Institute of Computer Science
Academy of Sciences of the Czech Republic

Interior-point method for nonlinear programming with complementarity constraints

L.Lukšan, C.Matonoha, J.Vlček

Technical report No. 1039

December 2008

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 2 688 42 44, fax: +420 2 858 57 89,
e-mail:e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Interior-point method for nonlinear programming with complementarity constraints

L.Lukšan, C.Matonoha, J.Vlček [1]

Technical report No. 1039

December 2008

Abstract:

In this report, we propose an algorithm for solving nonlinear programming problems with complementarity constraints, which is based on the interior-point approach. Main theoretical results concern direction determination and step-length selection. We use an exact penalty function to remove complementarity constraints. Thus a new indefinite linear system is defined with a tridiagonal low-right submatrix. Inexact solution of this system is obtained iteratively using indefinitely preconditioned conjugate gradient method. Furthermore, new merit function is defined, which includes barrier, exact penalty, and augmented Lagrangian terms. The algorithm was implemented in the interactive system for universal functional optimization UFO. Results of extensive numerical experiments are reported.

Keywords:
Nonlinear programming, complementarity constraints, interior-point methods, indefinite systems, indefinite preconditioners, preconditioned conjugate gradient method, merit functions, algorithms, computational experiments.

---

[1] This work was supported by the Grant Agency of the Czech Academy of Sciences, project No. IAA1030405, the Grant Agency of the Czech Republic, project No. 201/06/P397, and the institutional research plan No. AV0Z10300504 L.Lukšan is also from Technical University of Liberec, Hálkova 6, 461 17 Liberec.

1 Introduction

A general nonlinear programming problem with complementarity constraints can be written in the form

$$F(x) \to \min, \quad c_E(x) = 0, \quad c_I(x) \le 0, \quad c_K^T(x) c_L(x) = 0,$$

where $F : R^n \to R$, $c_E : R^n \to R^{m_E}$, $c_I : R^n \to R^{m_I}$, are twice continuously differentiable functions and $I = J \cup K \cup L$ is a disjunctive decomposition of I with $K = \{k_1, \dots, k_p\}$, $L = \{l_1, \dots, l_p\}$. This problem is difficult to solve by standard nonlinear programming methods since the Mangasarian-Fromowitz constraint qualification is not satisfied at any feasible point if $K \ne \emptyset$, $L \ne \emptyset$. Therefore, special methods have been developed by considering complementarity constraints $c_K^T(x) c_L(x) = 0$ separately. In this report, we describe an interior-point method that uses l_1 exact penalty function instead of complementarity constraints. To simplify the description and analysis of this method, we assume without a loss of generality that $E = J = \emptyset$ (constraints $c_E(x) = 0$, $c_J(x) \le 0$ can be treated by a usual way as, e.g. in [9], [11]). Thus we are concerned with the problem

$$F(x) \to \min, \quad c_K(x) \le 0, \quad c_L(x) \le 0, \quad c_K^T(x) c_L(x) = 0. \tag{1}$$

This problem can replaced by the problem

$$F(x) + \rho c_K^T(x) c_L(x) \to \min, \quad c_K(x) \le 0, \quad c_L(x) \le 0, \tag{2}$$

where $\rho > 0$, which has the same solution as (1) if ρ is sufficiently large. The advantage of this transformation consists in the fact that the constraints of problem (2) usually satisfy the Mangasarian-Fromowitz constraint qualification. Problem (2) can be solved by an interior-point method. Thus we solve a sequence of the following IP subproblems

$$F(x) + \rho s_K^T s_L - \mu e^T \ln(S_K) e - \mu e^T \ln(S_L) e \to \min, \quad c_K(x) + s_K = 0, \quad c_L(x) + s_L = 0, \tag{3}$$

where $\rho > 0$, $\mu > 0$ are parameters, $s_K > 0$, $s_L > 0$ are vectors of slack variables, and $S_K = \mathrm{diag}(s_K)$, $S_L = \mathrm{diag}(s_L)$. If we denote $s = (s_K, s_L)$, $u = (u_K, u_L)$, where u_k, u_L are vectors of Lagrange multipliers, then the Lagrange function of subproblem (3) has the form

$$\begin{aligned} L(x, s, u) \;\; = \;\; & F(x) + \rho s_K^T s_L - \mu e^T \ln(S_K) e - \mu e^T \ln(S_L) e \\ & + u_K^T (c_K(x) + s_K) + u_L^T (c_L(x) + s_L), \end{aligned} \tag{4}$$

