ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Applied Mathematics and Computation xxx (2004) xxx–xxx

APPLIED MATHEMATICS AND COMPUTATION

www.elsevier.com/locate/amc

An inexact Newton method combined with Hestenes multipliers' scheme for the solution of Karush–Kuhn–Tucker systems ☆

Silvia Bonettini [a], Emanuele Galligani [a,*], Valeria Ruggiero [b]

[a] *Dipartimento di Matematica, Università di Modena e Reggio Emilia, Via Campi 213/b, 41100 Modena, Italy*

[b] *Dipartimento di Matematica, Università di Ferrara, Via Machiavelli 35, 44100 Ferrara, Italy*

Abstract

In this work a Newton interior-point method for the solution of Karush–Kuhn–Tucker systems is presented.

A crucial feature of this iterative method is the solution, at each iteration, of the inner subproblem. This subproblem is a linear-quadratic programming problem, that can solved approximately by an inner iterative method such as the Hestenes multipliers' method.

A deep analysis on the choices of the parameters of the method (perturbation and damping parameters) has been done.

The global convergence of the Newton interior-point method is proved when it is viewed as an inexact Newton method for the solution of nonlinear systems with restriction on the sign of some variables.

☆ This research was supported by the Italian Ministry for Education, University and Research (MIUR), FIRB Project RBAU01JYPN.

* Corresponding author.

E-mail addresses: bonettini.silvia@unimo.it (S. Bonettini), galligan@unimo.it (E. Galligani), rgv@dns.unife.it (V. Ruggiero).

doi:10.1016/j.amc.2004.09.018

The Newton interior-point method is numerically evaluated on large scale test problems arising from elliptic optimal control problems which show the effectiveness of the approach.

Keywords: Newton interior-point method; Inexact Newton method; Large scale nonlinear programming problems; Hestenes multipliers' method; Elliptic control problems

1. Introduction

This work is concerned with the numerical solution of large scale nonlinear programming (NLP) problems arising, for instance, from the discretization of optimal control problems with partial differential equations and control and state constraints. In particular, we make reference to optimal control problems for semilinear elliptic equations subject to control and state inequality constraints, with boundary or distributed control. These problems have been studied in [18]: by a suitable finite–difference discretization scheme, a control problem is transcribed into a large scale finite–dimensional NLP problem (see also [16,17]), where the objective function often is a quadratic form, the elliptic state equation and the Dirichlet and/or Neumann boundary conditions become equality constraints and the control and state constraints are simple box constraints. The numerical solution of this large scale NLP problem can be determined by solving the constrained system of nonlinear equations obtained by the Karush–Kuhn–Tucker (KKT) optimality conditions of the problem. In this paper, we propose to solve the nonlinear system by an inexact Newton scheme, that includes the strategy of the interior-point (IP) method for the treatment of the constrained variables. For the linear system arising at each step of the iterative scheme, we can use an inner iterative method, devising an *adaptive* stopping rule that allows to avoid unnecessary inner iterations for the initial outer iterations and, at the same time, assures global convergence and local superlinear convergence of the whole method (see also [8]).

In the next section, we describe the steps of this inexact Newton method, pointing out the meaning of the parameters on which the scheme depends. In Section 3, we discuss about the convergence of the scheme. In Section 4, taking into account the special features of the NLP problem (equality and box constraints), we focus our attention to the inner linear system that has be solved at each iteration, given by the reduced Karush–Kuhn–Tucker indefinite system. Under suitable conditions that assure the nonsingularity of the system and the boundedness of its inverse, the system can be viewed as the Lagrange necessary conditions for the minimum point of a convex quadratic programming problem and it can be efficiently solved by Hestenes multipliers' scheme. Then, the solution of the KKT indefinite system is led to the solution of a sequence of symmetric positive definite systems. Numerical experiments on the elliptic control problems show that, generally, one or two iterations of the Hestenes scheme per outer iteration are sufficient to satisfy the inner adaptive stopping rule. As consequence, this inexact Newton method preserves the good behaviour of the classical interior-point methods and, at the same time, can take advantage of the use of efficient sparse Cholesky solvers for the solution of inner positive definite systems.

In the following, $\|\cdot\|$ denotes the Euclidean vector norm or the spectral matrix norm.

2. An inexact Newton method for Karush–Kuhn–Tucker systems

Consider the following nonlinear programming problem:

$$\begin{aligned} &\min f(\boldsymbol{x}),\\ &\boldsymbol{g}_1(\boldsymbol{x}) = 0,\\ &\boldsymbol{g}_2(\boldsymbol{x}) \geqslant 0, \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{g}_1(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^{neq}$, $\boldsymbol{g}_2(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$. We assume $f(\boldsymbol{x})$, $\boldsymbol{g}_1(\boldsymbol{x})$, $\boldsymbol{g}_2(\boldsymbol{x})$ are twice continuously differentiable and the first and second derivatives of the objective function and constraints are available.

By introducing slack variables, the problem (1) can be rewritten as

$$\begin{aligned} &\min f(\boldsymbol{x}),\\ &\boldsymbol{g}_1(\boldsymbol{x}) = 0,\\ &\boldsymbol{g}_2(\boldsymbol{x}) - \boldsymbol{s} = 0,\\ &\boldsymbol{s} \geqslant 0, \end{aligned} \tag{2}$$

whose KKT optimality conditions are

$$\begin{aligned} &\boldsymbol{\alpha} \equiv \nabla f(\boldsymbol{x}) - \nabla \boldsymbol{g}_1(\boldsymbol{x})\boldsymbol{\lambda}_1 - \nabla \boldsymbol{g}_2(\boldsymbol{x})\boldsymbol{\lambda}_2 = 0,\\ &\boldsymbol{\beta} \equiv -\boldsymbol{g}_1(\boldsymbol{x}) = 0,\\ &\boldsymbol{\gamma} \equiv -\boldsymbol{g}_2(\boldsymbol{x}) + \boldsymbol{s} = 0,\\ &\boldsymbol{\theta} \equiv \Lambda_2 S \boldsymbol{e}_m = 0 \end{aligned} \tag{3}$$

with

$$\boldsymbol{s} \geqslant 0 \quad \boldsymbol{\lambda}_2 \geqslant 0,$$

where $\boldsymbol{s}, \boldsymbol{\lambda}_2 \in \mathbb{R}^m$, $\Lambda_2 = \mathrm{diag}(\boldsymbol{\lambda}_2)$, $S = \mathrm{diag}(\boldsymbol{s})$. The vector $\boldsymbol{e}_m$ indicates the vector of m components whose values are equal to 1.

The system (3) can be written as

$$\begin{aligned} &\boldsymbol{H}(\boldsymbol{v}) = 0, \\ &\boldsymbol{s} \geqslant 0 \quad \boldsymbol{\lambda}_2 \geqslant 0, \end{aligned} \tag{4}$$

where

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \boldsymbol{s} \end{pmatrix} \quad \text{and} \quad \boldsymbol{H}(\boldsymbol{v}) = \begin{pmatrix} \left.\begin{matrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{matrix}\right\} \boldsymbol{H}_1(\boldsymbol{v}) \\ \boldsymbol{\theta} \end{pmatrix}.$$

The Jacobian matrix of $\boldsymbol{H}$ is the matrix

$$H'(\boldsymbol{v}) = \begin{pmatrix} Q & B & C & 0 \\ B^T & 0 & 0 & 0 \\ C^T & 0 & 0 & I \\ 0 & 0 & S & \Lambda_2 \end{pmatrix},$$

where $Q = \nabla^2 f(\boldsymbol{x}) - \sum_1^{neq} \lambda_{1,i} \nabla^2 g_{1,i}(\boldsymbol{x}) - \sum_1^m \lambda_{2,i} \nabla^2 g_{2,i}(\boldsymbol{x})$ is the Hessian matrix of the Lagrangian function of the problem (2), $B = -\nabla \boldsymbol{g}_1(\boldsymbol{x})$ and $C = -\nabla \boldsymbol{g}_2(\boldsymbol{x})$. Here $\nabla^2 f(\boldsymbol{x})$, $\nabla^2 g_{1,i}(\boldsymbol{x})$, $\nabla^2 g_{2,i}(\boldsymbol{x})$ are the Hessian matrices of the function $f(\boldsymbol{x})$ and of i-th component of the constraints $\boldsymbol{g}_1(\boldsymbol{x})$, and $\boldsymbol{g}_2(\boldsymbol{x})$ respectively; then, $\lambda_{1,i}$ and $\lambda_{2,i}$ are the i-th component of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ respectively.

In order to solve the system (4), we can use a Newton–type method that consists in computing the solution $\Delta \boldsymbol{v}^{(k)}$ of the linear system

$$H'(\boldsymbol{v}^{(k)}) \Delta \boldsymbol{v} = -H(\boldsymbol{v}^{(k)}), \tag{5}$$

and in updating the current iterate by

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{v}^{(k)} + \alpha_k \Delta \boldsymbol{v}^{(k)}.$$

As pointed out in [10], when we consider the last block of equations of the system (5), called *complementarity equations*, it can be observed that if $s_i^{(k)} = 0$ (or $\lambda_{2,i}^{(k)} = 0$), then $s_i^{(j)} = 0$ (or $\lambda_{2,i}^{(j)} = 0$), for all $j > k$. It means that, if the iterate reaches the boundary of the feasible region, it sticks on the boundary even if it is far from the solution. In order to avoid this drawback, the complementarity equations are modified by introducing a perturbation parameter ρ. This yields the perturbed KKT conditions

$$\begin{aligned} &\boldsymbol{H}(\boldsymbol{v}) = \rho \tilde{\boldsymbol{e}}, \\ &\boldsymbol{s} \geqslant 0 \quad \boldsymbol{\lambda}_2 \geqslant 0, \end{aligned} \tag{6}$$

where $\tilde{\boldsymbol{e}} = (\boldsymbol{0}_{n+neq+m}^T, \boldsymbol{e}_m^T)^T$.

The Newton IP method consists in finding the solution $\Delta \boldsymbol{v}^{(k)}$ of the perturbed Newton equation and in updating the current iterate

$$H'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v} = -H(\boldsymbol{v}^{(k)}) + \rho_k\tilde{\boldsymbol{e}}, \tag{7}$$

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{v}^{(k)} + \alpha_k\Delta\boldsymbol{v}^{(k)}, \tag{8}$$

where the parameter ρ_k goes to 0 when k diverges. The method determines a sequence $\{\boldsymbol{v}^{(k)}\}$ that strictly satisfies the constraints in (6) and the KKT conditions (4) only in the limit.

The crucial points in the analysis of the method are the choices of the parameters ρ_k and α_k and the solution of the linear system (7).

