, *Numerical optimization.* Second edition, Spriner Verlag, New-York, 1998.

[8] F. Facchinei, J.S. Pang, *Finite-Dimensional variational inequalities and complementarity problems.* Springer-Verlag, New-York, 2003.

[9] N.I.M. Gould, D. Orban, Ph.L Toint, *Numerical methods for large-scale nonlinear optimization.* Acta Numerica, Cambridge University Press (2005), pp. 299–361.

[10] Ph. E. Gill, W. Murray, M.H. Wright, *Practical optimization.* Second edition, Academic Press Inc., 1982.

[11] A.V. Fiacco, G.P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques.* Classics in Applied Mathematics, SIAM, Philadelphia, PA, second edition, 1990.

[12] R.T. Rockafellar, *Lagrange multipliers and optimality.* SIAM Review, 35, pp. 183–238.

[13] P.T. Boggs, J.W. Tolle, *Sequential quadratic programming.* Acta Numerica, Cambridge University Press (1995), pp. 1–51.

[14] N.I.M. Gould, Ph.L Toint, *SQP methods for large-scale nonlinear programming.* In System Modelling and Optimization, Methods, Theory and Applications. Kluwer Academic Publishers, (2000), pp. 149–178.

[15] S.J. Wright, *Primal-dual interior-point methods.* SIAM, Philadelphia, 1997.

[16] A. Forsgren, Ph. E. Gill, Wright, M.H. Wright, *Interior methods for nonlinear optimization.* SIAM, Review, 44 (2002), pp.525–597.

[17] A. Nemirovski, M.J.Todd, *Interior-point methods for optimization.* Acta Numerica, Cambridge University Press (2009), pp. 1–44.

[18] A. Fischer, *An NCP-Function and use for the solution of complementarity problems.* Recent Advandces in Nonsmooth Optimization (1995), pp. 88–105.

[19] M. C. Ferris,Ch. Kanzow, *Complementarity and related problems: A survey.* In Handbook of Applied Optimization. Oxford University Press, New-York (2002), pp. 514–530.

[20] Ch. Kanzow, H. Kleinmichel, *A class of Newton-type methods for equality and inequality constrained optimization.* Optimization Methods and Software, 5, (2) (1995), pp.173–198.

[21] Z.-H. Huang, D. Sun, G. Zhao, *A smootting Newton-type algorithm of stronger convergence for the quadratically constrained quadratic programming.* Computational Optimization and Applications, 35 Issue (2), (2006) 5, (2), pp.199–237.

[22] A. Fischer, *An special Newton type optimization method.* Optimization, 24, (1992), pp. 269–284.

[23] Y.G. Evtushenco, V.A. Purtov, *Sufficient conditions for a minimum for nonlinear programming problems.* Soviet Mathematics Doklady, 30, (1984), pp. 313–316.

[24] A.F. Izmailov, M.V. Solodov, *Numerical Methods of Optimization.* Fizmatlit/Nauka, Moscow, Russia, Second Edition – 2008. (In Russian).

[25] O. Stein, *Lifting mathematical programs with complementary constraints.* Mathematical Programming, 29, (2010), DOI: 10.1007/s10107-010-0345-y.

[26] A.F. Izmailov, A.L. Pogosyan, M.F. Solodov, *Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints mathematical programs with complementary constraints.* Computational Optimization and Application, (2010), DOI: 10.1007/s10589-010-9341-7.

[27] J.M. Ortega, W.C. Rheinboldt, *Iterative solution of nonlinear equation in several variables.* Academic Press, New-York, 1970.

[28] J.E. Dennis, R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice Hall, Englewood Cliff, 1983.

[29] T.J. Ypma, *Historical development of the Newton-Raphson method.* SIAM Review, 37 (4), (1995), pp. 531–551.

[30] C. Kanzow, *Some equation – based methods for nonlinear complementarity problem.* Optimization Method and Software, 3 (1), (1994), pp. 327-340.

[31] M. Kojima, *Strongly stable stationary solution in nonlinear programs.* In S. M. Robinson (Ed), Analysis and Computation of Fixed Points, Academic Press, New-York, (1980), pp. 93–138.

[32] J.-B.Hiriart-Urruty, *Optimisation et analyse convexe.* Presses Universitaires de France, Paris, 1998.

[33] W. Hock, K. Schittkowski, *Test Examples for Nonlinear Programming Codes.* Lecture Notes in Economics and Mathematical Systems, Vol. 187, Springer-Verlag, Berlin / Heidelberg / New York, 1981.

Vasile Moraru

Received April 9, 2011

Technical University of Moldova,
Faculty of Computers, Informatics and Microelectronics
Bd. Stefan cel Mare, 168, MD-2004, Chisinau, Moldova
Phone: (373-2) 50-99-01
E–mail: *moraru@mail.utm.md*