Denoting $U_K = \mathrm{diag}(u_K)$, $U_L = \mathrm{diag}(u_L)$ and $A_K = \nabla c_K(x)$, $A_L = \nabla c_L(x)$, we obtain the following necessary KKT conditions

$$\nabla_x L(x, s, u) = \nabla F(x) + A_K(x) u_K + A_L(x) u_L = 0,$$

$$\nabla_{s_K} L(x, s, u) = \rho S_L e - \mu S_K^{-1} e + U_K e = 0,$$

$$\nabla_{s_L} L(x,s,u) = \rho S_K e - \mu S_L^{-1} e + U_L e = 0,$$

$$\nabla_{u_K} L(x,s,u) = c_K(x) + s_K = 0,$$

$$\nabla_{u_L} L(x,s,u) = c_L(x) + s_L = 0,$$

or

$$\nabla F(x) + A_K(x) u_K + A_L(x) u_L = 0, \tag{5}$$

$$S_K U_K e + \rho S_K S_L e - \mu e = 0, \quad S_L U_L e + \rho S_K S_L e - \mu e = 0, \tag{6}$$

$$c_K(x) + s_K = 0, \quad c_L(x) + s_L = 0. \tag{7}$$

Applying the Newton method to the nonlinear system (5)–(7), we need to solve a sequence of linear KKT systems

$$\begin{bmatrix} G(x,u) & 0 & 0 & A_K(x) & A_L(x) \\ 0 & U_K + \rho S_L & \rho S_K & S_K & 0 \\ 0 & \rho S_L & U_L + \rho S_K & 0 & S_L \\ A_K^T(x) & I & 0 & 0 & 0 \\ A_L^T(x) & 0 & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s_K \\ \Delta s_L \\ \Delta u_K \\ \Delta u_L \end{bmatrix} = - \begin{bmatrix} g(x,u) \\ g_K(s,u) \\ g_L(s,u) \\ c_K + s_K \\ c_L + s_L \end{bmatrix}, \tag{8}$$

where

$$\begin{aligned} g(x,u) &= \nabla F(x) + A_K(x) u_K + A_L(x) u_L, \\ g_K(s,u) &= S_K U_K e + \rho S_K S_L e - \mu e, \\ g_L(s,u) &= S_L U_L e \ + \rho S_K S_L e \ - \mu e \end{aligned}$$

and

$$G(x,u) = \nabla^2 F(x) + \sum_{i \in K} u_i \nabla^2 c_i(x) + \sum_{i \in L} u_i \nabla^2 c_i(x).$$

The interior-point method for nonlinear programming with complementarity constraints can be roughly described in the following form. For given vectors $x \in R^n$, $s_K \in R^p$, $s_L \in R^p$, $u_K \in R^p$, $u_L \in R^p$ such that $s_K > 0$, $s_L > 0$ and given parameters $\mu > 0$, $\rho > 0$ we determine direction vectors Δx, Δs_K, Δs_L, Δu_K, Δu_L by solving linear system equivalent to (8) (more details are given in Section 2). Furthermore, we choose a step-length $\alpha > 0$ and set $x := x + \alpha \Delta x$, $s_K := s_K + \alpha \Delta s_K$, $s_L := s_L + \alpha \Delta s_L$, $u_K := u_K + \alpha \Delta u_K$, $u_L := u_L + \alpha \Delta u_L$ (more details are given in Section 3). Finally, we determine a new parameters $\mu > 0$, $\rho > 0$, see Section 4.

A similar idea with the exact penalty term $\rho c_K^T(x) c_L(x)$ instead of $\rho s_K^T s_L$ was used in [5], where conditions for global and superlinear convergence were studied.

2 Direction determination

System (8) is nonsymmetric, but it can be easily transformed to the symmetric form

$$\begin{bmatrix} G & 0 & 0 & A_K & A_L \\ 0 & S_K^{-1}(U_K+\rho S_L) & \rho I & I & 0 \\ 0 & \rho I & S_L^{-1}(U_L+\rho S_K) & 0 & I \\ A_K^T & I & 0 & 0 & 0 \\ A_L^T & 0 & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s_K \\ \Delta s_L \\ \Delta u_K \\ \Delta u_L \end{bmatrix} = - \begin{bmatrix} g \\ S_K^{-1} g_K \\ S_L^{-1} g_L \\ c_K + s_K \\ c_L + s_L \end{bmatrix} \tag{9}$$

by multiplying the second and the third equations by S_K^{-1} and S_L^{-1}, respectively. Denoting $A_I = [A_K, A_L]$,

$$g_I = \begin{bmatrix} g_K \\ g_L \end{bmatrix}, \quad s_I = \begin{bmatrix} s_K \\ s_L \end{bmatrix}, \quad u_I = \begin{bmatrix} u_K \\ u_L \end{bmatrix}$$

and

$$M_I^{-1} = \begin{bmatrix} S_K^{-1}(U_K+\rho S_L) & \rho I \\ \rho I & S_L^{-1}(U_L+\rho S_K) \end{bmatrix}, \tag{10}$$

we can write

$$\begin{bmatrix} G & 0 & A_I \\ 0 & M_I^{-1} & I \\ A_I^T & I & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s_I \\ \Delta u_I \end{bmatrix} = - \begin{bmatrix} g \\ S_I^{-1} g_I \\ c_I + s_I \end{bmatrix}. \tag{11}$$