We define $\rho_k = \sigma_k\mu_k$, where $\sigma_k \in [\sigma_{\min}, \sigma_{\max}] \subset (0, 1)$; if the following condition holds:

$$\mu_k \leqslant \mu_k^{(2)} \equiv \frac{\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|}{\sqrt{m}}, \tag{9}$$

then the solution $\Delta\boldsymbol{v}^{(k)}$ of the system (7) is a descent direction for $\|\boldsymbol{H}(\boldsymbol{v})\|^2$ ([4, p. 77]) and, from (7), it satisfies the residual condition of the inexact Newton method ([1]):

$$\|H'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v}^{(k)} + \boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant \eta_k\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|, \tag{10}$$

where the forcing term $\eta_k = \sigma_k \leqslant \sigma_{\max} < 1$. Furthermore, it is easy to prove that $\mu_k^{(1)} \equiv \frac{\boldsymbol{s}^{(k)\mathrm{T}}\boldsymbol{\lambda}_2^{(k)}}{m} \leqslant \mu_k^{(2)}$, where $\mu_k^{(1)}$ is strictly connected with the notion of adherence to the central path which is the basis of IP methods ([10]); then the choice of the perturbation parameter $\mu_k \in [\mu_k^{(1)}, \mu_k^{(2)}]$ assures that $\Delta\boldsymbol{v}^{(k)}$ satisfies the residual condition of the inexact Newton method and it is a descent direction for $\|\boldsymbol{H}(\boldsymbol{v})\|^2$. At the same time, the range of values of the perturbation parameter is enlarged in order to avoid *stagnation* of the current iterate on the boundary of the nonnegative orthant $(\boldsymbol{s}, \boldsymbol{\lambda}_2) \geqslant 0$ that occurs when the value of $\mu_k^{(1)}$ is too small and we are far away from the solution (see [5]).

Now consider the system (7). When the size is large, the computation of the exact solution can be too expensive then the system (7) can be solved *approximately*. We denote again by $\Delta\boldsymbol{v}^{(k)}$ the approximate solution of system (7). If the coefficient matrix has a special structure, an iterative scheme can exploit this feature. Nevertheless, the use of an iterative solver determines the necessity to state an adaptive termination rule so that the accuracy in solving the inner system depends on the quality of the current iterate of the outer method. This means that we can apply an iterative scheme until the final inner residual

$$\boldsymbol{r}^{(k)} = H'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v}^{(k)} + \boldsymbol{H}(\boldsymbol{v}^{(k)}) - \sigma_k\mu_k\tilde{\boldsymbol{e}} \tag{11}$$

satisfies a suitable stopping criterion; under convenient hypotheses reported below, the following criterion assures the global convergence of the whole scheme:

$$\|\boldsymbol{r}^{(k)}\| \leqslant \delta_k\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|. \tag{12}$$

This adaptive termination rule avoids unnecessary inner iterations when we are far from the solution.

Theorem 1. *If* $0 < \sigma_k \leqslant \sigma_{max} < 1$, $0 \leqslant \delta_k \leqslant \delta_{max} < 1$ *and* $\sigma_{max} + \delta_{max} < 1$, *the vector* $\Delta \boldsymbol{v}^{(k)}$, $k \geqslant 0$, *that satisfies* (11) *and* (12) *is a descent direction at* $\boldsymbol{v}^{(k)}$ *for* $\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2$ *and it satisfies the residual condition* (10) *with* $\eta_k = \sigma_k + \delta_k < \sigma_{max} + \delta_{max} < 1$.

Proof. The vector $\Delta \boldsymbol{v}^{(k)}$ is a descent direction if

$$2\boldsymbol{H}(\boldsymbol{v}^{(k)})^{\mathrm{T}} H'(\boldsymbol{v}^{(k)}) \Delta \boldsymbol{v}^{(k)} \leqslant 0.$$

From (11), (12) and (9), we have

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{v}^{(k)})^{\mathrm{T}} H'(\boldsymbol{v}^{(k)}) \Delta \boldsymbol{v}^{(k)} &= \boldsymbol{H}(\boldsymbol{v}^{(k)}) \boldsymbol{r}^{(k)} - \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2 + \sigma_k \mu_k \boldsymbol{H}(\boldsymbol{v}^{(k)})^{\mathrm{T}} \tilde{\boldsymbol{e}} \\ &= (\delta_k - 1)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2 + \sigma_k \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2 \\ &= -(1 - (\delta_k + \sigma_k))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2.\end{aligned}$$

Furthermore,

$$\|H'(\boldsymbol{v}^{(k)}) \Delta \boldsymbol{v}^{(k)} + \boldsymbol{H}(\boldsymbol{v}^{(k)})\| = \|\boldsymbol{r}^{(k)} + \sigma_k \mu_k \tilde{\boldsymbol{e}}\| \leqslant (\delta_k + \sigma_k))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|.$$

This completes the proof. □

The damping parameter α_k has to satisfy the *feasibility* and *centrality conditions* in order to assure the convergence of the method and have to guarantee a *sufficient decreasing* of $\|\boldsymbol{H}(\boldsymbol{v})\|^2$ at each iterate. In order to satisfy all the conditions, the damping parameter α_k is determined by the following sequence of steps.

1. Feasibility condition means that all the iterates $\boldsymbol{v}^{(k)}$ have to belong to the feasible region
$$\{\boldsymbol{v} \in \mathbb{R}^{n+neq+2m} \quad \text{s.t. } s_i > 0 \quad \text{and} \quad \lambda_{2,i} > 0 \quad \forall\, i = 1, \ldots, m\}.$$
So, if $\Delta s_i^{(k)} < 0$ (or $\Delta \lambda_{2,i}^{(k)} < 0$), $\alpha_k{}^{(1)}$ will be chosen such that $s_i^{(k+1)} > 0$ (or $\lambda_{2,i}^{(k+1)} > 0$).
2. Centrality conditions are expressed by the nonnegativity of the following functions ([10], see also [19, p. 402]):
$$\varphi(\alpha) \equiv \min_{i=1,m} \left(S^{(k)}(\alpha) \Lambda_2^{(k)}(\alpha) \boldsymbol{e}_m \right) - \gamma_k \tau_1 \left(\frac{\boldsymbol{s}^{(k)}(\alpha)^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)}(\alpha)}{m} \right) \geqslant 0, \tag{13}$$
$$\psi(\alpha) \equiv \boldsymbol{s}^{(k)}(\alpha)^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)}(\alpha) - \gamma_k \tau_2 \|\boldsymbol{H}_1(\boldsymbol{v}^{(k)}(\alpha))\| \geqslant 0, \tag{14}$$
where $\boldsymbol{s}^{(k)}(\alpha) = \boldsymbol{s}^{(k)} + \alpha \Delta \boldsymbol{s}^{(k)}$ and $\boldsymbol{\lambda}_2{}^{(k)}(\alpha) = \boldsymbol{\lambda}_2{}^{(k)} + \alpha \Delta\, \boldsymbol{\lambda}_2{}^{(k)}$; $\gamma_k \in \left[\frac{1}{2}, 1\right)$.

At each iterate we choose $\tilde{\alpha}_k$ such that conditions (13) and (14) are satisfied $\forall\alpha \in (0, \tilde{\alpha}_k] \subseteq (0, 1]$; then $\alpha_k^{(2)} = \min\left\{\tilde{\alpha}_k, \alpha_k^{(1)}\right\}$.

In order to satisfy inequalities (13) and (14) in the initial iterate, we have $\tau_1 \leqslant \frac{\min_{i=1,m}\left(S^{(0)}\Lambda_2^{(0)}\boldsymbol{e}_m\right)}{\left(\frac{\boldsymbol{s}^{(0)\mathrm{T}}\boldsymbol{\lambda}_2^{(0)}}{m}\right)}$, and $\tau_2 \leqslant \frac{\boldsymbol{s}^{(0)\mathrm{T}}\boldsymbol{\lambda}_2^{(0)}}{\|\boldsymbol{H}_1(\boldsymbol{v}^{(0)})\|}$, where we assume $\boldsymbol{s}^{(0)} > 0$, $\boldsymbol{\lambda}_2^{(0)} > 0$.

3. On the other hand, the final parameter α_k must be selected so that the outer iterative scheme is convergent to a solution of the system (4). For this aim, following the theory on inexact Newton methods ([9]), a *minimum reduction algorithm*, consisting in a line–search strategy with backtracking technique, is included in the method. Thus, $\alpha_k^{(2)}$ is reduced using the following strategy:
   - Set $\beta \in \left(0, \frac{1}{2}\right]$, $\theta \in (0, 1)$, $\alpha_k = \alpha_k^{(2)}$;
   - *while* $\|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \alpha_k\Delta\boldsymbol{v}^{(k)})\| > (1-\beta\alpha_k(1-(\sigma_k + \delta_k)))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|$

   $\alpha_k = \theta\alpha_k$
   end while

   If the backtracking procedure terminates after $\bar{t}$ steps, then $\alpha_k = \theta^{\bar{t}}\alpha_k^{(2)}$; furthermore, if $\alpha_k^{(2)}$ is bounded below by a scalar greater than zero, say $\alpha^{(2)}$, then also α_k is bounded below by a positive scalar, say $\bar{\alpha} > 0$, i.e. $\alpha_k \geqslant \bar{\alpha} > 0$, and the vector $\alpha_k\Delta\boldsymbol{v}^{(k)}$, $k \geqslant 0$, satisfies the residual condition of the inexact Newton method

   $$\|H'(\boldsymbol{v}^{(k)})\alpha_k\Delta\boldsymbol{v}^{(k)} + \boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant \eta_k\|H(\boldsymbol{v}^{(k)})\|,$$

   where $\eta_k = 1 - \alpha_k(1 - (\sigma_k + \delta_k)) \leqslant \bar{\eta} < 1$. Indeed, from (11), (12) and (9), we have

   $$\begin{aligned}\|H'(\boldsymbol{v}^{(k)})\alpha_k\Delta\boldsymbol{v}^{(k)} + \boldsymbol{H}(\boldsymbol{v}^{(k)})\| &\leqslant \|\alpha_k(-\boldsymbol{H}(\boldsymbol{v}^{(k)}) + \boldsymbol{r}^{(k)} + \sigma_k\mu_k\tilde{\boldsymbol{e}}) + \boldsymbol{H}(\boldsymbol{v}^{(k)})\| \\ &\leqslant (1 - \alpha_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + \alpha_k\|\boldsymbol{r}^{(k)} + \sigma_k\mu_k\tilde{\boldsymbol{e}}\| \\ &\leqslant (1 - \alpha_k(1 - (\sigma_k + \delta_k)))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|.\end{aligned}$$

Now we focus our attention to the solution of the linear system (7) that, by omitting the iteration index k, can be written as

$$\begin{cases} Q\Delta\boldsymbol{x} + B\Delta\boldsymbol{\lambda}_1 + C\Delta\boldsymbol{\lambda}_2 = -\boldsymbol{\alpha}, \\ B^{\mathrm{T}}\Delta\boldsymbol{x} = -\boldsymbol{\beta}, \\ C^{\mathrm{T}}\Delta\boldsymbol{x} + \Delta\boldsymbol{s} = -\boldsymbol{\gamma}, \\ S\Delta\boldsymbol{\lambda}_2 + \Lambda_2\Delta\boldsymbol{s} = -\boldsymbol{\theta} + \rho\boldsymbol{e}_m. \end{cases}$$