This system can be further simplified by the elimination of vector Δs_I. Using the second equation, we obtain

$$\Delta s_I = -M_I(\Delta u_I + S_I^{-1} g_I), \tag{12}$$

which after substitution into the third equation gives

$$\begin{bmatrix} G & A_I \\ A_I^T & -M_I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u_I \end{bmatrix} = - \begin{bmatrix} g \\ c_I + s_I - M_I S_I^{-1} g_I \end{bmatrix}. \tag{13}$$

Lemma 1 *Assume that the diagonal matrix*

$$D_K = D_L = U_K U_L + \rho(U_K S_K + U_L S_L) \tag{14}$$

is nonsingular. Then

$$M_I = \begin{bmatrix} D_K & 0 \\ 0 & D_L \end{bmatrix}^{-1} \begin{bmatrix} S_K(U_L+\rho S_K) & -\rho S_K S_L \\ -\rho S_K S_L & S_L(U_K+\rho S_L) \end{bmatrix}. \tag{15}$$

If diagonal matrices S_K, S_L, U_K, U_L are positive definite, then also M_I is positive definite.

Proof Since diagonal matrices commute and $D_K = D_L$, we can easily check by multiplication that

$$\begin{bmatrix} S_K^{-1}(U_K+\rho S_L) & \rho I \\ \rho I & S_L^{-1}(U_L+\rho S_K) \end{bmatrix} \begin{bmatrix} S_K(U_L+\rho S_K) & -\rho S_K S_L \\ -\rho S_K S_L & S_L(U_K+\rho S_L) \end{bmatrix} = \begin{bmatrix} D_K & 0 \\ 0 & D_L \end{bmatrix},$$

which confirms (15). If $S_K > 0$, $S_L > 0$, $U_K > 0$, $U_L > 0$, then matrices D_K, D_L, and

$$\begin{bmatrix} S_K U_L & 0 \\ 0 & S_L U_K \end{bmatrix}$$

are positive definite. Since

$$M_I = \begin{bmatrix} D_K & 0 \\ 0 & D_L \end{bmatrix}^{-\frac{1}{2}} \left(\begin{bmatrix} S_K U_L & 0 \\ 0 & S_L U_K \end{bmatrix} + \rho \begin{bmatrix} S_K^2 & -S_K S_L \\ -S_K S_L & S_L^2 \end{bmatrix} \right) \begin{bmatrix} D_K & 0 \\ 0 & D_L \end{bmatrix}^{-\frac{1}{2}},$$

it suffices to prove that the matrix

$$\begin{bmatrix} S_K^2 & -S_K S_L \\ -S_K S_L & S_L^2 \end{bmatrix}$$

is positive semidefinite. But it is true, since one has

$$[v_K^T, v_L^T] \begin{bmatrix} S_K^2 & -S_K S_L \\ -S_K S_L & S_L^2 \end{bmatrix} \begin{bmatrix} v_K \\ v_L \end{bmatrix} = (S_K v_K - S_L v_L)^T (S_K v_K - S_L v_L) \geq 0$$

for arbitrary vectors $v_K \in R^p$, $v_L \in R^p$. □

Linear system (13) with the matrix

$$K = \begin{bmatrix} G & A_I \\ A_I^T & -M_I \end{bmatrix} \tag{16}$$

can be solved either directly by the Bunch-Parlett decomposition (since matrix K is indefinite when M_I is positive semidefinite) or iteratively by the conjugate gradient method preconditioned by the matrix

$$C = \begin{bmatrix} D & A_I \\ A_I^T & -N_I \end{bmatrix} \tag{17}$$

where D is a positive definite diagonal matrix approximating G (e.g. a diagonal of G) and N_I is a suitable matrix. We assume that the matrix C is nonsingular, which implies that the matrix $N_I + A_I^T D^{-1} A_I$, the Schur complement of D in C, is nonsingular. In the subsequent considerations, we consider two cases where either $N_I = M_I$ or

$$N_I = \begin{bmatrix} D_K^{-1} S_K (U_L + \rho S_K) & 0 \\ 0 & D_L^{-1} S_L (U_K + \rho S_L) \end{bmatrix}. \tag{18}$$

In the first case, if $N_I = M_I$, the following theorems, proved in [9], demonstrate advantageous properties of preconditioner C.

Theorem 1 *Matrix KC^{-1} has at least m_I unit eigenvalues with m_I corresponding linearly independent eigenvectors. Remaining eigenvalues of matrix KC^{-1} are eigenvalues of the matrix $\tilde{G}\tilde{D}^{-1}$, where*

$$\tilde{G} = G + A_I M_I^{-1} A_I^T, \quad \tilde{D} = D + A_I M_I^{-1} A_I^T.$$

If matrices $\tilde{G}$, $\tilde{D}$ are positive definite, then all eigenvalues of KC^{-1} are positive.