From the complementarity equations we can deduce

$$\Delta\boldsymbol{s} = \Lambda_2^{-1}[-S\Delta\boldsymbol{\lambda}_2 - \boldsymbol{\theta} + \rho\boldsymbol{e}_m]$$

and then the system (7) can be rewritten in *reduced form*

$$\begin{pmatrix} Q & B & C \\ B^{\mathrm{T}} & 0 & 0 \\ C^{\mathrm{T}} & 0 & -\Lambda_2^{-1}S \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{\lambda}_1 \\ \Delta \boldsymbol{\lambda}_2 \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\alpha} \\ -\boldsymbol{\beta} \\ \boldsymbol{g}_2(x) - \rho \Lambda_2^{-1} \boldsymbol{e}_m \end{pmatrix}. \tag{15}$$

By a further substitution from the third block equation

$$\Delta \boldsymbol{\lambda}_2 = S^{-1}[\Lambda_2 C^{\mathrm{T}} \Delta \boldsymbol{x} + \Lambda_2 \boldsymbol{\gamma} - \boldsymbol{\theta} + \rho \boldsymbol{e}_m],$$

the system can be written in *condensed form*

$$\begin{pmatrix} A & B \\ B^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{\lambda}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{q} \end{pmatrix}, \tag{16}$$

with

$$A = Q + CS^{-1}\Lambda_2 C^{\mathrm{T}},$$
$$\boldsymbol{c} = -\boldsymbol{\alpha} - C(\boldsymbol{x})S^{-1}[\Lambda_2 \boldsymbol{\gamma} - \boldsymbol{\theta} + \rho \boldsymbol{e}_m],$$
$$\boldsymbol{q} = -\boldsymbol{\beta}.$$

Let the vector $\boldsymbol{r}^{(k)}$ of (11) be partitioned commensurately as $\boldsymbol{v}^{(k)}$ and $\boldsymbol{H}(\boldsymbol{v}^{(k)})$:

$$\boldsymbol{r}^{(k)} = \begin{pmatrix} \boldsymbol{r}_1^{(k)} \\ \boldsymbol{r}_2^{(k)} \\ \boldsymbol{r}_3^{(k)} \\ \boldsymbol{r}_4^{(k)} \end{pmatrix}; \tag{17}$$

when we solve approximately the system (7) in the form (15), we have $\boldsymbol{r}_4^{(k)} = 0$ while in the form (16), we have $\boldsymbol{r}_3^{(k)} = \boldsymbol{r}_4^{(k)} = 0$. In other words, in both cases, the block related to the complementarity equations is solved exactly.

3. Analysis of the convergence

We state the conditions such that the Newton IP method can be viewed as an inexact Newton method ([9]) with restriction on the sign of some variables.

Given $\epsilon \geqslant 0$, we define

$$\begin{aligned} \Omega(\epsilon) = \{\boldsymbol{v} : 0 \leqslant \epsilon \leqslant \|\boldsymbol{H}(\boldsymbol{v})\|^2 \\ \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(0)})\|^2, \quad \text{s.t. (13) and (14) hold}\}. \end{aligned} \tag{18}$$

$\Omega(\epsilon)$ is a closed set.

Let assume that the following conditions hold ([6], see also [10]):

C1 in $\Omega(0)$, $f(\boldsymbol{x})$, $\boldsymbol{g}_1(\boldsymbol{x})$, $\boldsymbol{g}_2(\boldsymbol{x})$ are twice continuously differentiable; the gradients of the equality constraints are linearly independent and $H_1'(\boldsymbol{v})$ is Lipschitz continuous;

C2 the sequences $\{\boldsymbol{x}^{(k)}\}$ and $\{\boldsymbol{\lambda}_2^{(k)}\}$ are bounded;
C3 in any compact subset of $\Omega(0)$ where $\boldsymbol{s}$ is bounded away from zero, the matrix $H'(\boldsymbol{v})$ is nonsingular.

In general, in literature, the condition C3 is replaced by a sufficient condition to assure that C3 holds.

The boundedness of the sequence $\{\boldsymbol{x}^{(k)}\}$ can be assured by enforcing box constraints $-l_i \leqslant x_i^{(k)} \leqslant l_i$ for sufficiently large $l_i > 0$, $i = 1,\ldots,n$.

Theorem 2. *Let* $\{\boldsymbol{v}^{(k)}\}$ *be a sequence generated as described in the previous section.*

If $\boldsymbol{v}^{(k)} \in \Omega(\epsilon)$, $\epsilon > 0$, *then*

(a) $(\boldsymbol{s}^{(k)})^T \boldsymbol{\lambda}_2{}^{(k)}$, $s_i^{(k)} \lambda_{2,i}^{(k)}$, $i = 1,\ldots,m$, *are bounded above and below away from zero for any* $k \geqslant 0$; $\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\|$ *is bounded above for any* $k \geqslant 0$;
(b) *if C1 and C2 hold, then* $\{\boldsymbol{v}^{(k)}\}$ *is bounded above and* $\boldsymbol{s}^{(k)}$ *and* $\boldsymbol{\lambda}_2^{(k)}$ *are componentwise bounded away from zero*;
(c) *if C1, C2 and C3 hold, then the sequence of matrices* $\{H'(\boldsymbol{v}^{(k)})^{-1}\}$ *is bounded*;
(d) *if C1, C2 and C3 hold, then the sequence* $\{\Delta\boldsymbol{v}^{(k)}\}$ *is bounded.*

Proof. (a) The above boundedness of $s_i^{(k)} \lambda_{2,i}^{(k)}$, $i = 1,\ldots,m$, and $(\boldsymbol{s}^{(k)})^{\mathrm{T}} \boldsymbol{\lambda}_2{}^{(k)}$ follows from the inequality

$$s_i^{(k)} \lambda_{2,i}^{(k)} \leqslant \left(\boldsymbol{s}^{(k)}\right)^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)} = \|S^{(k)} \Lambda_2^{(k)} \boldsymbol{e}_m\|_1 \leqslant \sqrt{m} \|S^{(k)} \Lambda_2^{(k)} \boldsymbol{e}_m\| \\ = \sqrt{m} \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant \sqrt{m} \|\boldsymbol{H}(\boldsymbol{v}^{(0)})\|. \tag{19}$$

Furthermore in $\Omega(\epsilon)$, $\epsilon > 0$, from (13) and (14), we have $s_i^{(k)} \lambda_{2,i}^{(k)} > 0$ for any $k \geqslant 0$ and $i = 1,\ldots,m$. Indeed, if we assume that $s_j^{(k)} \lambda_{2,j}^{(k)} = 0$ for some j, then $(\boldsymbol{s}^{(k)})^{\mathrm{T}} \boldsymbol{\lambda}_2{}^{(k)} = 0$ and $\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\| = 0$; but this contradicts $\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \geqslant \epsilon$, $\epsilon > 0$. From the inequality

$$\epsilon \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant \|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\| + \|S^{(k)} \Lambda_2^{(k)} \boldsymbol{e}_m\| \\ \leqslant ((\boldsymbol{s}^{(k)})^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)})/(\gamma_k \tau_2) + \|S^{(k)} \Lambda_2^{(k)} \boldsymbol{e}_m\|_1 = (1 + 1/(\gamma_k \tau_2))(\boldsymbol{s}^{(k)})^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)}, \tag{20}$$

it follows that, for $k \geqslant 0$ and $i = 1,\ldots,m$,

$$(\boldsymbol{s}^{(k)})^{\mathrm{T}} \boldsymbol{\lambda}_2^{(k)} \geqslant \epsilon\tau_2/(\tau_2 + 2) \tag{21}$$

and, from (13),

$$s_i^{(k)} \lambda_{2,i}^{(k)} \geqslant \epsilon\tau_1\tau_2/(2m(\tau_2 + 2)). \tag{22}$$

Finally,

$$\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\| \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(0)})\|.$$

(b) From assumptions C1 and C2 and from (3), we have

$$\|\boldsymbol{\alpha}^{(k)}\| = \|\nabla f(\boldsymbol{x}^{(k)}) + B^{(k)}\boldsymbol{\lambda}_1^{(k)} + C^{(k)}\boldsymbol{\lambda}_2^{(k)}\| \leqslant \|\boldsymbol{H}(\boldsymbol{x}^{(k)})\|.$$

Then, since $B^{(k)}$ is a full column–rank matrix, we can write

$$\boldsymbol{\lambda}_1^{(k)} = (B^{(k)^{\mathrm{T}}}B^{(k)})^{-1}B^{(k)^{\mathrm{T}}}(-\nabla f(\boldsymbol{x}^{(k)}) - C^{(k)}\boldsymbol{\lambda}_2^{(k)} + \boldsymbol{\alpha}^{(k)}).$$

For C1 and C2, the sequence $\{\boldsymbol{\lambda}_1^{(k)}\}$ is bounded.

Furthermore,

$$\|\boldsymbol{s}^{(k)}\| \leqslant \|\boldsymbol{s}^{(k)} - \boldsymbol{g}_2(\boldsymbol{x}^{(k)})\| + \|\boldsymbol{g}_2(\boldsymbol{x}^{(k)})\| \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + \|\boldsymbol{g}_2(\boldsymbol{x}^{(k)})\|.$$

Then the sequence $\{\boldsymbol{s}^{(k)}\}$ is bounded.

Since $s_i^{(k)}\lambda_{2,i}^{(k)}$ are bounded below away from zero and $\boldsymbol{s}^{(k)}$ is bounded above, for any k, it follows that $\boldsymbol{\lambda}_2^{(k)}$ is bounded below away from zero. Analogously, for the same argument, $\boldsymbol{s}^{(k)}$ is bounded away from zero.

(c) Rearranging the rows and the columns of the matrix H' $(\boldsymbol{v}^{(k)})$, we obtain the following matrix

$$\begin{bmatrix} \Lambda_2^{(k)} & S^{(k)} & 0 & 0 \\ I & 0 & C^{(k)^{\mathrm{T}}} & 0 \\ 0 & C^{(k)} & Q^{(k)} & B^{(k)} \\ 0 & 0 & B^{(k)^{\mathrm{T}}} & 0 \end{bmatrix}. \tag{23}$$

Since $\boldsymbol{s}^{(k)}$ and $\boldsymbol{\lambda}_2^{(k)}$ are bounded above and componentwise below away from 0, the matrix (23) can be factorized in the form $L^{(k)}U^{(k)}$, where $L^{(k)}$ is the matrix

$$\begin{bmatrix} I & 0 & 0 & 0 \\ \Lambda_2^{-1} & I & 0 & 0 \\ 0 & -C^{(k)}E^{(k)^{-1}} & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \tag{24}$$

and $U^{(k)}$ is the matrix

$$\begin{bmatrix} \Lambda_2^{(k)} & S^{(k)} & 0 & 0 \\ 0 & -E^{(k)} & C^{(k)^{\mathrm{T}}} & 0 \\ 0 & 0 & F^{(k)} & B^{(k)} \\ 0 & 0 & B^{(k)^{\mathrm{T}}} & 0 \end{bmatrix}, \tag{25}$$

with $E^{(k)} = \Lambda_2^{(k)^{-1}}S^{(k)}$ and $F^{(k)} = Q^{(k)} + C^{(k)}E^{(k)-1}C^{(k)\mathrm{T}}$. Since $L^{(k)}$ and $H'(\boldsymbol{v}^{(k)})$ are nonsingular bounded matrices, the block triangular matrix $U^{(k)}$ is a nonsingular matrix with nonsingular and bounded diagonal blocks. The inverse of the matrix $H'(\boldsymbol{v}^{(k)})$ is given by $(U^{(k)})^{-1}(L^{(k)})^{-1}$. Since all the blocks of the matrices $(U^{(k)})^{-1}$ and $(L^{(k)})^{-1}$ are bounded, then $H'(\boldsymbol{v}^{(k)})^{-1}$ is also bounded in $\Omega(\epsilon)$, $\epsilon > 0$, i.e.

$$\|H'(\boldsymbol{v}^{(k)})^{-1}\| \leqslant \overline{M} \tag{26}$$

for $\boldsymbol{v}^{(k)} \in \Omega(\epsilon)$, $\epsilon > 0$ and for $k \geqslant 0$, with $\overline{M}$ a positive scalar.