Theorem 2 *The dimension of the Krylov subspace defined by matrix KC^{-1} is at most $n+1$.*

Theorem 3 *Consider the conjugate gradient method with preconditioner C applied to system (13). Assume that matrices $\tilde{G}$, $\tilde{D}$ are positive definite and choose the initial estimation of Δx in such a way that the second equation is satisfied accurately (e.g. set $\Delta x = -D^{-1}A_I(A_I^T D^{-1} A_I)^{-1}(c_I + s_I - M_I S_I^{-1} g_I)$). Then:*

- *Vector Δx^* (the first part of the solution) is found after n iterations at most.*
- *Algorithm cannot fail before Δx^* is found.*
- *The norm $\|\Delta x - \Delta x^*\|$ converges to zero at least R-linearly with a quotient*

$$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1},$$

  where κ is the spectral condition number of matrix $\tilde{G}(\tilde{D})^{-1}$.
- *If $\Delta x = \Delta x^*$, then also $\Delta u_I = \Delta u_I^*$.*

A disadvantage of the choice $N_I = M_I$ is the fact that matrix M_I can be indefinite. This fact motivated us to use a positive diagonal of M_I given by (18). Nevertheless, preconditioner C with N_I given by (18) has not an excellent properties given by Theorem 1–Theorem 3 and computational efficiency is also lower in comparison with the choice $N_I = M_I$.

3 Stepsize selection

Having computed directions Δx, Δs_I, Δu_I, we need to select a suitable stepsize α for computing new vectors

$$x^+ = x + \min(\alpha, \overline{\alpha}_x)\Delta x, \quad s_I^+ = s_I + \min(\alpha, \overline{\alpha}_s)\Delta s_I, \quad u_I^+ = u_I + \min(\alpha, \overline{\alpha}_u)\Delta u_I, \tag{19}$$

where $\overline{\alpha}_x > 0$, $\overline{\alpha}_s > 0$, $\overline{\alpha}_u > 0$ are suitable upper bounds. Theoretically, the Newton method requires a full step $\alpha = 1$ (here we assume that upper bounds $\overline{\alpha}_x$, $\overline{\alpha}_s$, $\overline{\alpha}_u$ are sufficiently large). But the unit stepsize is sometimes unsuitable and has to be decreased. Usually, a merit function $P(\alpha)$ is used for this purpose and a stepsize α is chosen in such a way that $\alpha = \beta^j \min(1, \overline{\alpha}_x)$, where $0 < \beta < 1$, and $j \geq 0$ is the lowest integer for which $P(\alpha) < P(0)$. Motivated by [6], we use the following merit function

$$\begin{aligned} P(\alpha) \;=\; & F(x + \alpha\Delta x) \\ & + (u_K + \Delta u_K)^T (c_K(x + \alpha\Delta x) + s_K + \alpha\Delta s_K) \\ & + (u_L + \Delta u_L)^T (c_L(x + \alpha\Delta x) + s_L + \alpha\Delta s_L) \\ & + \rho(s_K + \Delta s_K)^T (s_L + \alpha\Delta s_L) + \rho(s_L + \Delta s_L)^T (s_K + \alpha\Delta s_K) \\ & - \mu\, e^T \ln(S_K + \alpha\Delta S_K)e - \mu\, e^T \ln(S_L + \alpha\Delta S_L)e \\ & + \frac{\sigma}{2}\|c_K(x + \alpha\Delta x) + s_K + \alpha\Delta s_K\|^2 + \frac{\sigma}{2}\|c_L(x + \alpha\Delta x) + s_L + \alpha\Delta s_L\|^2, \end{aligned} \tag{20}$$

where $\rho > 0$, $\mu > 0$, $\sigma \geq 0$. The following theorem holds.

Theorem 4 *Let $U_K + \rho S_L > 0$, $U_L + \rho S_K > 0$ and let the pair Δx, Δu_I be an inexact solution of system (13) so that*

$$\begin{bmatrix} G & A_I \\ A_I^T & -M_I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u_I \end{bmatrix} + \begin{bmatrix} g \\ c_I + s_I - M_I S_I^{-1} g_I \end{bmatrix} = \begin{bmatrix} r \\ r_I \end{bmatrix}, \tag{21}$$

where $r_I^T = [r_K^T, r_L^T]$. Then

$$\begin{aligned} P'(0) \;=\; & -\Delta x^T G \Delta x + \Delta x^T r \\ & - \Delta s_K^T S_K^{-1}(U_K + \rho S_L)\Delta s_K - \Delta s_L^T S_L^{-1}(U_L + \rho S_K)\Delta s_L \\ & - \sigma\|c_K + s_K\|^2 - \sigma\|c_L + s_L\|^2 + \sigma(c_K + s_K)^T r_K + \sigma(c_L + s_L)^T r_L. \end{aligned} \tag{22}$$

If

$$\sigma > -\frac{\Delta x^T G \Delta x + \Delta s_K^T S_K^{-1}(U_K + \rho S_L)\Delta s_K + \Delta s_L^T S_L^{-1}(U_L + \rho S_K)\Delta s_L}{\|c_K + s_K\|^2 + \|c_L + s_L\|^2} \tag{23}$$

and if (13) is solved with a sufficient precision, namely if

$$\begin{aligned} \Delta x^T r + \sigma(c_K + s_K)^T r_K + \sigma(c_L + s_L)^T r_L < {} & \Delta x^T G \Delta x \\ & + \Delta s_K^T S_K^{-1}(U_K + \rho S_L)\Delta s_K + \Delta s_L^T S_L^{-1}(U_L + \rho S_K)\Delta s_L \\ & + \sigma\|c_K + s_K\|^2 + \sigma\|c_L + s_L\|^2, \end{aligned} \tag{24}$$

then $P'(0) < 0$.