(d) Since (11), $\Delta\boldsymbol{v}^{(k)}$ has the following form:

$$\Delta\boldsymbol{v}^{(k)} = H'(\boldsymbol{v}^{(k)})^{-1}(-\boldsymbol{H}(\boldsymbol{v}^{(k)}) + \boldsymbol{r}^{(k)} + \sigma_k\mu_k\tilde{\boldsymbol{e}}). \tag{27}$$

From (26), (18), (9) and (12), we have that

$$\|\Delta\boldsymbol{v}^{(k)}\| \leqslant \overline{M}(1 + \delta_k + \sigma_k)\|\boldsymbol{H}(\boldsymbol{v}^{(0)})\| < 2\overline{M}\|\boldsymbol{H}(\boldsymbol{v}^{(0)})\|,$$

because $\delta_k + \sigma_k \leqslant \delta_{\max} + \sigma_{\max} < 1$. Then the proof is completed. □

In the following we analyze the three steps for the computation of the damping parameter α_k and we show that it is uniformly bounded away from zero (see Section 2).

In the step 1., it is easy to see that $\alpha_k^{(1)}$ is bounded away from zero, i.e. $\alpha_k^{(1)} \geqslant \alpha^{(1)} > 0$, since we set

$$\alpha_k^{(1)} = \min\left\{\min_{\Delta s_i^{(k)}<0} \frac{-s_i^{(k)}}{\Delta s_i^{(k)}}, \min_{\Delta\lambda_{2,i}^{(k)}<0} \frac{-\lambda_{2,i}^{(k)}}{\Delta\lambda_{2,i}^{(k)}}, 1\right\},$$

where, for any iteration k, $s_i^{(k)}$ and $\lambda_{2,i}^{(k)}$ are bounded away from zero and $\Delta s_i^{(k)}$ and $\Delta\lambda_{2,i}^{(k)}$ are bounded.

Now, we analyze the damping parameter in step 2.; the following theorem (see [11]) shows the existence of two positive numbers $\hat{\alpha}_k^{(2)}$ and $\check{\alpha}_k^{(2)}$ such that the centrality functions $\varphi(\alpha)$ and $\psi(\alpha)$ are nonnegative for $\alpha \in (0, \hat{\alpha}_k^{(2)}]$ and for $\alpha \in (0, \check{\alpha}_k^{(2)}]$ respectively.

Theorem 3. *Let $\{\boldsymbol{v}^{(k)}\}$ be a sequence generated as described in the previous section; let us also assume $\sigma_k \in [\sigma_{min}, \sigma_{max}] \subset (0,1)$ and $\delta_k \in [0, \delta_{max}] \subset [0,1)$, and*

$$\sigma_k > \delta_k(1 + \gamma_k\tau_2). \tag{28}$$

Then, if $\varphi^{(k)}(0) \geqslant 0$, there exists a positive number $\hat{\alpha}_k^{(2)} > 0$, such that $\varphi^{(k)}(\alpha) \geqslant 0$ for all $\alpha \in (0, \hat{\alpha}_k^{(2)}]$.

Then, if $\psi^{(k)}(0) \geqslant 0$, there exists a positive number $\check{\alpha}_k^{(2)} > 0$, such that $\psi^{(k)}(\alpha) \geqslant 0$ for all $\alpha \in (0, \check{\alpha}_k^{(2)}]$.

Proof. Set

$$N_i^{(k)} = \left|\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)} - \frac{\gamma_k\tau_1}{m}\Delta\boldsymbol{s}^{(k)\mathrm{T}}\Delta\boldsymbol{\lambda}_2^{(k)}\right| \quad i = 1, \ldots, m.$$

The fourth block equations of the linear system (7) in componentwise is

$$s_i^{(k)}\Delta\lambda_{2,i}^{(k)} + \lambda_{2,i}^{(k)}\Delta s_i^{(k)} = -s_i^{(k)}\lambda_{2,i}^{(k)} + \sigma_k\mu_k. \tag{29}$$

Summing for any $i = 1,\dots,m$, we have

$$\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)} + \boldsymbol{\lambda}_2^{(k)^{\mathrm{T}}}\Delta\boldsymbol{s}^{(k)} = -\boldsymbol{s}^{(k)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(k)} + m\sigma_k\mu_k. \tag{30}$$

Thus, for $\alpha \in (0,1]$, we can define

$$\begin{aligned}\varphi_i^{(k)}(\alpha) = {}& \left(s_i^{(k)} + \alpha\Delta s_i^{(k)}\right)\left(\lambda_{2,i}^{(k)} + \alpha\Delta\lambda_{2,i}^{(k)}\right) \\ & - \frac{\tau_1\gamma_k}{m}\left(\boldsymbol{s}^{(k)} + \alpha\Delta\boldsymbol{s}^{(k)}\right)^{\mathrm{T}}\left(\boldsymbol{\lambda}_2^{(k)} + \alpha\Delta\boldsymbol{\lambda}_2^{(k)}\right).\end{aligned}$$

By easy computation and by using (29) and (30), we can deduce

$$\begin{aligned}\varphi_i^{(k)}(\alpha) = {}& (1-\alpha)\left[s_i^{(k)}\lambda_{2,i}^{(k)} - \frac{\tau_1\gamma_k}{m}\boldsymbol{s}^{(k)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(k)}\right] + \alpha\sigma_k\mu_k(1-\tau_1\gamma_k) \\ & + \alpha^2\left(\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)} - \frac{\tau_1\gamma_k}{m}\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)}\right).\end{aligned}$$

Hence,

$$\begin{aligned}\varphi_i^{(k)}(\alpha) = {}& (1-\alpha)\varphi_i^{(k)}(0) + \alpha\sigma_k\mu_k(1-\tau_1\gamma_k) \\ & + \alpha^2\left(\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)} - \frac{\tau_1\gamma_k}{m}\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)}\right).\end{aligned} \tag{31}$$

Since $\varphi^{(k)}(0) \geqslant 0$, we have $\varphi_i^{(k)}(0) \geqslant 0$. Then

$$\begin{aligned}(1-\alpha)\varphi_i^{(k)}(0) = {}& \varphi_i^{(k)}(\alpha) - \alpha\sigma_k\mu_k(1-\tau_1\gamma_k) \\ & - \alpha^2\left(\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)} - \frac{\tau_1\gamma_k}{m}\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)}\right).\end{aligned}$$

Thus

$$\begin{aligned}\varphi_i^{(k)}(\alpha) & \geqslant \alpha\sigma_k\mu_k(1-\tau_1\gamma_k) + \alpha^2\left(\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)} - \frac{\tau_1\gamma_k}{m}\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)}\right) \\ & \geqslant \alpha\sigma_k\mu_k(1-\tau_1\gamma_k) - \alpha^2 N_i^{(k)}.\end{aligned}$$

Set $N^{(k)} = \max_{i=1,\dots,m} N_i^{(k)}$; for any α such that

$$\alpha \geqslant ((1-\tau_1\gamma_k)\sigma_k\mu_k)/N^{(k)} > 0, \tag{32}$$

we have $\varphi^{(k)}(\alpha) \geqslant 0$. Thus we define

$$\hat{\alpha}_k^{(2)} = \max_{\alpha\in(0,1]}\{\alpha : \varphi^{(k)}(t) \geqslant 0,\ \forall t \leqslant \alpha\}.$$

We prove now the second part of the theorem.

By assumptions C1, we have that $H_1'(\boldsymbol{v})$ is Lipschitz continuous with Lipschitz constant Γ.

Set

$$M^{(k)} = \left| \Delta \boldsymbol{s}^{(k)^{\mathrm{T}}} \Delta \boldsymbol{\lambda}_2^{(k)} - \gamma_k \tau_2 \frac{\Gamma}{2} \|\Delta \boldsymbol{v}^{(k)}\|^2 \right|$$

and let $\hat{\boldsymbol{r}}^{(k)}$ be the vector composed by the first three block components of the vector $\boldsymbol{r}^{(k)}$ defined in (11) and (17). By the mean value theorem for vector valued functions (e.g. see [2, p. 74]), we can write for $\alpha \in (0,1]$

$$\begin{aligned}
\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)}) &= \boldsymbol{H}_1(\boldsymbol{v}^{(k)}) + \alpha H_1'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v}^{(k)} \\
&\quad + \alpha\left(\int_0^1 (H_1'(\boldsymbol{v}^{(k)} + \xi\alpha\Delta\boldsymbol{v}^{(k)}) - H_1'(\boldsymbol{v}^{(k)}))\,\mathrm{d}\xi\right)\Delta\boldsymbol{v}^{(k)}, \\
&= (1-\alpha)\boldsymbol{H}_1(\boldsymbol{v}^{(k)}) + \alpha\hat{\boldsymbol{r}}^{(k)} \\
&\quad + \alpha\left(\int_0^1 (H_1'(\boldsymbol{v}^{(k)} + \xi\alpha\Delta\boldsymbol{v}^{(k)}) - H_1'(\boldsymbol{v}^{(k)}))\,\mathrm{d}\xi\right)\Delta\boldsymbol{v}^{(k)}.
\end{aligned} \tag{33}$$

From the Lipschitz continuity for the derivative of $\boldsymbol{H}_1(\boldsymbol{v})$, we obtain

$$\begin{aligned}
\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\| &\leqslant (1-\alpha)\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\| + \alpha\|\hat{\boldsymbol{r}}^{(k)}\| \\
&\quad + \alpha\left(\int_0^1 \Gamma\|\xi\alpha\Delta\boldsymbol{v}^{(k)}\|\,\mathrm{d}\xi\right)\|\Delta\boldsymbol{v}^{(k)}\|,
\end{aligned}$$

or, by (12)

$$\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\| \leqslant (1-\alpha)\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)})\| + \alpha\delta_k\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + \frac{\Gamma}{2}\alpha^2\|\Delta\boldsymbol{v}^{(k)}\|^2. \tag{34}$$

From the definition of $\psi^{(k)}(\alpha)$ and by using (30), we have that

$$\begin{aligned}
\psi^{(k)}(\alpha) &= \boldsymbol{s}^{(k)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(k)} + \alpha(-\boldsymbol{s}^{(k)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(k)} + \sigma_k\mu_k m) + \alpha^2\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)} \\
&\quad - \gamma_k\tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\|.
\end{aligned}$$