Proof Differentiating (20) by α, we obtain

$$\begin{aligned} P'(0) \;=\; & \Delta x^T(\nabla F(x) + A_K(u_K + \Delta u_K) + A_L(u_L + \Delta u_L)) \\ & + \Delta s_K^T(u_K + \Delta u_K) + \Delta s_L^T(u_L + \Delta u_L) \\ & + \rho\,\Delta s_K^T(s_L + \Delta s_L) + \rho\,\Delta s_L^T(s_K + \Delta s_K) \\ & - \mu\,\Delta s_K^T S_K^{-1} e - \mu\,\Delta s_L^T S_L^{-1} e \\ & + \sigma(c_K + s_K)^T(A_K^T \Delta x + \Delta s_K) + \sigma(c_L + s_L)^T(A_L^T \Delta x + \Delta s_L) \end{aligned} \tag{25}$$

Using the equality

$$\begin{bmatrix} G & 0 & 0 & A_K & A_L \\ 0 & S_K^{-1}(U_K + \rho S_L) & \rho I & I & 0 \\ 0 & \rho I & S_L^{-1}(U_L + \rho S_K) & 0 & I \\ A_K^T & I & 0 & 0 & 0 \\ A_L^T & 0 & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s_K \\ \Delta s_L \\ \Delta u_K \\ \Delta u_L \end{bmatrix}$$

$$+ \begin{bmatrix} \nabla F(x) + A_K u_K + A_L u_L \\ u_K + \rho s_L - \mu S_K^{-1} e \\ u_L + \rho s_K - \mu S_L^{-1} e \\ c_K + s_K \\ c_L + s_L \end{bmatrix} = \begin{bmatrix} r \\ 0 \\ 0 \\ r_K \\ r_L \end{bmatrix},$$

which is equivalent to (12) and (21), we obtain

$$\Delta x^T(\nabla F(x) + A_K(u_K + \Delta u_K) + A_L(u_L + \Delta u_L)) = -\Delta x^T G \Delta x + \Delta x^T r,$$
$$\Delta s_K^T(u_K + \Delta u_K) + \rho \Delta s_K^T(s_L + \Delta s_L) - \mu \Delta s_K^T S_K^{-1} e = -\Delta s_K^T S_K^{-1}(u_K + \rho S_L)\Delta s_K,$$
$$\Delta s_L^T(u_L + \Delta u_L) + \rho \Delta s_L^T(s_K + \Delta s_K) - \mu \Delta s_L^T S_L^{-1} e = -\Delta s_L^T S_L^{-1}(u_L + \rho S_L)\Delta s_L,$$
$$(c_K + s_K)^T(A_K^T \Delta x + \Delta s_K) = -\|c_K + s_K\|^2 + (c_K + s_K)^T r_K,$$
$$(c_L + s_L)^T(A_L^T \Delta x + \Delta s_L) = -\|c_L + s_L\|^2 + (c_L + s_L)^T r_L,$$

which after substituting into (25) gives (22). If (23) holds, then the right-hand side in (24) is positive so if (13) is solved with a sufficient precision, then (24) holds and $P'(0) < 0$ by (22). □

Merit function (20) contains a new penalty parameter σ. Condition (23) restricts the choice of parameter σ weakly. If matrix G is positive semidefinite, any value $\sigma \geq 0$ satisfies this condition. In the opposite case, the second term, which is always positive, decreases the value of $P'(0)$ and partially eliminates the influence of the first term.

Inequality (23) gives one possibility for the computation of parameter σ, which implies $P'(0) < 0$ if (24) holds. But it is usually more efficient for practical computation to choose parameter σ as a constant and replace matrix G by a positive definite diagonal matrix D if condition $P'(0) < 0$ does not hold. If D is the same as in preconditioner C (where $C_I = M_I$), then $KC^{-1} = I$ and we obtain the solution of (13) in the first CG step.

Now we concentrate to the determination of upper bounds $\overline{\alpha}_x$, $\overline{\alpha}_s$, $\overline{\alpha}_u$. We usually set

$$\overline{\alpha}_x = \frac{\overline{\Delta}}{\|\Delta x\|}, \tag{26}$$

where value $\overline{\Delta}$ is used as a safeguard against possible overflows. The upper bound $\overline{\alpha}_s$ assures positivity of s_I^+. Thus we should set $\overline{\alpha}_s \leq \overline{\alpha}_s^{(1)}$, where

$$\overline{\alpha}_s^{(1)} = \tau \min_{i \in I,\, \Delta s_i < 0} \left(-\frac{s_i}{\Delta s_i} \right),$$

and $0 < \tau < 1$ is a coefficient close to unit. Unfortunately, the same idea cannot be used for Lagrange multipliers, since they can be negative by (6) (if complementarity constraints are not satisfied). Instead of inequality $u_I^+ > 0$, we need to assure inequalities $U_K^+ + \rho S_L^+ > 0$, $U_L^+ + \rho S_K^+ > 0$ used in Theorem 4. These inequalities restrict both $\overline{\alpha}_s$ and $\overline{\alpha}_u$. Thus we set

$$\overline{\alpha}_u = \overline{\alpha}_s = \min(\overline{\alpha}_s^{(1)}, \overline{\alpha}_s^{(2)}, \overline{\alpha}_s^{(2)}), \tag{27}$$

where

$$\overline{\alpha}_s^{(2)} = \tau \min_{\substack{1 \leq i \leq p \\ \Delta u_{k_i} + \rho \Delta s_{l_i} < 0}} \left(-\frac{u_{k_i} + \rho s_{l_i}}{\Delta u_{k_i} + \rho \Delta s_{l_i}} \right),$$
$$\overline{\alpha}_s^{(3)} = \tau \min_{\substack{1 \leq i \leq p \\ \Delta u_{l_i} + \rho \Delta s_{k_i} < 0}} \left(-\frac{u_{l_i} + \rho s_{k_i}}{\Delta u_{l_i} + \rho \Delta s_{k_i}} \right).$$

Note that $U_K^+ + \rho S_L^+ > 0$, $U_L^+ + \rho S_K^+ > 0$ imply $U_K^+ + \rho^+ S_L^+ > 0$, $U_L^+ + \rho^+ S_K^+ > 0$ for every $\rho^+ \geq \rho$, so we can increase ρ in the next iteration.

4 The choice of parameters and their update

Interior-point methods for nonlinear programming with complementarity constraints are theoretically studied in [5]. It is shown that if the interior-point subproblems are solved with a sufficient precision and parameters μ and ρ are updated by a suitable way, then the interior-point method for nonlinear programming with complementarity constraints is globally convergent. Unfortunately, their strict rules for updating μ and ρ are not suitable for large problems with sparse matrices (since it is difficult to solve a large interior-point subproblem with a sufficient precision). Therefore, we use different strategies based on heuristic formulas which have been verified by computational experiments.

Our implementation of interior-point methods choose the value μ in such a way that

$$\mu = \max\left(\underline{\mu},\ \lambda \frac{s_K^T(u_K + \rho s_L) + s_L^T(u_L + \rho s_K)}{m_I}\right), \tag{28}$$

where $\underline{\mu} > 0$ is a small lower bound for the barrier parameter which serves as a safeguard and $0 < \lambda < 1$. This choice corresponds to a usual strategy used for standard nonlinear programming problems (where $\rho = 0$). Computational experience has shown that the algorithm performs best when components $s_{k_i}(u_{k_i} + \rho s_{l_i})$, $s_{l_i}(u_{l_i} + \rho s_{k_i})$, $1 \le i \le p$, of the dot-product in numerator approach zero at a uniform rate. The distance from uniformity can be measured by the ratio

$$\nu = 2\,p\,\frac{\min_{1\le i\le p}[s_{k_i}(u_{k_i} + \rho s_{l_i}) + s_{l_i}(u_{l_i} + \rho s_{k_i})]}{\sum_{i=1}^{p}[s_{k_i}(u_{k_i} + \rho s_{l_i}) + s_{l_i}(u_{l_i} + \rho s_{k_i})]}$$

(also called the centrality measure). Clearly, $0 < \nu \le 1$ and $\nu = 1$ if and only if the conditions (6) hold. The value λ is then computed by using ν. Heuristic formulas are usually used for this purpose. In our implementation, we have used the formula

$$\lambda = 0.1 \min\left(0.05\frac{1-\nu}{\nu}, 2\right)^3 \tag{29}$$

proposed in [11].

Parameter ρ should be increased if $|c_K^T(x) c_L^T(x)|$ (the violation of complementarity constraints) is much larger than $\|c_I^0(x)\|$, where $c_i^0(x) = \max(c_i(x), 0)$, $i \in I$. We use the condition

$$|c_K^T(x) c_L^T(x)| \le \underline{\rho} \max(10^{-8}, \|c_I^0\|), \tag{30}$$

where $\underline{\rho} > 0$ is a suitable constant. If this inequality holds, we set $\rho^+ = \rho$. In the opposite case, we set $\rho^+ = \min(\gamma\rho, \overline{\rho})$, where $\gamma > 1$ is a suitable coefficient and $\overline{\rho} > 0$ is a large upper bound which serves as a safeguard.