If we multiply (34) by $-\gamma_k\tau_1$, changing the sign, then we have a lower bound of $-\gamma_k\tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\|$ that gives

$$\begin{aligned}
\psi^{(k)}(\alpha) &\geqslant (1-\alpha)\psi^{(k)}(0) + \alpha(\sigma_k\mu_k m - \gamma_k\tau_2\delta_k\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|) \\
&\quad + \alpha^2\left(\Delta\boldsymbol{s}^{(k)^{\mathrm{T}}}\Delta\boldsymbol{\lambda}_2^{(k)} - \gamma_k\tau_2\frac{\Gamma}{2}\|\Delta\boldsymbol{v}^{(k)}\|^2\right).
\end{aligned}$$

Then, by the hypothesis $\psi^{(k)}(0) \geqslant 0$, $\mu_k \geqslant \frac{\left(\boldsymbol{s}^{(k)}\right)^{\mathrm{T}}\boldsymbol{\lambda}_2^{(k)}}{m}$ and (20), we obtain

$$\psi^{(k)}(\alpha) \geqslant \alpha\left(\left(\frac{\sigma_k}{1+\gamma_k\tau_2} - \delta_k\right)\gamma_k\tau_2\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - \alpha M^{(k)}\right).$$

If condition (28) holds, then for any α such that

$$\alpha \geqslant \left(\frac{\sigma_k}{1+\gamma_k\tau_2} - \delta_k\right)\gamma_k\tau_2\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|/M^{(k)} > 0, \tag{35}$$

we have $\psi^{(k)}(\alpha) \geqslant 0$. Thus we define

$$\check{\alpha}_k^{(2)} = \max_{\alpha\in(0,1]}\{\alpha : \psi^{(k)}(t) \geqslant 0,\ \forall t \leqslant \alpha\}.$$

This completes the proof. $\square$

Let us define

$$\tilde{\alpha}_k = \min\left\{\hat{\alpha}_k^{(2)}, \check{\alpha}_k^{(2)}, 1\right\} \in (0,1];$$

then, under the hypotheses of Theorem 2, $N^{(k)}$ and $M^{(k)}$ are uniformly bounded and

$$\tilde{\alpha}_k \geqslant \tilde{\alpha} > 0.$$

Consequently, we have

$$\alpha_k^{(2)} \equiv \min\{\tilde{\alpha}_k, \alpha_k^{(1)}\} \geqslant \alpha^{(2)} \equiv \min\{\tilde{\alpha}, \alpha^{(1)}\} > 0.$$

To select the final value of the damping parameter at the iteration k, in step 3. we perform the backtracking technique described in [9] until an *acceptable*

$$\alpha_k = \theta^{\bar{t}}\alpha_k^{(2)}$$

is found, where $\bar{t}$ is the smallest nonnegative integer such that α_k satisfies the backtracking condition

$$\|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \alpha_k\Delta\boldsymbol{v}^{(k)})\| \leqslant (1 - \beta\alpha_k(1 - (\sigma_k + \delta_k)))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|, \tag{36}$$

with $\theta, \beta \in (0,1)$.

We have to prove now that $\bar{t}$ is a finite number independent on k.

Theorem 4. *Under the hypotheses of Theorems* 2, 3, *the while-loop in step* 3. *terminates in a finite number of steps.*

Proof. From (29), (33) and (11), we have, for $\alpha \in (0,1]$ and for $i = 1,\ldots,m$:

$$(s_i^{(k)} + \alpha\Delta s_i^{(k)})(\lambda_{2,i}^{(k)} + \alpha\Delta\lambda_{2,i}^{(k)}) = s_i^{(k)}\lambda_{2,i}^{(k)} + \alpha(-s_i^{(k)}\lambda_{2,i}^{(k)} + \sigma_k\mu_k) + \alpha^2\Delta s_i^{(k)}\Delta\lambda_{2,i}^{(k)}$$

and

$$\boldsymbol{H}_1(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)}) = (1-\alpha)\boldsymbol{H}_1(\boldsymbol{v}^{(k)}) + \alpha\hat{\boldsymbol{r}}^{(k)} + \alpha\left(\int_0^1 \left(H_1'\left(\boldsymbol{v}^{(k)} + \xi\alpha\Delta\boldsymbol{v}^{(k)}\right) - H_1'(\boldsymbol{v}^{(k)})\right) \mathrm{d}\xi\right)\Delta\boldsymbol{v}^{(k)}.$$

We can write

$$\begin{aligned}
\boldsymbol{H}\left(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)}\right) &= \begin{pmatrix} \boldsymbol{H}_1\left(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)}\right) \\ \left(S^{(k)} + \alpha\Delta S^{(k)}\right)\left(\Lambda_2^{(k)} + \alpha\Delta\Lambda_2^{(k)}\right) \end{pmatrix} \\
&= (1-\alpha)\begin{pmatrix} \boldsymbol{H}_1(\boldsymbol{v}^{(k)}) \\ S^{(k)}\Lambda_2^{(k)}\boldsymbol{e}_m \end{pmatrix} + \alpha\begin{pmatrix} \hat{\boldsymbol{r}}^{(k)} \\ 0 \end{pmatrix} + \alpha\begin{pmatrix} 0 \\ \sigma_k\mu_k\boldsymbol{e}_m \end{pmatrix} \\
&\quad + \alpha\begin{pmatrix} \left(\int_0^1 \left(H_1'\left(\boldsymbol{v}^{(k)} + \xi\alpha\Delta\boldsymbol{v}^{(k)}\right) - H_1'(\boldsymbol{v}^{(k)})\right)\mathrm{d}\xi\right)\Delta\boldsymbol{v}^{(k)} \\ 0 \end{pmatrix} \\
&\quad + \alpha^2\begin{pmatrix} 0 \\ \Delta S^{(k)}\Delta\Lambda_2^{(k)}\boldsymbol{e}_m \end{pmatrix}.
\end{aligned}$$

Thus

$$\begin{aligned}
\|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\| &\leqslant (1-\alpha)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + \alpha\|\boldsymbol{r}^{(k)}\| + \alpha\sigma_k\mu_k\|\boldsymbol{e}_m\| \\
&\quad + \alpha\|\Delta\boldsymbol{v}^{(k)}\|\int_0^1 \|H_1'(\boldsymbol{v}^{(k)} + \xi\alpha\Delta\boldsymbol{v}^{(k)}) - H_1'(\boldsymbol{v}^{(k)})\|\,\mathrm{d}\xi \\
&\quad + \alpha^2\|\Delta S^{(k)}\Delta\Lambda_2^{(k)}\boldsymbol{e}_m\|.
\end{aligned}$$

From the Lipschitz continuity for the derivative of $\boldsymbol{H}_1(\boldsymbol{v})$, from (12), we have

$$\begin{aligned}
\|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\| &\leqslant (1-\alpha)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + \alpha(\sigma_k + \delta_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \\
&\quad + \alpha^2\left(\|\Delta S^{(k)}\Delta\Lambda_2^{(k)}\boldsymbol{e}_m\| + \frac{\Gamma}{2}\|\Delta\boldsymbol{v}^{(k)}\|^2\right).
\end{aligned}$$

Therefore, we can affirm that

$$\begin{aligned}
&(1 - \beta\alpha(1 - (\sigma_k + \delta_k)))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - \|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \alpha\Delta\boldsymbol{v}^{(k)})\| \\
&\quad \geqslant (1-\beta)\alpha(1 - (\sigma_k + \delta_k))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - \left(\alpha^2\left(1 + \frac{\Gamma}{2}\right)\|\Delta\boldsymbol{v}^{(k)}\|^2\right)
\end{aligned}$$

is nonnegative for $\alpha \in (0, \hat{\alpha}]$ with

$$\hat{\alpha} = \frac{(1-\beta)(1 - (\sigma_k + \delta_k))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|}{\left(1 + \frac{\Gamma}{2}\right)\|\Delta\boldsymbol{v}^{(k)}\|^2} > 0.$$

Since $\hat{\alpha}$ is bounded away from zero in $\Omega(\epsilon)$, $\epsilon > 0$, it is possible to find a nonnegative integer $\bar{t}$ such that $0 < \theta^{\bar{t}}\alpha_k^{(2)} \leqslant \min\{\hat{\alpha}, 1\}$; then the value $\alpha_k = \theta^{\bar{t}}\alpha_k^{(2)}$ is bounded below by a strictly positive number, say $\breve{\alpha}$.

This completes the proof. □

Set $\bar{\alpha} = \min\left\{\alpha^{(2)}, \breve{\alpha}\right\}$, we observe that, since

$$(1 - \beta\alpha_k(1 - (\sigma_k + \delta_k))) \leqslant (1 - \beta\bar{\alpha}(1 - (\sigma_{\max} + \delta_{\max}))) < 1,$$

inequality (36) asserts that

$$\|\boldsymbol{H}(\boldsymbol{v}^{(k+1)})\| < \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|. \tag{37}$$

We prove now the following result (see [11]).

Proposition 1. *Let $\varphi(\alpha)$ and $\psi(\alpha)$ be the centrality functions defined in (13) and (14); set*

$$\tau_1 = \frac{\min\limits_{i=1,m}(S^{(0)}\Lambda_2^{(0)}\boldsymbol{e}_m)}{\left(\frac{\boldsymbol{s}^{(0)\mathrm{T}}\boldsymbol{\lambda}_2^{(0)}}{m}\right)}; \quad \tau_2 = \frac{\boldsymbol{s}^{(0)\mathrm{T}}\boldsymbol{\lambda}_2^{(0)}}{\|\boldsymbol{H}_1(\boldsymbol{v}^{(0)})\|}$$

and let be given a sequence of parameters $\{\gamma_k\}$ with

$$1 > \gamma_0 \geqslant \gamma_1 \geqslant \cdots \geqslant \gamma_k \geqslant \gamma_{k+1} \geqslant \cdots \geqslant \frac{1}{2}.$$

Furthermore, since $\boldsymbol{s}^{(0)} > 0$, $\boldsymbol{\lambda}_2^{(0)} > 0$, then

$$\varphi^{(k)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \hat{\alpha}_k^{(2)}\right],$$

$$\psi^{(k)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \check{\alpha}_k^{(2)}\right]$$

for any $k = 0, 1, \ldots$

Proof. This results shows that the strict feasibility of the initial vectors $\boldsymbol{s}^{(0)} > 0$ and $\boldsymbol{\lambda}_2^{(0)} > 0$ is sufficient to guarantee the positivity of the centrality functions $\varphi(\alpha)$ and $\psi(\alpha)$ at each iterate k.

Indeed, for $k = 0$, the definitions of τ_1 and τ_2 give

$$\varphi^{(0)}(0) = (1-\gamma_0)\min_i(S^{(0)}\Lambda_2^{(0)}\boldsymbol{e}_m) > 0,$$

$$\psi^{(0)}(0) = (1-\gamma_0)\boldsymbol{s}^{(0)\mathrm{T}}\boldsymbol{\lambda}_2^{(0)} > 0.$$

Theorem 3 assures that there exist $\hat{\alpha}_0^{(2)} > 0$ and $\check{\alpha}_0^{(2)} > 0$ such that

$$\varphi^{(0)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \hat{\alpha}_0^{(2)}\right],$$

$$\psi^{(0)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \check{\alpha}_0^{(2)}\right].$$

Thus, we have $\varphi^{(0)}(\alpha_0) \geqslant 0$ and $\psi^{(0)}(\alpha_0) \geqslant 0$, where α_0 is the final value of the damping parameter obtained after the backtracking procedure.