Concerning parameter σ, we use a small constant value. If $P'(0) \ge 0$, than σ is not increased, but the iteration is restarted with G replaced by D as was pointed out in Section 3.

5 Description of the algorithm

The above considerations can be summarized in the algorithmic form.

Algorithm 1.

Data: Minimum precision for the direction determination $0 < \overline{\omega} < 1$. Line-search parameter $0 < \beta < 1$. Maximum step-length reduction $0 < \tau < 1$. Lower bound for the barrier parameter $\underline{\mu} > 0$. Level for changing the exact penalty parameter $\underline{\rho} > 0$. Upper bound for the exact penalty parameter $\overline{\rho} > 0$. Rate of the exact penalty parameter increase. Step bound $\overline{\Delta} > 0$.

Input: Sparsity pattern of matrices $\nabla^2 F$ and A_I. Initial choice of vector x.

Step 1: *Initiation.* Choose the values $\mu > 0$ (e.g. $\mu = 1$), $\rho > 0$ (e.g. $\rho = 1$) and $\sigma > 0$ (e.g. $\sigma = 0.01$). For $i \in I$ set $s_i := \max(-c_i(x), \delta_s)$ and $u_i := \delta_u$, where $\delta_s > 0$ (e.g. $\delta_s = 0.1$) and $\delta_u > 0$ (e.g. $\delta_u = 0.1$). Compute value $F(x)$ and vector $c_I(x)$. Set $k := 0$.

Step 2: *Termination.* Compute matrix $A_I := A_I(x)$ and vector $g := g(x, u)$. If complementarity constraints (1) and KKT conditions (5)–(7) are satisfied with a sufficient precision and μ is sufficiently small, then terminate the computation. Otherwise set $k := k + 1$.

Step 3: *Approximation of the Hessian matrix.* Compute approximation G of the Hessian matrix $G(x, u)$ by using differences of gradient $g(x, u)$ as in [3].

Step 4: *Direction determination.* Build linear system (13) and choose a suitable preconditioner of form (17). Determine positive definite diagonal matrix D as an approximation of the diagonal of G and factorize the matrix $A_I^T D^{-1} A_I + N_I$ by using the complete or incomplete Gill-Murray decomposition to obtain a representation of C^{-1}. Set $\omega = \min(\|g\|, 1/k, \overline{\omega})$ and determine direction vectors Δx, Δu_I as an inexact solution of (13) (with the precision ω) by using a preconditioned Krylov-subspace method. Compute vector Δs_I by (12). Compute directional derivative $P'(0)$ of the merit function $P(\alpha)$ by (25).

Step 5: *Restart.* If $P'(0) \geq 0$, determine positive definite diagonal matrix D by the procedure given in [8], set $G = D$ and go to Step 4.

Step 6: *Step-length selection.* Define maximum step-lengths $\overline{\alpha}_x$, $\overline{\alpha}_s$, $\overline{\alpha}_u$ by (26)–(27). Find the minimum integer $l \geq 0$ such that $P(\beta^l \overline{\alpha}) < P(0)$. Set $\alpha = \beta^l \overline{\alpha}$ and $x := x^+$, $s_I := s_I^+$, $u_I := u_I^+$, where x^+, s_I^+, u_I^+ are vectors given by (19). Compute value $F(x)$ and vector $c_I(x)$.

Step 7: *Parameters update.* Determine μ by (28), where λ is computed by (29). Multiply ρ by γ if (30) is not satisfied. Go to Step 2.

6 Computational experiments

Algorithm 1 was tested by using a set of 18 test problems with 100 variables. This set was obtained by a modification of test problems for equality constrained minimization given in [6] and [7] (Test18), which can be downloaded (together with report [7]) from `http://www.cs.cas.cz/luksan/test.html`. In our set, equalities $c_i(x) = 0$, $1 \le i \le m$, are replaced by complementarity constraints $c_i(x) \le 0$, $c_{i+p}(x) \le 0$, $c_i(x)c_{i+p}(x) = 0$, $1 \le i \le p = m/2$. In Algorithm 1, we have used values $\overline{\omega} = 0.9$, $\beta = 0.5$, $\tau = 0.95$, $\underline{\mu} = 1.0^{-15}$, $\underline{\rho} = 10^2$, $\underline{\rho} = 10^6$, $\gamma = 3$. The default value $\overline{\Delta} = 10^3$ was frequently decreased. We have used preconditioner C with $N_I = M_I$ in our tests (preconditioner C with N_I given by (18) gave worse results).

The results of the tests are listed in Table 1, where NIT is the number of iterations, NFV is the number of function evaluations, NFG is the number of gradient evaluations (NFG is greater than NFV since the second order derivatives are computed by using gradient differences), NCG is the number of CG iterations. The last row contains summary results for all 18 problems together with the total number of restarts NRS and the total computational time.