For $k = 1$, the centrality functions are

$$\varphi^{(1)}(\alpha) = \min_{i=1,\ldots,m}\left(S^{(1)}(\alpha)\Lambda_2^{(1)}(\alpha)\boldsymbol{e}_m\right) - \gamma_1\tau_1\left(\frac{\boldsymbol{s}^{(1)}(\alpha)^{\mathrm{T}}\boldsymbol{\lambda}_2^{(1)}(\alpha)}{m}\right),$$

$$\psi^{(1)}(\alpha) = \boldsymbol{s}^{(1)}(\alpha)^{\mathrm{T}}\boldsymbol{\lambda}_2^{(1)}(\alpha) - \gamma_1\tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(1)}(\alpha))\|,$$

where $\boldsymbol{s}^{(1)}(\alpha) = \boldsymbol{s}^{(1)} + \alpha\Delta\boldsymbol{s}^{(1)}$, $\boldsymbol{\lambda}_2^{(1)}(\alpha) = \boldsymbol{\lambda}_2^{(1)} + \alpha\Delta\boldsymbol{\lambda}_2^{(1)}$ and $\boldsymbol{v}^{(1)}(\alpha) = \boldsymbol{v}^{(1)} + \alpha\Delta\boldsymbol{v}^{(1)}$.

We have

$$\varphi^{(1)}(0) = \min_{i=1,\dots,m}\left(S^{(1)}\Lambda_2^{(1)}\boldsymbol{e}_m\right) - \gamma_1\tau_1\left(\frac{\boldsymbol{s}^{(1)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(1)}}{m}\right),$$

$$\psi^{(1)}(0) = \boldsymbol{s}^{(1)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(1)} - \gamma_1\tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(1)})\|.$$

Since

$$\varphi^{(0)}(\alpha_0) = \min_{i=1,\dots,m}\left(S^{(1)}\Lambda_2^{(1)}\boldsymbol{e}_m\right) - \gamma_0\tau_1\left(\frac{\boldsymbol{s}^{(1)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(1)}}{m}\right) \geqslant 0,$$

$$\psi^{(0)}(\alpha_0) = \boldsymbol{s}^{(1)^{\mathrm{T}}}\boldsymbol{\lambda}_2^{(1)} - \gamma_0\tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(1)})\| \geqslant 0$$

and $\gamma_1 \leqslant \gamma_0$, we have

$$\varphi^{(1)}(0) \geqslant \varphi^{(0)}(\alpha_0) \geqslant 0 \quad \text{and} \quad \psi^{(1)}(0) \geqslant \psi^{(0)}(\alpha_0) \geqslant 0.$$

Thus, Theorem 3 assures that there exist $\hat{\alpha}_1^{(2)} > 0$ and $\check{\alpha}_1^{(2)} > 0$ such that

$$\varphi^{(1)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \hat{\alpha}_1^{(2)}\right],$$

$$\psi^{(1)}(\alpha) \geqslant 0 \quad \text{for all } \alpha \in \left(0, \check{\alpha}_1^{(2)}\right].$$

Hence, we have $\varphi^{(1)}(\alpha_1) \geqslant 0$ and $\psi^{(1)}(\alpha_1) \geqslant 0$, where α_1 is the step-length obtained after the execution of the backtracking procedure.

Thus, in the next steps ($k = 2,3,\dots$) of the process we have

$$\varphi^{(k)}(0) \geqslant \varphi^{(k-1)}(\alpha_{k-1}) \geqslant 0 \quad \text{and} \quad \psi^{(k)}(0) \geqslant \psi^{(k-1)}(\alpha_{k-1}) \geqslant 0.$$

This completes the proof. □

Since the initial iterate $\boldsymbol{v}^{(0)}$ satisfies (13) and (14), then $\boldsymbol{v}^{(0)} \in \Omega(0)$. Then $\Delta\boldsymbol{v}^{(0)}$, satisfying (11) and (12), is well defined. The procedure described in Section 2 and Theorems 2–4 enable us to determine α_0 so that $\|\boldsymbol{H}(\boldsymbol{v}^{(1)})\| < \|\boldsymbol{H}(\boldsymbol{v}^{(0)})\|$ and (13) and (14) hold at $\boldsymbol{v}^{(1)}$. Then $\boldsymbol{v}^{(1)} \in \Omega(0)$. With the same argument, if $\boldsymbol{v}^{(k)} \in \Omega(0)$, then $\boldsymbol{v}^{(k+1)} \in \Omega(0)$. Therefore the sequence $\{\boldsymbol{v}^{(k)}\} \in \Omega(0)$, for $k \geqslant 0$.

Theorem 5. *Under the hypotheses of Theorems* 2, 3, *the Newton IP algorithm, with* $\epsilon = 0$, *generates a sequence* $\{\boldsymbol{v}^{(k)}\}$ *such that*:

(a) *if* $\boldsymbol{v}^*$ *is a limit point of the sequence* $\{\boldsymbol{v}^{(k)}\}$ *such that* $H'(\boldsymbol{v}^*)$ *is nonsingular, then the sequence* $\{\boldsymbol{H}(\boldsymbol{v}^{(k)})\}$ *converges to zero and* $\boldsymbol{v}^{(k)}$ *converges to* $\boldsymbol{v}^*$ *when k diverges*;

(b) *the sequence* $\{\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\}$ *converges to zero and each limit point of the sequence* $\{\boldsymbol{v}^{(k)}\}$ *satisfies the KKT conditions for (1) and (2);*

(c) *if the sequence* $\{\boldsymbol{v}^{(k)}\}$ *converges to* $\boldsymbol{v}^*$ *with* $H'(\boldsymbol{v}^*)$ *nonsingular matrix,* $\sigma_k = \mathcal{O}(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^{\xi})$, $0 < \xi < 1$, *and* $\delta_k = \mathcal{O}(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|)$, *then there exists an index* $\bar{k}$ *such that* $\alpha_k = 1$ *for* $k \geqslant \bar{k}$. *Thus, the Newton IP algorithm has a superlinear local convergence.*

Proof. Part (a) Since $\boldsymbol{v}^{(k)} \in \Omega(\epsilon)$, $\epsilon > 0$, then, by Theorem 2, $\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \neq 0$ and $H'(\boldsymbol{v}^{(k)})$ is nonsingular. Therefore, the method is well defined and determines a new point at each iteration k. Since theorems 2, 3 and 4, the Newton IP step $\alpha_k\Delta\boldsymbol{v}^{(k)}$ satisfies the residual condition of the inexact Newton method with forcing term uniformly bounded by 1 (see Section 2) and the condition on the reduction of the norm (37). Thus, from [20, p. 70] (or [9, Theor. 6.1]) we have the result.

Part (b) (see [8, Theor. 3.1]). The Newton IP method generates in $\Omega(\epsilon)$, $\epsilon > 0$ a sequence $\{\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\}$ which is monotone nonincreasing, then, bounded. Consequently, this sequence has a limit point, say, $H^* \in \mathbb{R}$. If it is equal to zero, we have the result. Suppose that $H^* \neq 0$, then the sequence $\{\boldsymbol{v}^{(k)}\}$ and its limit point belong to $\Omega(\epsilon)$, with $\epsilon = (H^*)^2 > 0$. If $\boldsymbol{v}^*$ is one this limit points, we have that $H'(\boldsymbol{v}^*)$ is a nonsingular matrix, then from part (a) of this theorem we deduce that $\boldsymbol{H}(\boldsymbol{v}^*) = 0$. This contradicts the assumption that $H^* > 0$. Hence, the sequence $\{\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\}$ must converges to zero.

Part (c) (see [7]). From (11), (9) and (12), we have

$$\|\Delta\boldsymbol{v}^{(k)}\| \leqslant \|H'(\boldsymbol{v}^{(k)})^{-1}\|(1 + \sigma_k + \delta_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|,$$

where $H'(\boldsymbol{v}^{(k)})^{-1}$ is a bounded matrix, then, for $k \geqslant \bar{k}$, we have, $\|\Delta\boldsymbol{s}^{(k)}\| = \mathcal{O}\big(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\big)$ and $\|\Delta\boldsymbol{\lambda}_2^{(k)}\| = \mathcal{O}\big(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\big)$. Then, for k sufficiently large, the conditions (13) and (14) are satisfied for $\alpha_k^{(2)} = 1$. Indeed, for k sufficiently large, $\Delta s_i^{(k)} < 0$ and $\Delta\lambda_{2,i}^{(k)} < 0$ are negligible with respect $s_i^{(k)}$ and $\lambda_{2,i}^{(k)}$ and then $\alpha_k^{(1)} = 1$.

Furthermore, from the definition of $\varphi_i^{(k)}(\alpha)$ and (31), we observe that

$$\begin{aligned}\varphi_i^{(k)}(1) &= s_i^{(k)}(1)\lambda_{2,i}^{(k)}(1) - (\gamma_k\tau_1/m)(\boldsymbol{s}^{(k)}(1))^{\mathrm{T}}\boldsymbol{\lambda}_2^{(k)}(1)\\ &\geqslant \sigma_k\mu_k(1 - \tau_1\gamma_k) - (1 + \tau_1\gamma_k/m)\|\Delta\boldsymbol{s}^{(k)}\|\|\Delta\boldsymbol{\lambda}_2^{(k)}\|.\end{aligned}$$

Since, from (9) and (20), we have

$$\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|/((1 + 1/(\gamma_k\tau_2))m) \leqslant \mu_k \leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|/\sqrt{m},$$

then $\mu_k = \mathcal{O}\big(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\big)$ and $\sigma_k\mu_k = \mathcal{O}\Big(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^{\xi+1}\Big)$, while $\|\Delta\boldsymbol{s}^{(k)}\|\|\Delta\boldsymbol{\lambda}^{(k)}\| = \mathcal{O}\Big(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2\Big)$. Hence the criterion (13) is satisfied for $\hat{\alpha}_k^{(2)} = 1$, with k sufficiently large.

As far as the criterion (14) is concerned,

$$\begin{aligned}\psi^{(k)}(1) &= (\boldsymbol{s}^{(k)}(1))^{\mathrm{T}}\boldsymbol{\lambda}_2^{(k)}(1) - \tau_2\|\boldsymbol{H}_1(\boldsymbol{v}^{(k)}(1))\| \\ &\geqslant m\sigma_k\mu_k - \left(\gamma_k\tau_2\delta_k\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| + (1+\gamma_k\tau_2)\|\Delta\boldsymbol{v}^{(k)}\|^2\right),\end{aligned}$$

so, for sufficiently large k, $\check{\alpha}_k^{(2)} = 1$ satisfies (14).

Then $\alpha_k^{(2)} = \min\left(\alpha_k^{(1)}, \hat{\alpha}_k^{(2)}, \check{\alpha}_k^{(2)}, 1\right) = 1$.

Now we prove that the backtracking procedure determines $\alpha_k = 1$ for sufficiently large k.

$$\begin{aligned}\|\boldsymbol{H}(\boldsymbol{v}^{(k)}(1))\| &= \|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \Delta\boldsymbol{v}^{(k)})\|, \\ &\leqslant \|\boldsymbol{H}(\boldsymbol{v}^{(k)} + \Delta\boldsymbol{v}^{(k)}) - (\boldsymbol{H}(\boldsymbol{v}^{(k)}) + H'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v}^{(k)})\| + \|\boldsymbol{H}(\boldsymbol{v}^{(k)}) \\ &\quad + H'(\boldsymbol{v}^{(k)})\Delta\boldsymbol{v}^{(k)}\|.\end{aligned}$$

For the Lemma 2.2 in [1] (see also the footnote in p. 403) and from the residual condition (10) with forcing term $\eta_k = \sigma_k + \delta_k$, it follows that

$$\begin{aligned}\|\boldsymbol{H}(\boldsymbol{v}^{(k)}(1))\| &\leqslant o(\|\Delta\boldsymbol{v}^{(k)}\|) + (\delta_k + \sigma_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \\ &= o(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|) + (\delta_k + \sigma_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|.\end{aligned}$$

Hence, we have

$$\begin{aligned}&(1-\beta(1-(\delta_k+\sigma_k)))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - \|\boldsymbol{H}(\boldsymbol{v}^{(k)}(1))\| \\ &\quad\geqslant (1-\beta)(1-(\delta_k+\sigma_k))\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - o(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|) \\ &\quad= (1-\beta)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - (1-\beta)(\delta_k+\sigma_k)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - o(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|) \\ &\quad= (1-\beta)\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| - (\mathcal{O}(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^{1+\xi}) + \mathcal{O}(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|^2)) - o(\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|) \\ &\quad\geqslant 0.\end{aligned}$$

Then, there exists an index $\bar{k} \geqslant 0$ such that $\alpha_k = 1$ for all $k \geqslant \bar{k}$. It follows that

$$\eta_k = 1 - \alpha_k(1-(\delta_k+\sigma_k)) = \delta_k + \sigma_k, \quad \text{for } k \geqslant \bar{k}$$

and then, from Corollary 3.5(a) in [1], the sequence $\{\boldsymbol{v}^{(k)}\}$ converges to $\boldsymbol{v}^*$ superlinearly. □

4. Solution of the KKT indefinite system in condensed form

When we have to solve NLP problems as those in [16–18], where the inequality constraints are simple box constraints, it is convenient to reduce the inner linear system (7) in the form (16); indeed, in this case, the term $CS^{-1}\Lambda_2C^{\mathrm{T}}$ of the matrix A is a diagonal matrix.

A crucial point for the well definition of the algorithm is that the matrix $H'(\boldsymbol{v})$ is nonsingular in any compact set of $\Omega(0)$ where $\boldsymbol{s}$ is bounded away from zero (hypothesis C3). This hypothesis holds if the coefficient matrix of the system (16) is nonsingular in the same sets. This matrix is nonsingular if and only if the matrix $Z^{\mathrm{T}}AZ$ is nonsingular, where Z is the $m \times (n - neq)$ matrix such that $B^{\mathrm{T}}Z = 0$ and $Z^{\mathrm{T}}Z = I$, i.e. the columns of Z form an orthogonal basis of the null space of B^{T} ([12]).

The following theorem states two sufficient conditions to assure that this matrix is nonsingular.

Theorem 6. *The coefficient matrix in (16) is nonsingular if one of the following conditions hold*:

- $C3'$ *the matrices A and $B^{\mathrm{T}}A^{-1}B$ are nonsingular*;
- $C3''$ *B^{T} is a full row–rank matrix and A is positive definite on the null space of B^{T}: $\mathcal{N}(B^{\mathrm{T}}) = \{\boldsymbol{x} \in \mathbb{R}^n : B^{\mathrm{T}}\boldsymbol{x} = 0\}$.*

Proof. If $C3'$ holds, it is immediate to prove that the following matrix is the inverse of the coefficient matrix in (16):

$$\begin{pmatrix} A^{-1} - A^{-1}B(B^{\mathrm{T}}A^{-1}B)^{-1}B^{\mathrm{T}}A^{-1} & A^{-1}B(B^{\mathrm{T}}A^{-1}B)^{-1} \\ (B^{\mathrm{T}}A^{-1}B)^{-1}B^{\mathrm{T}}A^{-1} & -(B^{\mathrm{T}}A^{-1}B)^{-1} \end{pmatrix}.$$

For the condition $C3''$, see [15, p. 424]. □

Under the hypothesis $C3''$, setting $\boldsymbol{y}_1 = \Delta\boldsymbol{x}$ and $\boldsymbol{y}_2 = \Delta\boldsymbol{\lambda}_1$, the system (16), can be viewed as the Lagrange necessary conditions for the minimum point of the following quadratic problem

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\boldsymbol{y}_1^{\mathrm{T}}A\boldsymbol{y}_1 - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{y}_1, \\ &\text{subject to} \quad B^{\mathrm{T}}\boldsymbol{y}_1 - \boldsymbol{q} = 0. \end{aligned}$$

This quadratic problem can be solved efficiently by Hestenes multipliers' scheme ([13, p. 308]), that consists in updating the dual variable by the rule

$$\boldsymbol{y}_2^{(j+1)} = \boldsymbol{y}_2^{(j)} + \chi(B^{\mathrm{T}}\boldsymbol{y}_1^{(j)} - \boldsymbol{q}),$$

where χ is a positive parameter (penalty parameter) and $\boldsymbol{y}_1^{(j)}$ minimize the augmented lagrangian function of the quadratic problem

$$\mathcal{L}_\chi(\boldsymbol{y}_1, \boldsymbol{y}_2) = \frac{1}{2}\boldsymbol{y}_1^{\mathrm{T}}A\boldsymbol{y}_1 - \boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{c} + \boldsymbol{y}_2^{\mathrm{T}}(B^{\mathrm{T}}\boldsymbol{y}_1 - \boldsymbol{q}) + \frac{\chi}{2}(B^{\mathrm{T}}\boldsymbol{y}_1 - \boldsymbol{q})^{\mathrm{T}}(B^{\mathrm{T}}\boldsymbol{y}_1 - \boldsymbol{q}).$$

This means that $\boldsymbol{y}_1^{(j)}$ is the solution of the linear system of order n

$$(A + \chi BB^{\mathrm{T}})\boldsymbol{y}_1 = -B\boldsymbol{y}_2^{(j)} + \boldsymbol{c} + \chi B\boldsymbol{q}. \tag{38}$$

Note that, since B^T has full row–rank, the null space of BB^T is equal to the null space of B^T, then the matrix A is positive definite on the null space of BB^T. Then, it is immediate the following theorem.

Theorem 7 [15, p. 408]. *There exists a positive parameter* χ^* *such that for all* $\chi > \chi^*$, *the matrix* $A + \chi B\, B^T$ *is positive definite.*

This result enables us to solve the system (38) by applying a Cholesky factorization.

In order to choose the parameter χ, we observe that, $\forall \boldsymbol{x} \neq 0$, we must have $\boldsymbol{x}^T(A + \chi BB^T)\boldsymbol{x} > 0$. When $B^T\boldsymbol{x} = 0$, we have, for the hypothesis C3″, $\boldsymbol{x}^T A\boldsymbol{x} > 0$. If $B^T\boldsymbol{x} \neq 0$, $\boldsymbol{x}^T BB^T\boldsymbol{x} > 0$. Then, it follows that

$$\chi > \max\left(0, \max_{x \notin \mathcal{N}(B^T)} \frac{-\boldsymbol{x}^T A\boldsymbol{x}}{\boldsymbol{x}^T BB^T\boldsymbol{x}}\right).$$

Since $\|A\| \geqslant (-\boldsymbol{x}^T A\boldsymbol{x})/(\boldsymbol{x}^T\boldsymbol{x})$ for any natural norm and also for the Frobenius norm $\|\cdot\|_F$, and $\boldsymbol{x}^T BB^T\boldsymbol{x}/(\boldsymbol{x}^T\boldsymbol{x}) \geqslant \tau_{\min}$, where $\tau_{\min}$ is the minimum nonzero eigenvalue of BB^T or of B^TB, we can choose as χ the following value:

$$\chi > \frac{\|A\|_F}{\tau_{\min}}.$$

In general it is difficult to determine an estimate of $\tau_{\min}$. Numerical evidence shows that a good approximation of $\tau_{\min}$ is $\min(1,t_{\min})$, where $t_{\min}$ is the minimum diagonal entry of the matrix B^TB, although $t_{\min} \geqslant \tau_{\min}$. Furthermore, in order to avoid that the value of χ is too small (the matrix is not positive definite) or too large (too ill–conditioned system), it is convenient to use safeguards. In the numerical experiments of the next section, the following value of χ produced good results:

$$\chi = \min\left(\max\left(10^7, \frac{\max\{\|A\|_F, 1\}}{\min\{t_{\min}, 1\}}\right), 10^8\right). \tag{39}$$

5. Numerical results

In order to evaluate the effectiveness of the Newton IP method, a Fortran 90 code, implementing the method, has been carried out on HP zx6000 workstation with Itanium2 processor 1.3 GHz and 2 Gb of RAM. The code has been compiled with a +O3 optimization option of the Fortran HP compiler.

In this code, the Hessian matrix Q of the lagrangian function and the Jacobian matrices B^T and C^T of the equality and inequality constraints are stored in a column compressed format ([21]). Then, in order to compute the matrices $A = Q + C\,S^{-1}\Lambda_2 C^T$ and $A + \chi B\,B^T$, the code executes a preprocess procedure.

The preprocess routine builds a data structure storing the indices of the nonzero entries of the above matrices. For any nonzero entry, in the same data structure we also store the pairs of indices of the elements of B and B^T(and C and C^T respectively) that give a nonzero contribution in the scalar products. The preprocess routine also computes the symbolic Cholesky factorization of $A + \chi B B^T$. Indeed, at each Hestenes iteration, it is necessary to solve a linear system where the coefficient matrix of order n, $A + \chi B B^T$, is sparse, symmetric, positive definite. To exploit the sparsity of $A + \chi BB^T$, its factorization is obtained by a Fortran package (version 0.3) of Ng and Peyton (included in the package LIPSOL, downloadable from http://www.cam.rice.edu/~zhang/lipsol). This package computes a priori the symbolic factor of $A + \chi B B^T$, using the multiple minimum degree ordering of Liu to minimize the fill-ins in this factor and the supernodal block factorization to take advantage of modern computer architectures ([14]). The *a priori* symbolic factorization is executed in the preprocess routine.

The code uses $\mu_k^{(1)}$ as perturbation parameter; at each outer iteration the damping parameter α_k is initially chosen equal to 1, then it is eventually reduced in order to satisfy the feasibility condition, the centrality conditions (13) and (14) and the backtracking strategy (36) with $\beta = 10^{-4}$. The factor of reduction of α_k is $\theta = 0.5$.