P	NIT	NFV	NFG	NCG	F	$\|c_I^0\|$	$\|c_K^T c_L\|$	$\|g\|$
1	35	35	210	977	3.98714	0.0	0.3E-13	0.1E-11
2	71	71	994	8354	2084.88	0.9E-12	0.3E-08	0.2E-09
3	12	12	72	41	14.1685	0.0	0.4E-21	0.6E-06
4	33	34	198	165	454.645	0.2E-14	0.2E-20	0.4E-09
5	46	55	460	465	4.890021E-01	0.0	0.6E-11	0.1E-06
6	19	19	266	88	6037.6532	0.4E-15	0.2E-29	0.5E-08
7	16	16	112	29	-34.9980	0.0	0.3E-25	0.3E-09
8	128	189	896	1858	9743.49	0.4E-15	0.9E-14	0.1E-08
9	450	2007	3157	5911	9.99304	0.3E-01	0.1E-05	0.2E-03
10	13	13	78	64	2.23397	0.6E-16	0.1E-11	0.3E-10
11	74	75	444	8471	1.663530E-16	0.5E-11	0.1E-16	0.6E-09
12	33	33	231	2928	3.748598E-11	0.0	0.1E-10	0.2E-10
13	39	102	312	1928	339.382	0.0	0.4E-27	0.4E-08
14	72	72	504	3544	2.141127E-21	0.0	0.2E-19	0.4E-15
15	126	128	756	13551	1.083434E-17	0.0	0.6E-17	0.6E-12
16	42	49	210	4848	2.846946E-17	0.0	0.2E-17	0.1E-14
17	32	42	160	2278	29.4314	0.2E-12	0.9E-13	0.8E-07
18	108	146	540	3849	32.5028	0.5E-64	0.9E-11	0.4E-07
Σ	1349	3098	9600	59349	NRS = 63		TIME=1.72	

Table 1: Test 18 – Problems with 100 variables

7 Conclusions

The results proposed in Table 1 imply several conclusions:

- The idea used in this report seems to be reasonable. Algorithm 1 solved all problems except Problem 9 with a sufficient precision. Problem 9 was solved after changing parameters $\underline{\rho}$, $\overline{\rho}$ and γ.
- Linear system (13) is usually worse conditioned than similar system obtained by interior-point methods for standard nonlinear programming problems. Thus the number of CG iterations is larger in comparison with problems where complementarity constraints are not present.
- We have used a simple procedure for updating the exact penalty parameter ρ and have observed that the efficiency of the method strongly depends on parameters $\underline{\rho}$, $\overline{\rho}$ and γ. For this reason, the efficiency of Algorithm 1 could be increased by using more sophisticated procedure, which could be the main field for future research. We have also used procedures proposed in [5], which require sufficient precision in solving IP subproblems, but the results obtained were not satisfactory.

References

[1] I.Bongartz, A.R.Conn, N.Gould, P.L.Toint: CUTE: constrained and unconstrained testing environment. ACM Transactions on Mathematical Software, 21, 123-160, 1995.

[2] R.H.Byrd, J.Nocedal, R.A.Waltz: Feasible interior methods using slacks for nonlinear optimization. Report No. OTC 2000/11, Optimization Technology Center, December 2000.

[3] A.R.Curtis, M.J.D.Powell, J.K.Reid: On the estimation of sparse Jacobian matrices. IMA Journal of Aplied Mathematics 13, 117-119, 1974.

[4] A.El-Bakry, R.Tapia, T.Tsuchiya, Y.Zhang: On the formulation and theory of Newton interior-point method for nonlinear programming. Journal of Optimization Theory and Applications, 89, 507-541, 1996.

[5] S.Leyffer, G.Lopez-Calva, J. Nocedal: Interior Methods for Mathematical Programs with Complementarity Constraints.

[6] L.Lukšan, J.Vlček: Indefinitely Preconditioned Inexact Newton Method for Large Sparse Equality Constrained Nonlinear Programming Problems. Numerical Linear Algebra with Applications, 5, 219-247, 1998.

[7] L.Lukšan, J.Vlček: Sparse and partially separable test problems for unconstrained and equality constrained optimization. Report V-767, Institute of Computer Science, Czech Academy of Sciences, Prague 1998.

[8] L.Lukšan, J.Vlček: Numerical experience with iterative methods for equality constrained nonlinear programming problems. Optimization Methods and Software, 16, 257-287, 2001.

[9] L.Lukšan, C.Matonoha, J.Vlček: Interior-point method for nonlinear nonconvex optimization. Numerical Linear Algebra with Applications 11 (2004) 431-453.

[10] L.Lukšan, M.Tůma, J.Hartman, J.Vlček, N.Ramešová, M.Šiška, C.Matonoha: UFO 2006 - Interactive system for universal functional optimization. Technical Report V-977, Institute of Computer Science, Czech Academy of Sciences, Prague 2006.

[11] J.Vanderbei, D.F.Shanno: An interior point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications, 13, 231-252, 1999.