The Newton IP method stops when

$$\|\boldsymbol{H}(\boldsymbol{v}^{(k)})\| \leqslant 10^{-8},$$

or when (see [22])

$$\frac{|\mathrm{gap}|}{1 + |\mathrm{gap}|} \leqslant 10^{-8},$$

where "gap" is the difference between the primal function $f(\boldsymbol{x})$ and the dual function

$$d(\boldsymbol{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = f(\boldsymbol{x}) - \boldsymbol{\lambda}_2^T \boldsymbol{g}_2(\boldsymbol{x}) - \boldsymbol{\lambda}_1^T \boldsymbol{g}_1(\boldsymbol{x}) - \nabla f(\boldsymbol{x})^T \boldsymbol{x} + \begin{pmatrix} \boldsymbol{\lambda}_1^T & \boldsymbol{\lambda}_2^T \end{pmatrix} \begin{pmatrix} \nabla \boldsymbol{g}_1(\boldsymbol{x})^T \\ \nabla \boldsymbol{g}_2(\boldsymbol{x})^T \end{pmatrix} \boldsymbol{x}.$$

The inner Hestenes solver stops if the following rule is satisfied

$$\|\boldsymbol{r}^{(k)}\| \leqslant \max\left(5 \cdot 10^{-8}, \delta_k \|\boldsymbol{H}(\boldsymbol{v}^{(k)})\|\right),$$

or if a maximum number is reached; in the code, the maximum number is fixed equal to 6.

Numerical experiments have been carried out using the code on a set of test problems described in [16], [17] and [18]. In these cases, the matrix $C\, S^{-1} \Lambda_2\, C^T$ is a diagonal matrix and in the preprocess phase only the matrix BB^T has to be considered.

In Table 1, we report the references of the considered test problems. The number of variables n and the number of the equality constraints neq depend on a parameter N which represents the number of the mesh points for each dimension of the square domain of the control problem. The suffix in the name of the tests problems is the value of N.

In Table 2, for each test problems, we report the values of n, neq, the number of nonzero entries $nnzq$ and $nnzb$ of Q and B respectively, the number nnz of the nonzero entries of the lower triangular part of $A + \chi BB^{\mathrm{T}}$ and the number $nnzl$ of the nonzero entries of the Cholesky factor of $A + \chi BB^{\mathrm{T}}$. We observe that, because of the structure of B, the matrix–matrix product BB^{T} does not

Table 1
Description of the test-problems

Test problems	References
TP1-N	[18, 4.2, p. 191] $M = 1$, $K = 0.8$, $b = 1$, $u_1 = 1.7$, $u_2 = 2$, $\psi(x) = 7.1$
TP2-N	[17, 4.4, Example 4, p. 153]
TP3-N	[16, 5.2, Example 5.5, p. 47]
TP4-N	[16, 5.2, Example 5.7, p. 51]

Table 2
Values of n, neq, nonzero entries in Q, in B, nonzero entries in the triangular part of $A + \chi BB^{\mathrm{T}}$ and in its Cholesky factor

Test problems	n	neq	$nnzq$	$nnzb$	nnz	$nnzl$
TP1-99	19,602	9801	39,204	58,410	72,816	715,465
TP1-199	79,202	39,601	158,404	236,810	295,620	3,409,660
TP1-299	178,802	89,401	357,604	535,210	1,158,029	8,900,195
TP1-399	318,402	159,201	636,804	953,610	2,064,029	20,090,160
TP1-499	498,002	249,001	996,004	149,2010	3,230,029	28,768,781
TP2-99	19,998	10,197	19,602	38,214	128,401	717,837
TP2-199	79,998	40,397	79,202	156,414	516,801	3,414,432
TP2-299	179,998	90,597	178,802	354,614	1,165,201	8,907,367
TP2-399	319,998	160,797	318,402	632,814	2073601	20,099,732
TP3-99	10,593	10,197	10,593	30,789	70,783	622,759
TP3-199	41,193	40,397	41,193	121,589	281,583	3,181,444
TP3-299	91,793	90,597	91,793	272,389	632,383	8,374,469
TP3-399	162,393	160,797	162,393	483,189	1,123,183	16,252,152
TP3-499	252,993	250,997	252,993	753,989	1,753,983	26,855,490
TP4-99	10,593	10,197	10,197	30,789	70,783	622,759
TP4-199	41,193	40,397	40,397	121,589	281,588	3,181,444
TP4-299	91,793	90,597	90,597	272,389	632,383	8,374,469
TP4-399	162,393	160,797	160,797	483,189	1,123,183	16,252,152
TP4-499	252,993	250,997	250,997	753,989	1,753,983	26,855,490

give rise to an excessive number of nonzero entries and the matrix $A + \chi BB^T$ is very sparse with a density at most equal to 0.06%. Furthermore the ratio of the nonzero entries in the Cholesky factor and in the matrix $A + \chi BB^T$ is at most equal to 15.3.

In Table 3, we report the results of the Newton IP method when we use the Hestenes multipliers' iterative scheme. In this table, *it* represents the number of outer iterations of Newton IP method. The total number of inner iterations of the Hestenes method is reported in brackets. The execution time, expressed in seconds, is subdivided into two parts, the *preprocess time* and the time for computing the solution (*solution time*). Table 3 shows that the Newton IP–Hestenes scheme is able to solve high–dimensional and sparse NLP problems. The code is efficient from the point of view of the memory usage and of the execution time. Actually, the more expensive computational task is the preprocess phase, which is dependent on the strategy used to perform the matrix–matrix products needed in the method.

Furthermore, in all test problems, the total execution time of the Newton IP–Hestenes method is significantly less than the one of the Newton IP method which uses a direct inner solver as the MA27 subroutine of the Harwell Sub-

Table 3
Results of Newton IP-Hestenes

Test problems	Newton IP–Hestenes	
	it	Preprocess time + solution time
TP1-99	28(29)	5.77 + 2.7
TP1-199	48(49)	118.03 + 25.11
TP1-299	81(111)	686.30 + 131.49
TP1-399	102(153)	2292.11 + 477.5
TP1-499	101(166)	5496.66 + 699.3
TP2-99	13(29)	6.09 + 1.51
TP2-199	15(50)	118.07 + 10.38
TP2-299	16(46)	634.28 + 30.59
TP2-399	17(47)	2109.27 + 80.5
TP3-99	29(32)	2.22 + 2.03
TP3-199	54(59)	36.38 + 22.87
TP3-299	181(186)	206.35 + 246.8
TP3-399	327(341)	833.79 + 961.08
TP3-499	501(527)	1933.793 + 2768.707
TP4-99	21(23)	3.02 + 1.46
TP4-199	26(27)	47.83 + 10.87
TP4-299	39(45)	162.15 + 52.88
TP4-399	36(39)	831.0 + 105.29
TP4-499	65(87)	2062.11 + 360.03

routine Library. This routine solves the symmetric system by a sparse symmetry preserving Bunch–Parlett triangular factorization ([3]).

For instance, the execution time of Newton IP–MA27 method is 24.71 and 304.11 seconds for TP1-99 and TP1-199 respectively with 25 and 26 iterations in the two cases respectively. For the test problems TP3-99, TP3-199, TP4-99 and TP4-199, we have 27.38, 349.66, 22.52, 250.28 seconds respectively with 29, 37, 24, 27 iterations in all these cases respectively. In the other cases of Table 3, the Newton IP–MA27 method fails for exceeded memory requirements. Finally, we observe that in the larger test problems, the method with the direct solver, when it works, requires less iterations.

6. Conclusions

We have discussed the use of an inexact Newton method for the solution of nonlinear programming problems arising from elliptic control problems. At each step of this method we introduced Hestenes multipliers' scheme as iterative solver for the inner indefinite KKT system. Consequently the solution of this system is led to a solution of a sequence of positive definite systems. We devised conditions to assure the global convergence of the whole method. As shown in Table 3, few inner iterations of Hestenes scheme per outer iteration are sufficient to satisfy the inner stopping rule.

References

[1] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, SIAM J. Numer. Anal. 19 (1982) 400–408.
[2] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[3] I.S. Duff, J.K. Reid, MA27-a set of Fortran subroutines for solving sparse symmetric sets of linear equations, Tech Report AERE R10533, HMSO, London, 1982.
[4] C. Durazzi, On the Newton interior-point method for nonlinear programming problems, J. Optim. Theory Appl. 104 (2000) 73–90.
[5] C. Durazzi, E. Galligani, Nonlinear programming methods for solving optimal control problems, in: F. Giannessi, A. Maugeri, P.M. Pardalos (Eds.), Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models, Nonconvex Optimization and Its Applications, vol. 58, Kluwer Academic Publ., (2001) (2001) 71–99.
[6] C. Durazzi, V. Ruggiero, Global convergence of the Newton interior-point method for nonlinear programming, J. Optim. Theory Appl. 120 (2004) 199–208.
[7] C. Durazzi, V. Ruggiero, A Newton inexact interior-point method for large scale nonlinear optimization problems, Annali Univ. Ferrara, Sez. VII, Sc. Matem. IL, (2003) 333–357.
[8] C. Durazzi, V. Ruggiero, G. Zanghirati, Parallel interior-point method for linear and quadratic programs with special structure, J. Optim. Theory Appl. 110 (2001) 289–313.
[9] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, SIAM J. Optim. 4 (1994) 393–422.

[10] A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, Y. Zhang, On the formulation and theory of Newton interior-point method for nonlinear programming, J. Optim. Theory Appl. 89 (1996) 507–541.
[11] E. Galligani, Analysis of the convergence of an inexact Newton method for solving Karush–Kuhn–Tucker systems, Atti Sem. Matem. Fis. Univ. Modena, in press.
[12] N.I.M. Gould, On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem, Math. Program. 32 (1985) 90–99.
[13] M.R. Hestenes, Optimization Theory. The Finite Dimensional Case, J. Wiley & Sons, New York, 1975.
[14] J.W. Liu, E.G. Ng, B.W. Peyton, On finding supernodes for sparse matrix computations, SIAM J. Matrix Anal. Appl. 14 (1993) 242–252.
[15] D.G. Luenberger, Linear and Nonlinear Programming, second ed., Addison-Wesley, Reading, MA, 1984.
[16] H. Maurer, H.D. Mittelmann, Optimization techniques for solving elliptic control problems with control and state constraints: Part 1. Boundary control, Comput. Optim. Appl. 16 (2000) 29–55.
[17] H. Maurer, H.D. Mittelmann, Optimization techniques for solving elliptic control problems with control and state constraints: Part 2. Distributed control, Comput. Optim. Appl. 18 (2001) 141–160.
[18] H.D. Mittelmann, H. Maurer, Solving elliptic control problems with interior point and SQP methods: control and state constraint, J. Comput. Appl. Math. 120 (2000) 175–195.
[19] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.
[20] W.C. Rheinboldt, Methods for Solving Systems of Nonlinear Equations, second ed., SIAM, Philadelphia, PA, 1998.
[21] Y. Saad, Iterative Methods for Sparse Linear System, PSW Publ. Co., Boston, MA, 1996.
[22] R.J. Vanderbei, D.F. Shanno, An interior-point algorithm for nonconvex nonlinear programming, Comput. Optim. Appl. 13 (1999) 231–252.